Genetic Risk Prediction for Common Diseases

Methodology and Applications

Raluca Mihăescu

The work described in this thesis was conducted at the Department of Epidemiology of the Erasmus MC, Rotterdam, the Netherlands.

The studies presented in this thesis were funded by a VIDI grant of the Netherlands Organisation for Scientific Research (NWO) and by a grant from the Centre for Medical Systems Biology (CMSB) in the framework of the Netherlands Genomics Initiative (NGI).

The Rotterdam Study is supported by the Erasmus Medical Center and Erasmus University Rotterdam; the Netherlands Organisation for Scientific Research (NWO), the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Netherlands Heart Foundation, the Ministry of Education, Culture and Science, the Ministry of Health, Welfare and Sports, the European Commission (DG XII) and the Municipality of Rotterdam. The GWAS database of the Rotterdam Study was funded through the Netherlands Organisation of Scientific Research NWO (grant numbers 175.010.2005.011, 911.03.012) and the Research Institute of Diseases in the Elderly (RIDE).

Financial support for the publication of this thesis was kindly provided by the Department of Epidemiology of the Erasmus Medical Center and the Erasmus University Rotterdam.

Cover: Alexandru Ioniță, Bucureşti, Romania
Lay-out: Legatron Electronic Publishing, Rotterdam
Printing: Ipskamp Drukkers, Enschede

ISBN: 978-94-6191-799-7

Genetic Risk Prediction for Common Diseases

Methodology and Applications

Genetische risicovoorspelling van veelvoorkomende ziekten methodologie en toepassingen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.G. Schmidt

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

Friday, 27th of September 2013 at 9.30 o' clock

by

Raluca Mihăescu
born in Iaşi, Romania

ERASMUS UNIVERSITEIT ROTTERDAM

DOCTORAL COMMITTEE

Promoters:	Prof.dr. A.C.J.W. Janssens
	Prof.dr. M.G. Hunink
Other Members:	Prof.dr. H. Tiemeier
	Prof.dr. E.W. Steyerberg
	Prof.dr. M.C. Cornel

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."

Marcel Proust

To my parents

PUBLICATIONS AND MANUSCRIPTS IN THIS THESIS

Chapter 2

Mihaescu R, van Zitteren M, van Hoek M, Sijbrands EJ, Uitterlinden AG, Witteman JC, Hofman A, Hunink MG, van Duijn CM, Janssens AC. Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve. American Journal of Epidemiology. 2010;172(3):353-61.

Chapter 3

Mihaescu R, Moonesinghe R, Khoury MJ, Janssens AC. Predictive genetic testing for the identification of high-risk groups: a simulation study on the impact of predictive ability. Genome Medicine. 2011;3(7):51.

Chapter 4

Kundu S, **Mihaescu R**, Catelijne Meijer, Bakker R, Janssens AC. Estimating the predictive ability of genetic risk models using data from genome-wide association studies. In preparation.

Chapter 5

Mihaescu R, Pencina MJ, Alonso A, Lunetta KL, Heckbert SR, Benjamin EJ, Janssens AC. Incremental value of rare genetic variants for the prediction of multifactorial diseases. Genome Medicine. 2013;5(8):76.

Chapter 6

Iglesias AI, **Mihaescu R**, Ioannidis JP, Khoury MJ, Little J, van Duijn CM, Janssens AC. Completeness of reporting in genetic risk prediction studies: A review of published articles based on the GRIPS statement. Submitted.

Chapter 7

Mihaescu R, van Hoek M, Sijbrands EJ, Uitterlinden AG, Witteman JC, Hofman A, van Duijn CM, Janssens AC. Evaluation of risk prediction updates from commercial genome-wide scans. Genetics in Medicine. 2009;11(8):588-94.

Chapter 8

Mihaescu R, Navarro P, Prokopenko I, Bragg-Gresham J, Strassburger K, Yengo L, van Hoek M, Amin N, Balkau B, Boehnke M, Campbell H, Charpentier G, Frayling TM, Gögele M, Grallert H, Hayward C, Hicks AA, Hofman A, Illig T, Lindgren CM, Marullo L, Oostra BA, Pichler I, Polasek O, Rayner NW, Rudan I, Sijbrands EJ, Uitterlindenn AG, Wild SH, Willems SM, Jackson AU, Bonnycastle LL, Tuomilehto J, Cauchi S, Rathmann W, Scott L, Froguel P, van Duijn CM, Pramstaller PP, Wright AF, Wilson JF, McCarthy MI, Janssens AC. Comparison of performance of genetic risk prediction models for type 2 diabetes in nine European samples. Submitted.

Chapter 9

Kalf RR, **Mihaescu R**, Kundu S, de Knijff P, Green RC, Janssens AC. Variations in predicted risks in personal genome testing for common complex diseases. Genetics in Medicine. 2013; doi:10.1038/gim.2013.80

Chapter 10

Willems SM, **Mihaescu R**, Sijbrands EJ, van Duijn CM, Janssens AC. A methodological perspective on genetic risk prediction studies in type 2 diabetes: recommendations for future research. Current Diabetes Reports. 2011;11(6):511-8.

Chapter 11

Mihaescu R, Meigs J, Sijbrands E, Janssens AC. Genetic risk profiling for prediction of type 2 diabetes. PLoS Currents. 2011;3:RRN1208.

CONTENTS

Chapter 1

General Introduction

Genetic tests have traditionally been used for the prediction and diagnosis of Mendelian disorders in which the phenotypes are largely determined by mutations at individual loci, i.e. the specific chromosomal location of a gene or other DNA sequence. These disorders are generally rare. Common diseases, in contrast, are caused by a combination of multiple loci and environmental risk factors such as diet, smoking, exercise and stress. Genetic tests are being increasingly investigated for their ability to improve the prediction and management of common diseases. The advances in gene discovery and bioinformatics tools have fueled these efforts. Improvement in management of common diseases would benefit a large number of individuals and reduce the costs of health care at the population level, making genetic tests very useful. This introduction gives an overview of the advances in genetic discoveries, the methodological aspects of genetic risk prediction and the use of genetic tests in the commercial setting. The introduction continues with an elaboration on the current challenges in genetic risk prediction and concludes with a summary of the research presented in this thesis.

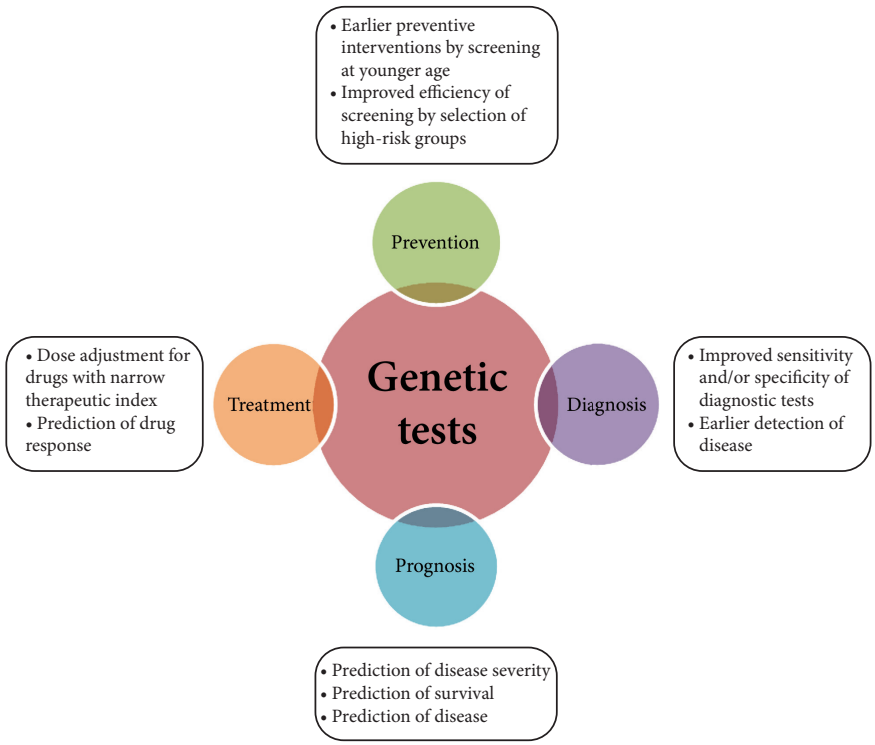
Progress in genomics and gene discovery

Genetic discoveries in common diseases have seen an enormous progress with the advent of genome wide association studies (GWASs). GWASs are designed to detect small DNA variations, single nucleotide polymorphisms (SNPs), that are associated with risk of common diseases. A SNP is a DNA sequence variation occurring in a single nucleotide of the genome (A, T, C, or G) that differs between individuals. Since their introduction in 2007, GWASs have been very fruitful; the number of statistically significant SNP-disease associations in 2013 being more than 8700 for numerous traits (1), including type 2 diabetes, cardiovascular disease, cancer and many other common diseases (2-4). Variants discovered in GWASs are common and typically have small effects on disease risk, with most of the SNPs discovered in recent GWASs having an odds ratio (OR) close to 1.1 (5). Future GWASs efforts are thought to uncover many SNPs with similar effect sizes. New statistical methods and technologies are explored to detect stronger genetic risk variants.

In the future, next generation sequencing will provide a massive throughput in gathering of genetic information since, in contrast to genotyping platforms that are based on common polymorphisms only, it reads every nucleotide and produce thousands or millions of sequences at once. This technology has applications in the sequencing of the whole genome or of the exome (i.e. the coding part of the genome). It is expected that next generation sequencing will expedite the identification of rare causal variants for Mendelian diseases and unravel rare variants with stronger effects for common diseases (6-22).

The rapid success of GWASs in identifying SNPs and the introduction of whole exome and genome sequencing technologies have boosted the expectations of personalized medicine (23). It is envisioned that genetic tests will personalize medicine through individualized lifestyle and preventive interventions for high-risk individuals and targeted treatment to the individual genetic profile (see **Figure 1**) (24-28). Examples of possible applications include the identification of genetic profiles associated with increased risk to develop disease and of profiles that predict the prognosis of disease, disease recurrence rates or disease severity (29). Another domain that holds much promise for the application of genetic tests is pharmacogenomics (30), which includes the identification of genetic profiles to select the effective medication or the dose (i.e., the genotype at loci that control the metabolization of drugs in the liver), the identification of genetic profiles that are associated with increased side-effects of medication and toxicity, and of genetic profiles that determine the availability of the drugs to specific tissues and organs.

Figure 1. Potential role of genetic testing in the clinical and public health settings (24-28).



This thesis focuses on a broad application of genetic tests, namely genetic risk prediction. Before discussing genetic risk prediction it is essential to comment on the fundamental difference between the architecture of Mendelian and common diseases, which has an impact on the applicability of genetic tests for personalized medicine.

Genetic architecture of Mendelian diseases

Mendelian diseases are caused, by definition, by a single mutation in a causal gene. A mutation in the causal gene is associated with a gain or loss of function of the gene that plays a major part in the development of disease. An example is the CAG repeat expansion in a gene that codes for huntingtin, a ubiquitous protein, that leads to development of Huntington disease in all mutation carriers (31). Most Mendelian mutations are fully penetrant, meaning that each individual carrying the mutation eventually develops the disease, but many causal genetic variants for monogenic diseases have an incomplete penetrance. This is observed for example in autosomal dominant Alzheimer's disease where some individuals that carry mutations in the presenilin 2 (*PSEN2*) gene do not develop the disease (32). In conclusion, for monogenic diseases the causal chain between the genetic mutation, the changes in protein structure and/or function and the development of disease is relatively straightforward.

Genetic architecture of common diseases

Unlike monogenic diseases, multifactorial diseases are caused by complex interactions between multiple genetic and environmental risk factors. Furthermore, in contrast to the straight link between mutations and disease in monogenic diseases, for most common variants the pathophysiological mechanisms have not been identified, partly because many of these variants lay in non-coding regions of the genome. The total contribution of genetic factors for the risk of disease is given by the heritability of disease. The heritability of disease is defined as the proportion of phenotypic variance in a population attributable to additive genetic factors (33). Despite the moderate to high heritability estimates for common diseases and the large number of common susceptibility variants discovered by GWAS, only a small part of the heritability of common diseases can be currently accounted for (3, 29, 34). One explanation is that susceptibility variants with very small effects on disease risk have not been picked up by GWASs because of the stringent criteria of the significance threshold by correcting for multiple testing. For example, the significantly associated SNPs explain 1% of the heritability of schizophrenia (35), while all SNPs considered simultaneously explain 30% of heritability (36). The "missing" heritability may also be explained by the fact that other genetic risk factors that play a major role in the risk for common diseases exist, but current technologies have not been as successful in discovering

them. These factors include structural variations (which comprise copy number variation, deletions and insertions), haplotypes and rare variants (37, 38). Furthermore, evidence for the role of other types of genetic factors on disease risk, such as mitochondrial DNA and micro RNAs, has emerged (39, 40). Another aspect of the genetic architecture of common diseases is the existence of interactions between the various risk factors. Examples include gene-gene and gene-environment interactions (41), parent-of-origin effects, i.e., the differential expression of alleles based on their parent of origin (42), and the epigenetic mechanisms that include heritable methylation patterns that influence gene expression (43). Genetics of common diseases has thus currently many aspects that remain poorly understood. The maximum predictive ability of genetic tests has not been reached because only a small part of the genetic risk factors that account for the heritability of diseases are currently known. With advanced technologies and understanding of genetic factors it is expected that genetic prediction can improve.

Genetic risk prediction

Prediction of monogenic diseases is done by testing individuals for known causal mutations. Usually, identification of the mutations implies that the individual will develop the disease. However, there are also cases when carriers of a mutation will not develop the disease or will develop it later in life. For example, three causal genes have been identified for autosomal dominant Alzheimer's disease: amyloid-β protein precursor (*APP*) gene on chromosome 21 (44), presenilin 1 (*PSEN1*) gene on chromosome 14 (45), and *PSEN2* gene on chromosome 1 (46, 47). Almost all carriers of a mutation in the *APP* or *PSEN1* genes will eventually develop the disease, but the mutations in the *PSEN2* gene, as mentioned earlier, are not completely penetrant with carriers showing large variations in the age of onset (32). In the above example, the three genes have been identified in a few hundred families in the world, which together account only for a small percentage of dominant Alzheimer's disease. Sometimes sequencing of the genes in affected individuals is needed to facilitate testing in healthy relatives because risk variants may be family specific.

In contrast to monogenic diseases, prediction of risk for common diseases is based on numerous genetic variants. Due to their low effect sizes, common variants are not useful for risk prediction when used alone (48-50). Therefore, multiple common variants taken together are increasingly being investigated for their potential to predict risk of disease (51-55). The cumulative effect of multiple genetic variants can be quantified by constructing models that include many genetic variants, either combined in risk scores or analyzed together by means of regression analysis. Risk scores are simple counts of risk alleles or weighted count scores, where weights are usually effect sizes estimated from meta-analysis or individual GWAS. The

complexity of the risk models distinguishes the risk prediction for common diseases from the prediction of monogenic disease.

Stepwise evaluation of genetic tests for risk prediction

In 2004, the Centers for Disease Control (CDC) and Prevention launched the EGAPP initiative to evaluate the transition of genetic tests from research to clinical and public health practice. The EGAPP Working Group has published six recommendation statements on genomic applications for breast cancer, colorectal cancer, heart disease, and depression. CDC's Office of Public Health Genomics supported the ACCE Model Project, which proposed a multi-stage evaluation of genetic tests genetic tests. The ACCE model process is composed of a standard set of 44 targeted questions that address the analytic validity of the test, i.e. the test accuracy and reliability in identifying genetic risk variants; the clinical validity, i.e. the test accuracy and reliability in predicting risk of disease; and the clinical utility, i.e. the net benefit of testing in improving health outcomes (see **Table 1** in the Appendix) (56).

These are in line with other criteria for the evaluation of new risk factors (57, 58).

Measures used to evaluate genetic risk prediction models

The different parts of the ACCE model, i.e. the analytical validity, the clinical validity and the clinical utility, are evaluated using the following measures for model assessment and validation: general measures of models fit; calibration measures; sensitivity, specificity, positive and negative predictive value; discrimination measures; and reclassification measures. We discuss below each of these measures.

Measures of model fit and calibration

The *general measures of model fit* indicate how likely the selected model would give rise to the observed data. One example is Akaike Information Criterion (AIC). Apart from showing how models fit to the data the AIC takes also into account the number of predictors in the model (59), which means that between two models with similar predictions, the model with fewer variables would be preferred.

*Calibratio*n indicates how close the risks predicted by the model are to the observed risks. The most commonly used measure of calibration is the Hosmer-Lemeshow (H-L) chi-square test (60). The H-L test compares predicted risk within the corresponding observed frequency, usually within deciles of risk.

Clinical validity

Clinical validity includes sensitivity, specificity, positive and negative predictive value; discrimination, integrated discrimination improvement and validation.

Sensitivity, specificity, positive and negative predictive value are used to evaluate the clinical validity of genetic tests for Mendelian disorders. They can also be used in the case of common diseases, for specific risk cutoff thresholds. The risk cutoff can be a certain value of a risk score or a value of the predicted risk. Sensitivity represents the percentage of individuals with the disease that have a risk score or a predicted risk above the threshold value, specificity the percentage of individuals free of disease below the threshold, positive predictive value the percentage of individuals with the disease among all individuals above the threshold and the negative predictive value the percentage of individuals free of disease among all individuals below the threshold.

Discrimination represents the ability of the model to identify individuals who will eventually develop the disease from individuals who remain free of disease. The area under the receiver operating characteristic curve (AUC) is used to assess discrimination (61). Its value can be interpreted as the probability for the model to assign a higher risk to an individual who will develop disease compared to one who remains free of disease, if a random pair is drawn. Genetic risk models based on known SNPs typically have low to moderate predictive ability for most diseases, as assessed by the AUC, except when one or more variants have a strong effect on disease risk (4). For example, the AUC of a prediction model for prostate cancer using 33 SNPs was 0.59 (62), but the AUC for age-related macular degeneration using 13 risk variants was 0.82 (51). In the former example the ORs for the SNPs included in the model was at most 1.8, while in the latter six SNPs had an OR above 2, with some around 4.

Frequently, genetic tests are evaluated for their ability to improve existing clinical risk models. When SNPs are added to a baseline model, the difference in AUC between the baseline model and the model that includes the SNPs is measured and denoted here as ΔAUC. The increment in AUC with the addition of genetic variants to clinical risk factors is generally very low, usually below 0.10 (63-71). The predictive ability of genetic risk models for common diseases is likely to remain modest, because non-genetic risk factors have a substantial impact on disease risk as well (72, 73).

Another measure of the added value of a new biomarker is the integrated discrimination improvement (IDI) (74), which indicates how much the difference between the mean of the predicted risks for individuals who will develop the disease and the mean for those remaining free of disease increases when a new biomarker is added to a model. Although a reclassification

measure, IDI is a global measure of correct reclassification over all possible cutoff values, and in that sense is comparable to the AUC, and is thus a measure of clinical validity.

Another essential step in the assessment of the clinical validity is the *validation* of the risk model (75). A risk model is more generalizable if calibration and discrimination are similar across samples (76). Two types of validations are possible, that is internal and external validation. Internal validation is performed in the same sample that was used to develop the model. Several techniques for internal validation exist including split-sample, cross-validation and bootstrap re-sampling. Essentially all these methods consist of developing the model in a part of the initial sample and testing it in the remaining part of the sample. External validation means testing the model in an external sample. This is an essential step in model evaluation because it indicates if the model is generalizable across samples or populations.

Clinical utility

Clinical utility is generally assessed by means of *reclassification* measures. Reclassification evaluates the extent to which improvements of risk prediction models influence medical decisions. When prediction models are used for making treatment decisions, predicted risks are categorized using clinically relevant risk cutoff thresholds such as the Framingham risk score for cardiovascular disease (77). Reclassification measures include the reclassification table (78) and the net reclassification improvement (NRI) (74). The percentage of total reclassification represents the percentage of individuals that change risk category out of the total population. NRI is calculated as the sum of differences in proportion of individuals moving up minus the proportion moving down for individuals who will develop the disease, and the proportion of individuals moving down minus the proportion moving up for those who will remain free of disease (74). Overall, improvement in reclassification is indicated by a NRI significantly greater than 0 (74).

Direct-to-consumer personal genome testing

The interest in using genetic variants for risk prediction or guidance of therapeutic interventions has also been picked up by commercial companies who provide genetic tests directly to consumers. These tests are based on genome-wide scans (e.g. 23andMe (79), deCODEme (80)) or even whole genome sequencing (e.g. Knome (81)). From only a swab sample, such companies would predict risks for a wide range of diseases including diabetes, breast cancer and heart disease. Genetic risks are calculated on the basis of literature data. The companies take an average risk from an epidemiological study and multiply this with the ORs from published meta-analyses or large scale GWASs. Information about clinical risk factors is typically not

incorporated into the calculations of disease risk. When available, some companies use sex, ethnicity and age matched population risks to depart from. Several reports of individuals who sent a saliva sample to several companies showed contradictory risk predictions (82, 83). Disagreement between results may reflect the use of different SNPs to predict risk of disease and the use of different average risks. A thorough evaluation of these direct-to-consumer (DTC) tests following the ACCE model is needed to correctly inform the public on their utility.

CURRENT CHALLENGES

Genetic risk prediction for common diseases is a novel field that faces many challenges. The following challenges are addressed in this thesis: First, there is yet no consistency in the methods used to evaluate genetic tests for risk prediction and the utility of the newly developed methods needs to be further evaluated. Second, simulations, which are gaining an increasing role in the field, both for testing genetic models and to compare and develop measures for model performance, need to be evaluated and improved. Third, the literature on genetic risk prediction has seen a tremendous increase and this requires a regulation of reporting in order to facilitate the access and the evaluation of the findings.

Assessment of performance and development of measures

There is a general agreement that methodology used to investigate genetic models varies and needs to be improved (57, 58, 84). Much polemic surrounds one of the most used measures of model assessment, the AUC, which is criticized because even statistically significant new biomarkers yield minimal improvement in AUC (66, 70, 85). Because AUC is considered insensitive, other measures that are deemed more suitable for assessing improvement in prediction have been developed (84, 86). These include the measures of reclassification, which currently receive much attention and are used to determine the clinical utility of genetic risk models (78).

Surprisingly, studies that report both reclassification and AUC highlight the former, especially when reclassification is observed in the absence of improved AUC (63, 66). Whether reclassification reflects improvement in risk prediction when AUC does not change is yet not clear. Furthermore, the relationship between AUC and other model performance measures needs to be further explored before concluding that AUC is too stringent. Also, the settings for which the new measures are useful and their place in the evaluation of genetic risk prediction modeling need to be further investigated and existing models should be validated in various settings before recommendations about their use in practice can be done.

The value of modeling studies

Progress in genetic risk prediction is still going on. The advance of genomic technologies can facilitate the discovery of a large number of genetic variants and the rapid pace of discoveries implies that prediction models must often be updated. For scientific research this may mean that empirical studies may be needed to test the predictive power of the new genetic risk variants in samples other than those used for genetic discovery (75). Furthermore, the genetic variants may be population specific, which means that the predictive performance of a genetic test will vary with the ethnicity. Modeling studies can provide an estimation of the predictive ability of accumulating genetic information before empirical studies are performed. A number of simulation and analytical methods have already been developed (72, 87-90) and the AUC values estimated by these methods closely replicate the empirical values from prediction studies (91), indicating that modeling methods can be used to investigate the potential of genetic variants for risk prediction. Besides testing the values of SNPs for risk prediction the simulations may be used to test the role for risk prediction of gene-gene and gene-environment interactions or of other genetic risk factors. A recently published paper suggests that inclusion of interaction effects is unlikely to improve risk prediction for common diseases unless many such interactions with large effects are uncovered (92). Rare variants, on the other hand, are expected to display strong effects on disease risk and are envisioned to advance personalized interventions (35, 93-95). Since they are only becoming available their role for risk prediction in empirical studies remains yet to be investigated.

The rapid pace of genetic discoveries has also an impact on the DTC tests offered by commercial companies. The risk estimation offered to a consumer would change in time with the addition of new discoveries in the panel used to calculate the risk. Besides, the information incorporated in the risk prediction would generally still need to be tested in population cohorts. Modeling studies may prove a useful instrument for the assessment of the validity of DTC tests in the absence of population cohorts and can be used to compare the outputs of risk prediction tests between companies.

To summarize, simulations can be used to provide an estimate of the number of genetic variants and their performance parameters that would make them useful for risk prediction of common diseases, to study different modeling strategies and model performance measures, and to evaluate genetic DTC-tests.

Reporting in genetic risk prediction literature

Reporting in genetic risk prediction studies varies widely. The Genetic RIsk Prediction Studies (GRIPS) statement, a 25 items guideline recently developed by the Human Genome

Epidemiology Network (HuGENet) (96), aims to maximize the transparency, quality and completeness of reporting on research methodology and findings address genetic risk prediction studies. The GRIPS statement recommends that genetic risk prediction studies include a clear description of the key information specifically relevant for the interpretation and validation of genetic risk models. Examples include specification of the genetic variant that is considered as the risk variant, handling of genetic variants in the risk model construction, metrics used to assess the performance of the model. The extent to which these aspects are addressed in the published genetic risk prediction studies is unknown.

THIS THESIS

The focus of this thesis was twofold, first to evaluate traditional and recent measures of model performance in genetic risk models using simulation strategies, and second to assess the predictive performance of genetic risk prediction models for common diseases in epidemiological studies and the clinical utility of DTC-tests.

The first part of the thesis covers the methodology of genetic risk prediction. **Chapter 2** discusses the recently proposed reclassification measures and compares them to the traditional measures of model performance like the AUC and the risk distributions. The aim was to investigate if reclassification reflects improvement in risk prediction when addition of genetic variants does not increase the AUC. For this purpose, the relationship between AUC and various measures of reclassification was investigated using empirical data from the Rotterdam study (97) and simulated data. In the empirical data, Type 2 Diabetes was used as an example and a prediction model based on clinical factors was compared to the model with added genetic polymorphisms. **Chapter 3** continues with the exploration of sensitivity, specificity, positive and negative predictive value versus the AUC of genetic risk models along the range of threshold values that can be chosen to define high-risk groups. This chapter investigates how the relation between the prevalence of the risk group and the disease prevalence influences the model performance measures. Additionally a simulation based on published odds ratios and frequencies for six genetic polymorphisms predicting age-related macular degeneration risk was carried out. **Chapter 4** evaluates simulation models of genetic risk models based on information about SNPs from meta-analyses and GWAS data and compares the predictive performance of these models with the empirical performance of models from published studies. The aim of this study was to investigate if simulation methods can accurately estimate the AUC reported in empirical studies by using epidemiological data from published GWAS and meta-analyses. The ability to predict risk for common diseases by using rare variants is estimated in **Chapter 5**. To this end, a large dataset was simulated and risk models were constructed based either on

common variants alone or on common and rare variants. The impact of the frequency and effect of the rare variants on their predictive ability was assessed. These analyses were repeated across a wide range of predictive ability of the baseline risk model. Hypothetical data that replicated the empirical distributions of risk factors included in a clinical model for atrial fibrillation (AF) was also used. To these clinical factors genetic variants with varying ORs and frequencies were added and the improvement in model performance was assessed by ΔAUC, IDI and NRI. Finally, **Chapter 6** evaluates published genetic risk prediction studies based on the GRIPS. The aim was to map the weaknesses of genetic risk predictions papers, in order to bring awareness and improve the quality of reporting for future papers. A review of studies published from January until December 2010 was conducted.

In the second part of this thesis empirical examples were used to evaluate the current utility of genetic tests for risk prediction. **Chapter 7** analyses the DTC tests offered by commercial companies, with a focus on the repeated updates that these companies present to their consumers once new information about genetic risk variants becomes available. It investigates the extent to which updating of risk predictions leads to reclassification of individuals from below to above average disease risk or vice versa. The example of Type 2 Diabetes risk prediction was used. Analyses were performed using data from the Rotterdam Study (97). **Chapter 8** assesses the generalizability of genetic risk prediction models across nine European samples, using SNPs associated with increased risk of Type 2 Diabetes. Three prediction models were constructed in each sample: one based on clinical factors, one on genetic factors and one that combined the clinical and genetic risk factors. Each model was then tested in all other samples. The generalizability of the risk models was assessed by comparing calibration and discrimination measures between the nine European samples. **Chapter 9** compares the predictions between three DTC companies (23andMe, deCODEme and Navigenics) for a wide range of diseases. This is the first large scale study on the predictive ability of personal genome tests that examines whether discrepancies between companies are anecdotal or systematic. Predicted risks and the predictive ability of personal genome tests were assessed and compared. Predicted risks were calculated using the formulas of the companies and the predictive ability was quantified by AUC. **Chapter 10** assesses the methodological aspects of genetic risk prediction that could explain the differences in prediction performance between reported results in published articles, taking the example of Type 2 Diabetes. Genetic risk prediction studies were reviewed from a methodological perspective by focusing on factors in the choice of study design and population that may have impacted the observed predictive ability. **Chapter 11** continues the exploration of genetic risk prediction models for Type 2 Diabetes to provide an overview of the variants and model assessment methods used in scientific papers and the criteria used by commercial companies to select variants and calculate risks based on DTC tests.

REFERENCES

1. Hindorff LA MJEBI, Wise A, Junkins HA, Hall PN, Klemm AK, and Manolio TA. **A Catalog of Published Genome-Wide Association Studies.** Available at: www.genome.gov/gwastudies. Accessed 18-08-2012.
2. Ellinor PT, Lunetta KL, Albert CM, et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet*. 2012;44 (6):670-5.
3. Visscher PM, Brown MA, McCarthy MI, et al. **Five years of GWAS discovery.** *Am J Hum Genet*. 2012;**90** (1):7-24.
4. O'Donnell CJ, Nabel EG. **Genomics of cardiovascular disease.** *N Engl J Med*. 2011;**365** (22):2098-109.
5. Pawitan Y, Seng KC, Magnusson PK. **How many genetic variants remain to be discovered?** *PLoS One*. 2009;**4**(12):e7969.
6. Michot C, Le Goff C, Goldenberg A, et al. **Exome sequencing identifies PDE4D mutations as another cause of acrodysostosis.** *Am J Hum Genet*. 2012;**90**(4):740-5.
7. Zhang Z, Xia W, He J, et al. **Exome sequencing identifies SLCO2A1 mutations as a cause of primary hypertrophic osteoarthropathy.** *Am J Hum Genet*. 2012;**90**(1):125-32.
8. Chen WJ, Lin Y, Xiong ZQ, et al. **Exome sequencing identifies truncating mutations in PRRT2 that cause paroxysmal kinesigenic dyskinesia.** *Nat Genet*. 2011;**43**(12):1252-5.
9. Marti-Masso JF, Ruiz-Martinez J, Makarov V, et al. **Exome sequencing identifies GCDH (glutaryl-CoA dehydrogenase) mutations as a cause of a progressive form of early-onset generalized dystonia.** *Hum Genet*. 2012;**131**(3):435-42.
10. Sloan JL, Johnston JJ, Manoli I, et al. **Exome sequencing identifies ACSF3 as a cause of combined malonic and methylmalonic aciduria.** *Nat Genet*. 2011;**43**(9):883-6.
11. Regalado ES, Guo DC, Villamizar C, et al. **Exome sequencing identifies SMAD3 mutations as a cause of familial thoracic aortic aneurysm and dissection with intracranial and other arterial aneurysms.** *Circ Res*. 2011;**109**(6):680-6.
12. Dickinson RE, Griffin H, Bigley V, et al. **Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency.** *Blood*. 2011;**118**(10):2656-8.
13. Comino-Mendez I, Gracia-Aznarez FJ, Schiavi F, et al. **Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma.** *Nat Genet*. 2011;**43**(7):663-7.
14. O'Sullivan J, Bitu CC, Daly SB, et al. **Whole-Exome sequencing identifies FAM20A mutations as a cause of amelogenesis imperfecta and gingival hyperplasia syndrome.** *Am J Hum Genet*. 2011;**88** (5):616-20.
15. Puente XS, Quesada V, Osorio FG, et al. **Exome sequencing and functional analysis identifies BANF1 mutation as the cause of a hereditary progeroid syndrome.** *Am J Hum Genet*. 2011;**88**(5):650-6.
16. Glazov EA, Zankl A, Donskoi M, et al. **Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia.** *PLoS Genet*. 2011;**7**(3):e1002027.
17. Haack TB, Danhauser K, Haberberger B, et al. **Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency.** *Nat Genet*. 2010;**42**(12):1131-4.
18. Ng SB, Bigham AW, Buckingham KJ, et al. **Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.** *Nat Genet*. 2010;**42**(9):790-3.
19. Ng SB, Buckingham KJ, Lee C, et al. **Exome sequencing identifies the cause of a mendelian disorder.** *Nat Genet*. 2010;**42**(1):30-5.

20. Need AC, McEvoy JP, Gennarelli M, et al. **Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia.** *Am J Hum Genet*. 2012;**91**(2):303-12.

21. Xu B, Roos JL, Dexheimer P, et al. **Exome sequencing supports a de novo mutational paradigm for schizophrenia.** *Nat Genet*. 2011;**43**(9):864-8.

22. Yi X, Liang Y, Huerta-Sanchez E, et al. **Sequencing of 50 human exomes reveals adaptation to high altitude.** *Science*. 2010;**329**(5987):75-8.

23. Salari K, Watkins H, Ashley EA. **Personalized medicine: hope or hype?** *Eur Heart J*. 2012.

24. Collins FS, Green ED, Guttmacher AE, et al. **A vision for the future of genomics research.** *Nature*. 2003;**422**(6934):835-47.

25. Brand A, Brand H, Schulte in den Baumen T. **The impact of genetics and genomics on public health.** *Eur J Hum Genet*. 2008;**16**(1):5-13.

26. Collins FS, McKusick VA. **Implications of the Human Genome Project for medical science.** *JAMA*. 2001;**285**(5):540-4.

27. Hartman M, Loy EY, Ku CS, et al. **Molecular epidemiology and its current clinical use in cancer management.** *Lancet Oncol*. 2010;**11**(4):383-90.

28. Guttmacher AE, Collins FS. **Realizing the promise of genomics in biomedical research**. *JAMA*. 2005;**294** (11):1399-402.

29. Jostins L, Barrett JC. **Genetic risk prediction in complex disease.** *Hum Mol Genet*. 2011;**20**(R2):R182-8.

30. Swen JJ, Huizinga TW, Gelderblom H, et al. **Translating pharmacogenomics: challenges on the road to the clinic.** *PLoS Med*. 2007;**4**(8):e209.

31. Reiner A, Dragatsis I, Dietrich P. **Genetics and neuropathology of Huntington's disease.** *Int Rev Neurobiol*. 2011;**98**:325-72.

32. Sherrington R, Froelich S, Sorbi S, et al. **Alzheimer's disease associated with mutations in presenilin 2 is rare and variably penetrant.** *Hum Mol Genet*. 1996;**5**(7):985-8.

33. Visscher PM, Hill WG, Wray NR. **Heritability in the genomics era--concepts and misconceptions.** *Nat Rev Genet*. 2008;**9**(4):255-66.

34. So HC, Gui AH, Cherny SS, et al. **Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases.** *Genet Epidemiol*. 2011;**35**(5):310-7.

35. Visscher PM, Goddard ME, Derks EM, et al. **Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses.** *Mol Psychiatry*. 2011.

36. Purcell SM, Wray NR, Stone JL, et al. **Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.** *Nature*. 2009;**460**(7256):748-52.

37. Manolio TA, Collins FS, Cox NJ, et al. **Finding the missing heritability of complex diseases.** *Nature*. 2009;**461**(7265):747-53.

38. Makowsky R, Pajewski NM, Klimentidis YC, et al. **Beyond missing heritability: prediction of complex traits.** *PLoS Genet*. 2011;**7**(4):e1002051.

39. Schapira AH. **Mitochondrial diseases.** *Lancet*. 2012;**379**(9828):1825-34.

40. Chin L, Gray JW. **Translating insights from the cancer genome into clinical practice.** *Nature*. 2008;**452** (7187):553-63.

41. Moore JH, Williams SM. **Epistasis and its implications for personal genetics.** *Am J Hum Genet*. 2009;**85** (3):309-20.

42. Galvan A, Ioannidis JP, Dragani TA. **Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer.** *Trends Genet*. 2010;**26**(3):132-41.

43. **Moving AHEAD with an international human epigenome project.** *Nature*. 2008;**454**(7205):711-5.

44. Goate A, Chartier-Harlin MC, Mullan M, et al. **Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease.** *Nature*. 1991;**349**(6311):704-6.

45. Sherrington R, Rogaev EI, Liang Y, et al. **Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease.** *Nature*. 1995;**375**(6534):754-60.

46. Levy-Lahad E, Wasco W, Poorkaj P, et al. **Candidate gene for the chromosome 1 familial Alzheimer's disease locus.** *Science*. 1995;**269**(5226):973-7.

47. Rogaev EI, Sherrington R, Rogaeva EA, et al. **Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene.** *Nature*. 1995;**376**(6543):775-8.

48. Vineis P, Schulte P, McMichael AJ. **Misconceptions about the use of genetic tests in populations.** *Lancet*. 2001;**357**(9257):709-12.

49. Holtzman NA, Marteau TM. **Will genetics revolutionize medicine?** *N Engl J Med*. 2000;**343**(2):141-4.

50. Janssens AC, Gwinn M, Valdez R, et al. **Predictive genetic testing for type 2 diabetes.** *BMJ*. 2006;**333** (7567):509-10.

51. Grassmann F, Fritsche LG, Keilhauer CN, et al. **Modelling the genetic risk in age-related macular degeneration.** *PLoS One*. 2012;**7**(5):e37979.

52. Uusitupa MI, Stancakova A, Peltonen M, et al. **Impact of positive family history and genetic risk variants on the incidence of diabetes: the Finnish Diabetes Prevention Study.** *Diabetes Care*. 2011;**34**(2):418-23.

53. Aly M, Wiklund F, Xu J, et al. **Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm-1 cohort study.** *Eur Urol*. 2011;**60**(1):21-8.

54. Ruchat SM, Vohl MC, Weisnagel SJ, et al. **Combining genetic markers and clinical risk factors improves the risk assessment of impaired glucose metabolism.** *Ann Med*. 2010;**42**(3):196-206.

55. van der Net JB, Janssens AC, Defesche JC, et al. **Usefulness of genetic polymorphisms and conventional risk factors to predict coronary heart disease in patients with familial hypercholesterolemia.** *Am J Cardiol*. 2009;**103**(3):375-80.

56. Haddow JE, Palomaki GE. ACCE: **A Model Process for Evaluating Data on Emerging Genetic Tests. In: Khoury M LJ, Burke W, ed. Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease.** : Oxford University Press, 2003:217-33.

57. Hlatky MA. **Framework for evaluating novel risk markers.** *Ann Intern Med*. 2012;**156**(6):468-9.

58. Hlatky MA, Greenland P, Arnett DK, et al. **Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association.** *Circulation*. 2009;**119**(17):2408-16.

59. Akaike H. **A new look at the statistical model identification.** *IEEE Trans Automat Contr*. 1974;AC-**19**:716-23.

60. Hosmer DW, Lemeshow S. **Applied Logistic Regression.** New York: Wiley, 2000.

61. Hanley JA, McNeil BJ. **The meaning and use of the area under a receiver operating characteristic (ROC) curve.** *Radiology*. 1982;**143**(1):29-36.

62. Kader AK, Sun J, Reck BH, et al. **Potential Impact of Adding Genetic Markers to Clinical Parameters in Predicting Prostate Biopsy Outcomes in Men Following an Initial Negative Biopsy: Findings from the REDUCE Trial.** *Eur Urol*. 2012.

63. Kathiresan S, Melander O, Anevski D, et al. **Polymorphisms associated with cholesterol and risk of cardiovascular events.** *N Engl J Med*. 2008;**358**(12):1240-9.

64. Lango H, Consortium UKTDG, Palmer CN, et al. **Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk.** *Diabetes*. 2008;**57**(11):3129-35.

65. Lyssenko V, Almgren P, Anevski D, et al. **Genetic prediction of future type 2 diabetes.** *PLoS Med*. 2005;2 (12):e345.

66. Lyssenko V, Jonsson A, Almgren P, et al. **Clinical risk factors, DNA variants, and the development of type 2 diabetes.** *N Engl J Med*. 2008;**359**(21):2220-32.

67. Meigs JB, Shrader P, Sullivan LM, et al. **Genotype score in addition to common risk factors for prediction of type 2 diabetes.** *N Engl J Med*. 2008;**359**(21):2208-19.

68. Morrison AC, Bare LA, Chambless LE, et al. **Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study.** *Am J Epidemiol*. 2007;**166**(1):28-35.

69. Seddon JM, Reynolds R, Maller J, et al. **Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables.** *Invest Ophthalmol Vis Sci*. 2009;**50**(5):2044-53.

70. van Hoek M, Dehghan A, Witteman JC, et al. **Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study.** *Diabetes*. 2008;**57**(11):3122-8.

71. Vaxillaire M, Veslot J, Dina C, et al. **Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study.** *Diabetes*. 2008;**57**(1):244-54.

72. Janssens AC, Aulchenko YS, Elefante S, et al. **Predictive testing for complex diseases using multiple genes: fact or fiction?** *Genet Med*. 2006;**8**(7):395-400.

73. Wray NR, Yang J, Goddard ME, et al. **The genetic interpretation of area under the ROC curve in genomic profiling.** *PLoS Genet*. 2010;**6**(2):e1000864.

74. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., et al. **Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond.** *Stat Med*. 2008;**27**(2):157-72; discussion 207-12.

75. Altman DG, Vergouwe Y, Royston P, et al. **Prognosis and prognostic research: validating a prognostic model.** *Bmj*. 2009;**338**:b605.

76. Harrell FE, Jr., Lee KL, Mark DB. **Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.** *Stat Med*. 1996;**15**(4):361-87.

77. Paynter NP, Chasman DI, Pare G, et al. **Association between a literature-based genetic risk score and cardiovascular events in women.** *Jama*. **303**(7):631-7.

78. Janes H, Pepe MS, Gu W. **Assessing the value of risk predictions by using risk stratification tables.** *Ann Intern Med*. 2008;**149**(10):751-60.

79. 23andMe. Available at: http://www.23andme.com/. Accessed November 20, 2008.

80. deCODEme. Available at: http://www.decodeme.com/. Accessed November 20, 2008.

81. Knome Web. Available at: http://www.knome.com/. Accessed November 20, 2008.

82. Kuehn BM. **Inconsistent results, inaccurate claims plague direct-to-consumer gene tests.** *Jama*. 2010;**304**(12):1313-5.

83. Swan M. **Multigenic condition risk assessment in direct-to-consumer genomic services.** *Genet Med*. 2010;**12**(5):279-88.

84. Pepe MS, Feng Z, Huang Y, et al. **Integrating the predictiveness of a marker with its performance as a classifier.** *Am J Epidemiol*. 2008;**167**(3):362-8.

85. Vaxillaire M, Veslot J, Dina C, et al. **Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study.** *Diabetes*. 2008;**57**(1):244-54.

86. Cook NR. **Use and misuse of the receiver operating characteristic curve in risk prediction.** *Circulation*. 2007;**115**(7):928-35.

87. Pepe MS, Gu JW, Morris DE. **The potential of genes and other markers to inform about risk.** *Cancer Epidemiol Biomarkers Prev*. 2010;**19**(3):655-65.

88. Lu Q, Elston RC. **Using the optimal receiver operating characteristic curve to design a predictive genetic test, exemplified with type 2 diabetes.** *Am J Hum Genet*. 2008;**82**(3):641-51.

89. Moonesinghe R, Liu T, Khoury MJ. **Evaluation of the discriminative accuracy of genomic profiling in the prediction of common complex diseases.** *Eur J Hum Genet*. 2010;**18**(4):485-9.

90. Gail MH. **Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk.** *J Natl Cancer Inst*. 2008;**100**(14):1037-41.

91. Kundu S, Karssen LC, Janssens AC. **Analytical and simulation methods for estimating the potential predictive ability of genetic profiling: a comparison of methods and results.** *Eur J Hum Genet*. 2012.

92. Aschard H, Chen J, Cornelis MC, et al. **Inclusion of Gene-Gene and Gene-Environment Interactions Unlikely to Dramatically Improve Risk Prediction for Complex Diseases.** *Am J Hum Genet*. 2012.

93. Bodmer W, Bonilla C. **Common and rare variants in multifactorial susceptibility to common diseases.** *Nat Genet*. 2008;**40**(6):695-701.

94. Gorlov IP, Gorlova OY, Frazier ML, et al. **Evolutionary evidence of the effect of rare variants on disease etiology.** *Clin Genet*. 2011;**79**(3):199-206.

95. Wang K, Dickson SP, Stolle CA, et al. **Interpretation of association signals and identification of causal variants from genome-wide association studies.** *Am J Hum Genet*. 2010;**86**(5):730-42.

96. Janssens AC, Ioannidis JP, Bedrosian S, et al. **Strengthening the reporting of genetic risk prediction studies (GRIPS): explanation and elaboration.** *Eur J Hum Genet*. 2011.

97. Hofman A, Breteler MM, van Duijn CM, et al. **The Rotterdam Study: objectives and design update.** *Eur J Epidemiol*. 2007;**22**(11):819-29.

PART I
Methodology

Chapter 2

Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve

Raluca Mihaescu, Moniek van Zitteren, Mandy van Hoek, Eric J.G. Sijbrands, André G. Uitterlinden, Jacqueline C.M. Witteman, Albert Hofman, M.G. Myriam Hunink, Cornelia M. van Duijn, A. Cecile J.W. Janssens

American Journal of Epidemiology. 2010;172(3):353-61

ABSTRACT

Reclassification is observed even when there is no or minimal improvement in the area under the ROC curve (AUC) and it is unclear whether this indicates improved clinical utility. We investigated total reclassification, net reclassification improvement (NRI) and integrated discrimination improvement (IDI) for different ΔAUC using empirical and simulated data. Empirical analyses compared prediction of type 2 diabetes risk based on age, sex and body mass index with prediction updated with 18 established genetic risk factors. Simulated data were used to investigate measures of reclassification against ΔAUC of 0.005, 0.05 and 0.10. Total reclassification and NRI were calculated for all possible cutoff values. AUC of type 2 diabetes risk prediction improved from 0.63 to 0.66 when 18 polymorphisms were added, whereas total reclassification ranged from 0% to 22.5%, depending on the cutoff value chosen. In the simulation study, total reclassification, NRI and IDI increased with higher ΔAUC. When ΔAUC was low (0.005), NRI values were close to zero, IDI was 0.08% ($P > 0.05$), but total reclassification ranged from 0 to 6.7%. Reclassification increases with increasing AUC, but predominantly varies with the chosen cutoff values. Reclassification observed in the absence of AUC increase is unlikely to improve clinical utility.

There is increasing interest in investigating the improvement in risk predictions by adding novel genetic variants to traditional risk prediction models for common diseases (1-4). Improvement of disease risk prediction is commonly assessed by comparing the area under the receiver operating characteristic curve (AUC, or the *c* statistic) of the prediction models without and with these added variables of interest. AUC assesses the extent to which predicted risks discriminate between individuals who will develop the disease and those who will not (5). Many researchers criticize AUC because this measure lacks an apparent intuitive interpretation, and because even statistically significant new biomarkers yield minimal improvement in AUC (1-3). For these reasons, AUC is considered insensitive and not suitable for assessing improvement in prediction, and the use of other measures is encouraged (6, 7).

Currently, much attention is directed towards measures of reclassification (8). Reclassification assesses the extent to which improvements of risk prediction models influence medical decisions. When prediction models are used for making treatment decisions, predicted risks are categorized using clinically relevant risk cutoff thresholds. Improvement of a prediction model has impact on medical decisions when individuals end up in another category under a new model compared to the initial model. The percentage of individuals who change risk categories is referred to as the percentage of reclassification.

Several empirical studies have shown substantial percentages of individuals reclassify when novel risk indicators are added to prediction models (9, 10). In some studies, reclassification was observed even when there was no or minimal improvement in AUC (6, 11). Yet, more recently it has been acknowledged that not every reclassification may be correct, which has led to the development of alternative measures such as percentage of correct reclassification, net reclassification improvement (NRI) and integrated discrimination improvement (IDI) (12-14). These measures differ predominantly in their definition of correct reclassification. Reclassification can be considered correct when the observed risk in a reclassified group falls within or is closer to the new risk category (12, 13), but also when individuals who will develop disease change to a higher risk category and those who will not, change to a lower risk category (14). Another difference between the measures is that IDI is a global measure of correct reclassification over all possible cutoff values, and in that sense is comparable to the AUC, whereas the other measures assess correct reclassification for specific cutoff values.

Measures of reclassification are increasingly used in studies investigating the added value of genetic factors beyond traditional risk factors (**Table 1**) (2, 12, 15, 16). Remarkably, these studies tend to put stronger emphasis on reclassification than on AUC, particularly when reclassification is observed in the absence of improved AUC (2, 15). It is not clear whether reclassification reflects improvement in risk prediction when AUC does not change. In this study, we investigated the

relationship between AUC and various measures of reclassification for different ΔAUC and for all possible risk cutoff levels, using empirical data from the Rotterdam study and simulated data. In the Rotterdam study, we compared risk predictions for type 2 diabetes based on age, sex and body mass index (BMI), with predictions based on these factors plus 18 genetic polymorphisms. In the simulated data we additionally investigated measures of reclassification against different improvements in AUC.

MATERIALS AND METHODS

Empirical data

The design and data collection of the Rotterdam Study have been described elsewhere (17). In short, the Rotterdam Study is a prospective, population-based, cohort study among 7,983 inhabitants of a Rotterdam suburb, designed to investigate determinants of chronic diseases. Participants were aged 55 years and older. Baseline examinations took place from 1990 until 1993, and follow-up examinations were performed in 1993-1994, 1997-1999 and 2002-2004. Between these exams, continuous surveillance on major disease outcomes was conducted. Details on the genotyped variants and diagnostic criteria for diabetes have been published elsewhere (1, 18, 19). Patients with a recorded diagnosis of type 1 diabetes were excluded from the present analyses. BMI was calculated as weight in kilograms divided by height in meters squared. Age and BMI were obtained from the baseline assessment. The medical ethics committee of the Erasmus Medical Center approved the study protocol and all participants gave their written informed consent.

Simulated data

For the construction of simulated data sets, we adopted a modeling procedure described in detail elsewhere (20). In short, the procedure creates a dataset in such a way that the frequencies and odds ratios (OR) of the genotypes and the disease risk match specified values. For simplicity, we assumed each individual polymorphism had only two genotypes, one of which was associated with an increased risk of disease and the other being the referent group at baseline risk. Genetic variants were modeled as independent, while the joint risk of multiple genotypes followed a multiplicative risk model without any statistical interactions. For a population of 10,000 individuals we simulated two prediction models. The first model was based on 20 polymorphisms, all with arbitrary OR of 1.2 and genotype frequency of 20%, which gave an AUC of 0.61. In the second model we added genetic markers to the first model so the increment in AUC (ΔAUC) was 0.005, 0.05 and 0.10, respectively. To achieve this, we added 1, 25 and 60

markers with the same ORs and genotype frequencies as in the first model. The disease risk was arbitrarily set at 10%.

Statistical analyses

In both the empirical and simulated data, predicted risks were obtained using logistic regression analyses. In the empirical study, we constructed a prediction model for type 2 diabetes (prevalent and incident cases) based on age, sex and BMI, and one based on age, sex and BMI plus 18 genetic polymorphisms, with all polymorphisms entered as categorical variables, allowing effect sizes to differ between heterozygous and homozygous carriers of the risk alleles. The population of the present analyses and the included polymorphisms have been described elsewhere (1). In the simulation study, we constructed risk prediction models based on 20, 21, 45 and 80 polymorphisms.

Cutoff values were varied between 1% and 99% with increments of 1%. At each possible cutoff value predicted risks were classified into two risk categories. One category included individuals with a risk higher or equal to the cutoff value and the other included those with a risk lower than the cutoff value. AUC ranges from 0.5 (total lack of discrimination) to 1.0 (perfect discrimination). The percentage of total reclassification was measured as the percentage of individuals that change risk category out of the total population. NRI was calculated as the sum of differences in proportion of individuals moving up minus the proportion moving down for cases, and the proportion of individuals moving down minus the proportion moving up for non-cases (14). The components of NRI indicate the net benefit of reclassification improvement in cases and non-cases. Positive and negative values represent the net percentage of individuals with improved or worse classification, respectively. Overall, improvement in reclassification is indicated by a NRI significantly greater than 0 (14). IDI was calculated as the difference in mean predicted probabilities between cases and non-cases between the two models (14). IDI can be seen as an improvement in average sensitivity weighted by average one minus specificity (14). Formulas for the reclassification measures used in this study are listed in the Web Appendix. Because both the cutoff levels and most reclassification measures are presented as percentages, we write zero decimals when percentages refer to the cutoff levels and one decimal when they concern the observed percentages of reclassification.

Table 1. AUC and reclassification measures reported for genetic polymorphisms added to type 2 diabetes and cardiovascular disease risk prediction models.

Study	Disease risk, %[a]	Cutoff values for risk categories, %	Clinical risk prediction model[a]	Number of added poly-morphisms[b]	AUC for the initial model[b]	95% CI
Type 2 diabetes						
Meigs et al. (16)	10.47	0 to <2 2 to <8 ≥8	Sex	18	0.534	0.502, 0.565
			Sex and family history	18	0.595	0.560, 0.630
			Clinical risk factors[c]	18	0.900	0.880, 0.919
Lyssenko et al. (2)	12.8 (Malmö study)	0 to ≤10 >10 to ≤20 >20	Clinical risk factors[e]	11	0.743	
	5 (Botnia study)		Clinical risk factors [e]	11	0.786	
Rotterdam study	20.0	All possible thresholds[f]	Age, sex and BMI	18	0.631	0.612, 0.650
Cardiovascular disease						
Paynter et al. (12)	3.2	0 to <5 5 to <10 10 to 20 ≥20	ATP III[h]	1	0.803	0.766, 0.840
			Reynolds Risk Score[i]	1	0.807	0.770, 0.844
Kathiresan et al. (15)	5.6	0 to ≤10 >10 to ≤20 >20	ATP III[h]	9	0.800	

Abbreviations: ATP III, Third Report of the National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Cholesterol in Adults; AUC, area under the receiver operating characteristic curve; CI, confidence interval; IDI, integrated discrimination improvement; NR, not reported; NRI, net reclassification improvement.
[a] In the Rotterdam Study, calculated as the number of prevalent and incident cases divided by the sample size. [b] Initial risk prediction model included only clinical factors, whereas the updated risk prediction model included clinical factors and one or more genetic susceptibility variants. [c] Calculated from reclassification tables available from the cited papers. [d] The model was adjusted for age, sex, family history, body mass index, fasting glucose level, systolic blood pressure, high density lipoprotein cholesterol level, and triglyceride level. [e] The initial model included age, sex, family history of diabetes, body mass index, blood pressure, high density lipoprotein cholesterol (available in Botnia only), triglycerides, waist circumference (available in Botnia only), and fasting plasma glucose. [f] Cutoff values were used to define 2 risk categories; one included individuals with a risk higher than or equal to the cutoff value, and the other included those with a risk lower than the cutoff value. [g] Range of reclassification measures and corresponding P values. [h] The ATP III covariates include the natural logarithm of age, systolic blood pressure, total and high density lipoprotein cholesterol, smoking status, antihypertensive medication use, and history of diabetes. [i] Reynolds Risk Score covariates include age, smoking status, family history of myocardial infarction, hemoglobin A1c levels (diabetic patients only), and the natural logarithm of systolic blood pressure and total and high density lipoprotein cholesterol.

	AUC for the updated model[b]	95% CI	ΔAUC	P value	Reclassification, %	NRI, %	P value	IDI, %	P value
	0.581	0.546, 0.617	0.047	0.01	2.56[d]	4.10	0.004		
	0.615	0.579, 0.652	0.020	0.11	9.54[d]	2.60	0.22		
	0.901	0.881, 0.920	0.001	0.49	6.28[d]	2.13	0.17		
	0.753		0.010	0.0001	15.59[d]	4.5[d]	$2.5x10^{-5}$[d]	NR	$3.7x10^{-14}$
	0.801		0.015	0.23	9.44[d]	8.79[d]	0.13[d]	NR	0.001
	0.661	0.643, 0.679	0.030	$2x10^{-5}$	0 to 22.5[g]	−0.03 to 9.2[g]	0.79 $6.8x10^{-14}$[g]	2.04	$6.5x10^{-20}$
	0.805	0.768, 0.842	0.002		2.7	2.7	0.02	0.001	0.11
	0.809	0.772, 0.846	0.002		2.6	−0.2	0.59	0.0	0.18
	0.800		0.00		4.39[d]	6[d]	0.01[d]	NR	0.02

In the simulation study, reclassification measures and AUCs are presented as averages of 100 simulations. The obtained confidence intervals for AUCs and IDI were extremely small (varying the third digit after the decimal point) and therefore only confidence intervals for total reclassification and NRI were presented in this paper. Differences in AUCs were tested with Analyze-it software version 2.20 (www.analyze-it.com), which uses the method of DeLong, DeLong and Clarke-Pearson for ROC curve analyses (21, 22). Logistic regression analyses in the empirical study were performed using the SPSS software version 15.0.1 (SPSS Inc., Chicago, Illinois). All other analyses were performed using the R programming language version 2.8.0 (23). AUC was obtained as the c-statistic by the R function *somers2*, which is available in the Hmisc library of the R software (24).

RESULTS

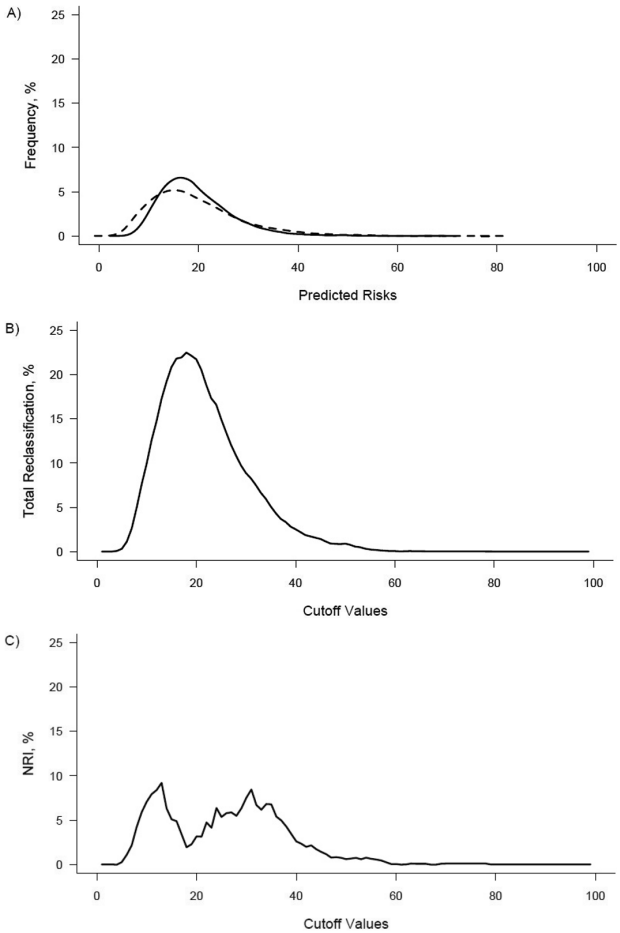
Prediction of type 2 diabetes

The average risk of type 2 diabetes in the population was 20%. Updating prediction based on age, sex and BMI with 18 polymorphisms, redistributed predicted risks towards more extreme values (**Figure 1A**). This redistribution improved the discriminative accuracy, which was also reflected in the AUC values. The AUC was 0.63 (95% confidence interval (CI): 0.61, 0.65) for prediction based on age, sex and BMI, and 0.66 (95% CI: 0.64, 0.68) for prediction based on age, sex, BMI and 18 polymorphisms ($P < 0.001$). The percentages of total reclassification varied with the cutoff values (**Figure 1B**). The maximum percentage of total reclassification was 22.5% when the cutoff value was 18%. The NRI values ranged from -0.03% ($P=0.79$) to 9.2% ($P<0.001$) and 8.5 ($P<0.001$), when the cutoff values were 4%, 13% and 31% (**Figure 1C**). All reclassification measures were zero for cutoff values lower than 4% and higher than 78%. The IDI was 2.04% ($P<0.001$).

Simulation study

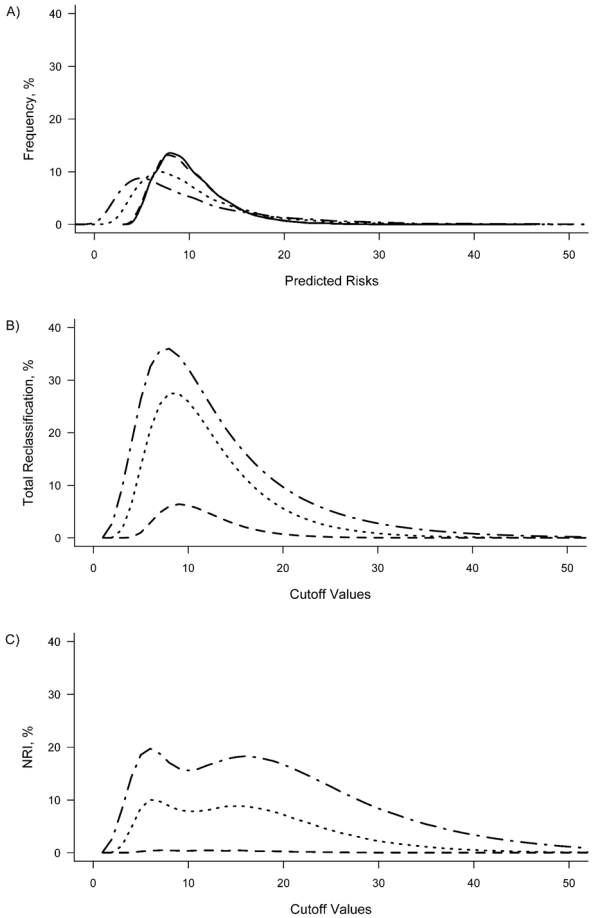
Figure 2 shows risk distributions diverged with increasing AUC. Evidently, when ΔAUC was 0.005 the two risk distributions overlapped, but when ΔAUC increased to 0.05 and 0.10, the risk distribution of the updated models covered more extreme values (**Figure 2A**). The percentages of total reclassification varied with the cutoff values and were higher with increasing ΔAUC at each cutoff value (**Figure 2B**). Total reclassification ranged from 0.0 to 6.7% (95% CI: 6.2, 7.2 at cutoff value of 9%) when ΔAUC was 0.005, to 27.2% (95% CI: 26.9, 27.6 at cutoff of 8%) when ΔAUC was 0.05, and to 35.6% (95% CI: 35.3, 36.0 at cutoff value of 8%) when ΔAUC was 0.10.

Figure 1. Risk distribution and reclassification measures for prediction of type 2 diabetes risk in the Rotterdam study, performed from 1990-1993 to 2002-2004.



A) Continuous line: distribution of predicted risks from the model based on age, sex, and body mass index; dashed line: distribution of predicted risks from the model based on age, sex, body mass index, and 18 polymorphisms. B) Percentage of total reclassification per cutoff value for risk stratification. C) Net reclassification improvement (NRI) per cutoff value. Cutoff values were used to define 2 risk categories; one included individuals with a risk higher than or equal to the cutoff value, and the other included those with a risk lower than the cutoff value. NRI was calculated as the sum of differences in the proportion of individuals moving up minus the proportion moving down for cases, and the proportion of individuals moving down minus the proportion moving up for noncases.

Figure 2. Risk distribution and reclassification measures for prediction models in the simulation study.



A) Distribution of predicted risks for the model based on 20 polymorphisms (continuous line) and for the updated models when change in the area under the receiver operating characteristic curve (ΔAUC) is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed-and-dotted line). B) Percentage of total reclassification per cutoff value for risk stratification when DAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed-and-dotted line). C) Net reclassification improvement (NRI) per cutoff value for risk stratification when ΔAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed-and-dotted line). Each polymorphism has an odds ratio of 1.2 and a frequency of 20%. Disease risk is 10%, and sample size is 10,000. Percentage of total reclassification and NRI are presented as mean values obtained from 100 simulations. Cutoff values are used to define 2 risk categories; one included individuals with a risk higher than or equal to the cutoff value, and the other included those with a risk lower than the cutoff value. NRI was calculated as the sum of differences in the proportion of individuals moving up minus the proportion moving down for cases, and the proportion of individuals moving down minus the proportion moving up for noncases.

NRI also varied with the cutoff value and ΔAUC (**Figure 2C**). When ΔAUC was 0.005, NRI ranged from -0.7% (95% CI: -0.8,-0.65; P>0.05) to 1.5% (95% CI: 1.4, 1.7; P>0.05); when ΔAUC was 0.05, NRI ranged from -0.015% (95% CI: -0.02, -0.01; P>0.05) to 10.5% (95% CI: 10.2, 10.8; P<0.001); and when ΔAUC was 0.10, NRI ranged from 0 (95% CI: -0.002, 0.002; P>0.05) to 19.9% (95% CI: 19.5, 20.2; P<0.001), depending on the chosen cutoff value. The individual components of NRI (i.e. net benefit in cases and in non-cases) varied with the cutoff value (**Figure 3**). When ΔAUC was 0.10, the net benefit in cases was negative below and positive above the cutoff value of 10% [range -9.4% (95% CI: -9.7, -9.2; P<0.001) to 25.8% (95% CI: 25.2, 26.3; P<0.001)], whereas the reverse was observed among non-cases [range -7.6% (95% CI: -7.8, -7.5; P<0.001) to 28.8% (95% CI: 28.3, 29.3; P<0.001)]. IDI was 0.08% (P>0.05), 2% (P<0.001) and 5% (P<0.001) when ΔAUC was 0.005, 0.05 and 0.10, respectively. The distribution of reclassification measures did not change when the OR and genotype frequencies of the added variants were varied (refer to the **Web Figure 1 and 2**). When the AUC of the original model was increased from 0.61 to 0.75, the distribution of reclassification measures changed, but the percentage of reclassification still varied with the cutoff threshold chosen (refer to the **Web Figure 3 and 4**).

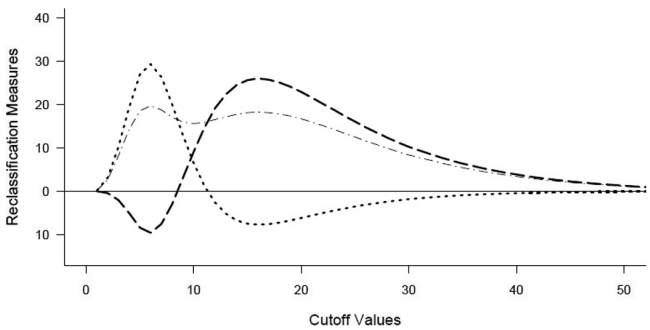
DISCUSSION

In this paper, we demonstrated the amount of reclassification depends on the increase in AUC achieved by updating a prediction model but even more on the chosen cutoff value. In the empirical study, prediction of type 2 diabetes risk based on age, sex and BMI improved by Continuous line shows the distribution of predicted risks from the model based on the age, sex and body mass index. Dashed line shows the distribution of predicted risks from the model based on age, sex, body mass index and 18 polymorphisms (**1A**). Continuous line shows the percentages of total reclassification per cutoff value for risk stratification (**1B**). Continuous line shows the net reclassification improvement (NRI) per cutoff value (**1C**). Cutoff values are used to define two risk categories, one category included individuals with a risk higher or equal to the cutoff value and the other included those with a risk lower than the cutoff value. NRI was calculated as the sum of differences in proportion of individuals moving up minus the proportion moving down for cases, and the proportion of individuals moving down minus the proportion moving up for non-cases.

Distribution of predicted risks for the model based on 20 polymorphisms (continuous line) and for the updated models when ΔAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed and dotted line) (**2A**); Percentages of total reclassification per cutoff value for risk stratification when ΔAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed and dotted line) (**2B**); Net reclassification improvement (NRI) per cutoff value for risk stratification when

ΔAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed and dotted line) (**2C**). Each polymorphism has an odds ratio of 1.2 and frequency of 20%. Disease risk is 10% and sample size is 10,000. Percentage of total reclassification and NRI are presented as mean values obtained from 100 simulations. Cutoff values are used to define two risk categories, one category included individuals with a risk higher or equal to the cutoff value and the other included those with a risk lower than the cutoff value. NRI was calculated as the sum of differences in proportion of individuals moving up minus the proportion moving down for cases, and the proportion of individuals moving down minus the proportion moving up for non-cases.

Figure 3. Net reclassification improvement and its components.



Dashed-and-dotted line: net reclassification improvement per cutoff value for risk stratification when change in the area under the receiver operating characteristic curve is 0.10; dotted line: percentage of correct (noncases moving down) minus incorrect (noncases moving up) reclassification of noncases for each cutoff value; dashed line: percentage of correct (cases moving up) minus incorrect (cases moving down) reclassification of cases for each cutoff value. Each value was obtained from the mean of 100 simulations. Cutoff values were used to define 2 risk categories; one included individuals with a risk higher than or equal to the cutoff value, and the other included those with a risk lower than the cutoff value.

The dashed and dotted line shows the net reclassification improvement per cutoff value for risk stratification when ΔAUC is 0.10. The dotted line shows the percentages of correct (non-cases moving down) minus incorrect (non-cases moving up) reclassification in non-cases for each cutoff value. The dashed line shows the percentages of correct (cases moving up) minus incorrect (cases moving down) reclassification in cases for each cutoff value. Each value is obtained from the mean of 100 simulations. Cutoff values are used to define two risk categories, one category included individuals with a risk higher or equal to the cutoff value and the other included those with a risk lower than the cutoff value adding 18 polymorphisms as indicated

by the improvement in AUC (0.63 to 0.66). Yet, the percentage of total reclassification ranged from 0% to 22.5% depending on the cutoff threshold, with a maximum at cutoff values around the median of the risk distribution. The simulation study showed all reclassification measures increased with increasing ΔAUC, at every cutoff value.

Our comparative analyses allowed several inferences to be drawn about the relationship between reclassification measures and AUC. First, the cutoff value had a substantial impact on the amount of reclassification. For example, in our simulation study, for an increase in AUC of 0.005, 6.7% of individuals changed their predicted risk categories when the cutoff was 9%, representing the highest amount of reclassification, but less than 1.0% of subjects reclassified when the cutoff was 20%. Similarly, when the increase in AUC of the updated model was 0.10, 35.0% and 10.0% of individuals were reclassified. In general, more reclassification was observed for cutoff values around the median of the risk distribution than for cutoff values in the tails. Because the amount of reclassification varied with the cutoff value chosen, calculation of reclassification is only meaningful when the cutoff scores defining risk categories with different medical implications, e.g. lead to different preventive or therapeutic choices. Only under the condition that cutoff thresholds are clinically meaningful would reclassification become a measure of clinical utility.

Second, when the increase in AUC was minimal (e.g., ΔAUC=0.005 or lower), reclassification was still observed at some cutoff values when the percentage of total reclassification was calculated, but not when NRI or IDI were calculated. When AUC does not change, the initial and the updated prediction model predict equally good or bad. The percentage of total reclassification indicates that at the individual level some cases and non-cases may move up and others may move down, yet IDI and NRI show that there is no net benefit of this reclassification at the population level. Thus, when reclassification is observed in the absence of any AUC improvement, reclassification means an updated model simply makes different errors than the initial model, not fewer errors.

Third, percentages of reclassification and NRI are calculated for specific cutoff values, whereas IDI is a summary measure over all possible cutoff thresholds. Because IDI assesses improvement in prediction across all possible cutoff values, this measure is interpreted as a weighted area under the receiver operating characteristic curve, comparable to AUC (14, 25). In our simulation study, changes in AUC were reflected in changes in IDI. When ΔAUC was 0.005, IDI was close to zero, but when ΔAUC increased to 0.10, IDI increased accordingly and became highly statistically significant. Because no specific clinical cutoff thresholds are considered, IDI, like AUC, is not a measure of clinical utility but rather of clinical validity.

Fourth, we showed that across the range of cutoff values NRI followed a bimodal distribution (**Figure 2C**), which is explained by the differences in the net benefit in cases and in non-cases (**Figure 3**). When a single threshold is used and this threshold is lower than the disease risk, reclassification is markedly improved in non-cases and slightly worsened in cases. At cutoff values higher than the disease risk, reclassification is markedly improved in cases and worsened in non-cases. In the above example, where the disease risk was 10%, NRI had two peaks corresponding to the maximum net benefit among non-cases, when the cutoff was 6%, and the maximum net benefit among cases, when the cutoff was 16% (**Figure 3**). When it is more important to avoid unnecessary treatment (i.e., due to expensive intervention and serious side effects) in individuals who will not develop the disease, a threshold lower than the disease risk should be considered, and when it is more important that all at risk individuals who might develop disease receive preventive interventions (i.e., due to significant impact on outcome and safe interventions), a threshold higher than the disease risk may be chosen. Therefore, information on improvement in risk prediction in cases and non-cases separately may be more valuable than the global NRI measure.

Finally, reclassification measures did not change when the same ΔAUC was realized by fewer genetic risk factors with higher allele frequencies and stronger OR (refer to the **Web Figure 1 and 2**). However, the distributions of reclassification measures were different when the AUC of the starting model was higher (refer to the **Web Figure 3 and 4**). When the AUC of the starting model was higher, reclassification was observed over a wider range of cutoff values. Therefore, the amount of reclassification observed for any given ΔAUC and cutoff threshold varied with the magnitude of the AUC. This can be explained by the fact that the underlying risk distributions of cases and non-cases are more spread out (refer to the **Web Figure 3A and 4**).
Analyzing reclassification for all possible cutoff values suggests the optimal cutoff value could be based on its impact on medical decisions. However, the choice of the optimal cutoff value will be determined by weighing benefits and harms of false positive and false negative decisions. Also, the number of risk categories should be determined by clinical considerations, namely the number of available preventive treatment strategies. For our empirical study on type 2 diabetes, there were no guidelines on clinically useful risk categories. While a previous study on reclassification of type 2 diabetes risk had arbitrarily considered three risk categories, defined by two cutoff thresholds (16), we used only one risk threshold. It is important to note that percentages of reclassification will be higher when more cutoff values are used (13).

In recent empirical studies, reclassification is often given stronger emphasis than AUC (6, 13), but our findings clearly demonstrate that these measures provide different information. For a prediction model to be a useful tool for clinical practice or public health, the model should have

both appreciable clinical validity and clinical utility. Because AUC and IDI are not calculated for specific clinical thresholds, both measures reflect the clinical validity of a test. Percentages of total reclassification and NRI do vary with the choice of cutoff values for given levels of IDI and AUC improvement. AUC and reclassification measures provide complementary information about improvement in risk prediction.

REFERENCES

1. van Hoek M, Dehghan A, Witteman JC, et al. **Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study.** *Diabetes*. 2008;**57**(11):3122-3128.
2. Lyssenko V, Jonsson A, Almgren P, et al. **Clinical risk factors, DNA variants, and the development of type 2 diabetes.** *N Engl J Med*. 2008;**359**(21):2220-2232.
3. Vaxillaire M, Veslot J, Dina C, et al. **Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study.** *Diabetes*. 2008;**57**(1):244-254.
4. Morrison AC, Bare LA, Chambless LE, et al. **Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study.** *Am J Epidemiol*. 2007;**166**(1):28-35.
5. Hanley JA, McNeil BJ. **The meaning and use of the area under a receiver operating characteristic (ROC) curve.** *Radiology*. 1982;**143**(1):29-36.
6. Cook NR. **Use and misuse of the receiver operating characteristic curve in risk prediction.** *Circulation*. 2007;**115**(7):928-935.
7. Pepe MS, Feng Z, Huang Y, et al. **Integrating the predictiveness of a marker with its performance as a classifier.** *Am J Epidemiol*. 2008;**167**(3):362-368.
8. Janes H, Pepe MS, Gu W. **Assessing the value of risk predictions by using risk stratification tables.** *Ann Intern Med*. 2008;**149**(10):751-760.
9. Zethelius B, Berglund L, Sundstrom J, et al. **Use of multiple biomarkers to improve the prediction of death from cardiovascular causes.** *N Engl J Med*. 2008;**358**(20):2107-2116.
10. Schnabel RB, Sullivan LM, Levy D, et al. **Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study.** *Lancet*. 2009;**373**(9665):739-745.
11. Simmons RK, Sharp S, Boekholdt SM, et al. **Evaluation of the Framingham risk score in the European Prospective Investigation of Cancer-Norfolk cohort: does adding glycated hemoglobin improve the prediction of coronary heart disease events?** *Arch Intern Med*. 2008;**168**(11):1209-1216.
12. Paynter NP, Chasman DI, Buring JE, et al. **Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3.** *Ann Intern Med*. 2009;**150**(2):65-72.
13. Cook NR, Ridker PM. **Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures.** *Ann Intern Med*. 2009;**150**(11):795-802.
14. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., et al. **Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond.** *Stat Med*. 2008;**27**(2):157-172; discussion 207-212.
15. Kathiresan S, Melander O, Anevski D, et al. **Polymorphisms associated with cholesterol and risk of cardiovascular events.** *N Engl J Med*. 2008;**358**(12):1240-1249.
16. Meigs JB, Shrader P, Sullivan LM, et al. **Genotype score in addition to common risk factors for prediction of type 2 diabetes.** *N Engl J Med*. 2008;**359**(21):2208-2219.
17. Hofman A, Breteler MM, van Duijn CM, et al. **The Rotterdam Study: objectives and design update**. *Eur J Epidemiol*. 2007;**22**(11):819-829.
18. **Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus.** *Diabetes Care*. 1997;**20**(7):1183-1197.
19. Alberti KG, Zimmet PZ. **Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation.** *Diabet Med*. 1998;**15**(7):539-553.

20. Janssens AC, Aulchenko YS, Elefante S, et al. **Predictive testing for complex diseases using multiple genes: fact or fiction?** *Genet Med*. 2006;**8**(7):395-400.

21. **Analyse-it Software**, Ltd. 2009. (http://www.analyse-it.com/). (Accessed February 4, 2010).

22. DeLong ER, DeLong DM, Clarke-Pearson DL. **Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.** *Biometrics*. 1988;**44**(3):837-845.

23. Ihaka GR. R: A **Language for Data Analysis and Graphics.** *J Comput Graph Stat* 1996;**5**:299-314.

24. Harrell FE. Design and Hmisc R function library. 2006. (http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/RS). (Accessed June 3, 2009).

25. Ware JH, Cai T. **Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond'** by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). *Stat Med*. 2008;**27**(2):185-187.

SUPPLEMENTARY MATERIAL

Definitions of Reclassification Measures

Total reclassification is the percentage of individuals that chance from one risk category based on the original prediction model to a different risk category based on the updated model (6).

Net Reclassification Improvement (NRI) separately considers the reclassification in cases and non-cases. Cases are correctly classified when they move to a higher risk category and wrongly classified when they move to a lower category. Non-cases move correctly to a lower category and wrongly to a higher. NRI is the sum of the net correct moves: the proportion of cases moving up minus the proportion of cases moving down, plus the proportion of non-cases moving down minus the proportion of non-cases moving up (14).

$$NRI = (\hat{p}_{up,\ cases} - \hat{p}_{down,\ cases}) + (\hat{p}_{down,\ non\text{-}cases} - \hat{p}_{up,\ non\text{-}cases})$$

$$\hat{p}_{up,\ cases} = \frac{\#cases\ moving\ up}{\#cases}$$

$$\hat{p}_{down,\ cases} = \frac{\#cases\ moving\ down}{\#cases}$$

$$\hat{p}_{down,\ non\text{-}cases} = \frac{\#non\text{-}cases\ moving\ down}{\#non\text{-}cases}$$

$$\hat{p}_{uo,\ non\text{-}cases} = \frac{\#non\text{-}cases\ moving\ up}{\#non\text{-}cases}$$

Integrated Discrimination Improvement (IDI) compares the average of the predicted probability in cases and non-cases between the original and updated models (14).

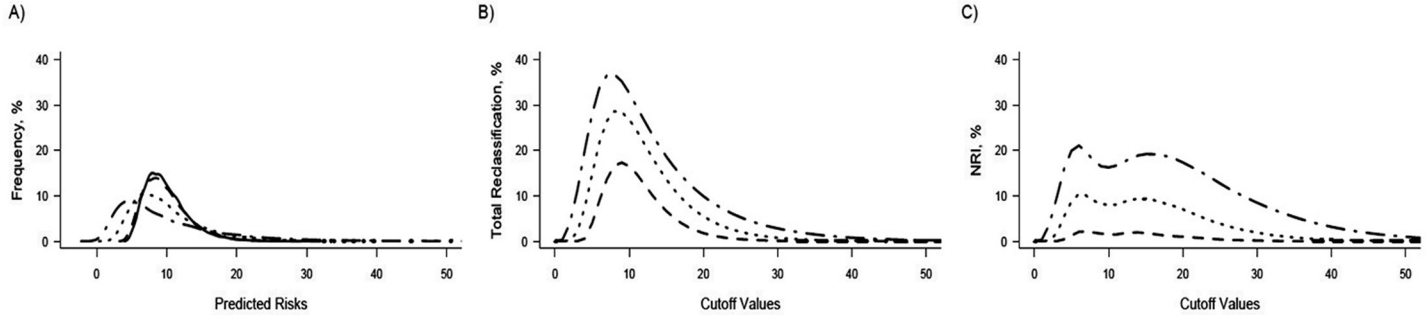
$$IDI = (\overline{\hat{p}}_{new,\ cases} - \overline{\hat{p}}_{old,\ cases}) - (\overline{\hat{p}}_{new,\ non\text{-}cases} - \overline{\hat{p}}_{old,\ non\text{-}cases})$$

where $\overline{\hat{p}}_{new,\ cases}$ is the mean of predicted probabilities based on the updated model for cases, $\overline{\hat{p}}_{old,\ cases}$ is the corresponding mean based on the original model, $\overline{\hat{p}}_{old,\ non\text{-}cases}$ is the mean of predicted probabilities based on the updated model for non-cases, and $\overline{\hat{p}}_{old,\ non\text{-}cases}$ is the corresponding mean based on the original model.

Simulation Scenarios

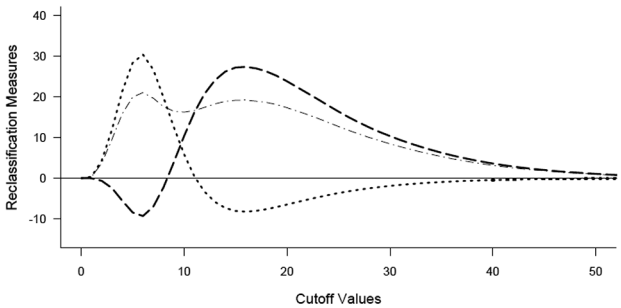
In order to determine whether the pattern of change in reclassification measures is influenced by other factors than the increment in AUC, we simulated two scenarios in which we increased the odds ratios and disease frequencies of the added genetic risk factors (**Figures 1 and 2**), or the disease frequency and AUC of the first model (**Figures 3 and 4**).

Figure S1. Risk distribution and reclassification measures for prediction models with higher odds ratios and genotype frequencies of the added polymorphisms.



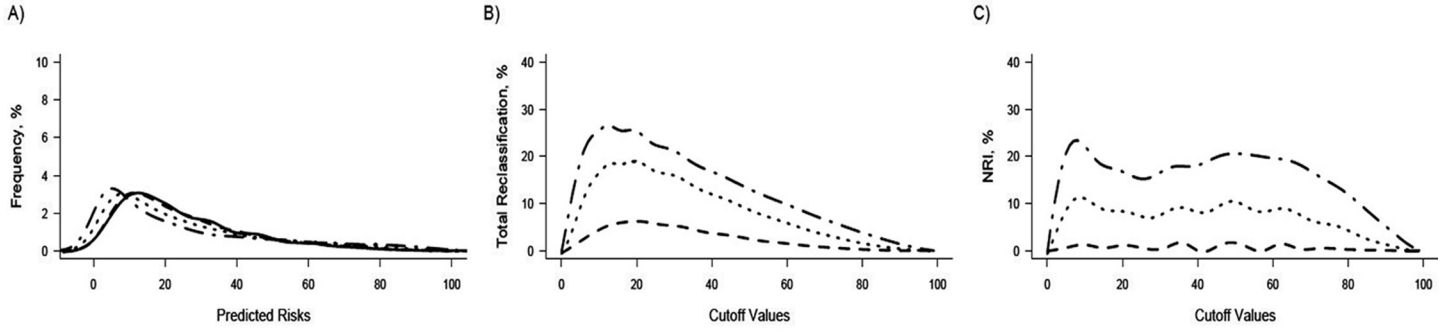
Distribution of predicted risks for the model based on 20 polymorphisms (continuous line) and for the updated models when ΔAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed and dotted line) (**1A**); Percentages of total reclassification per cutoff value for risk stratification when ΔAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed and dotted line) (**1B**); Net reclassification improvement (NRI) per cutoff value for risk stratification when ΔAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed and dotted line) (**1C**). The first model is based on 20 polymorphisms, each with an odds ratio of 1.2 and frequency of 20% (AUC=0.61). One, 6 and 15 genetic risk factors are added to the first model, each with an odds ratio of 1.4 and frequency of 40%. Disease risk is 10% and sample size is 10,000. Percentage of total reclassification and NRI are presented as mean values obtained from 100 simulations. Cutoff values are used to define two risk categories, one category included individuals with a risk higher or equal to the cutoff value and the other included those with a risk lower than the cutoff value.

Figure S2. Net reclassification improvement and its components for prediction models with higher odds ratios and genotype frequencies of the added polymorphisms.



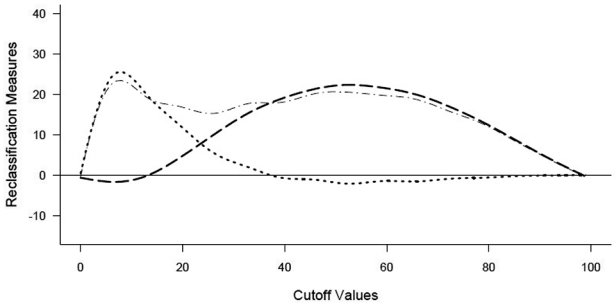
The dashed and dotted line shows the net reclassification improvement per cutoff value for risk stratification when ΔAUC is 0.10. The dotted line shows the percentages of correct (non-cases moving down) minus incorrect (non-cases moving up) reclassification in non-cases for each cutoff value. The dashed line shows the percentages of correct (cases moving up) minus incorrect (cases moving down) reclassification in cases for each cutoff value. Each value is obtained from the mean of 100 simulations. Cutoff values are used to define two risk categories, one category included individuals with a risk higher or equal to the cutoff value and the other included those with a risk lower than the cutoff value.

Figure S3. Risk distribution and reclassification measures for prediction models with higher AUC of the starting model and higher disease risk.



Distribution of predicted risks for the model based on 20 polymorphisms (continuous line) and for the updated models when ΔAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed and dotted line) (**3A**); Percentages of total reclassification per cutoff value for risk stratification when ΔAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed and dotted line) (**3B**); Net reclassification improvement (NRI) per cutoff value for risk stratification when ΔAUC is 0.005 (dashed line), 0.05 (dotted line), and 0.10 (dashed and dotted line) (**3C**). The first model is based on 20 polymorphisms, each with an odds ratio of 1.6 and frequency of 20% (AUC=0.75). Five, 55 and 120 genetic risk factors are added to the first model, each with an odds ratio of 1.2 and frequency of 20%. Disease risk is 25% and sample size is 10,000. Percentage of total reclassification and NRI are presented as mean values obtained from 100 simulations. Cutoff values are used to define two risk categories, one category included individuals with a risk higher or equal to the cutoff value and the other included those with a risk lower than the cutoff value.

Figure S4. Net reclassification improvement and its components for prediction models with higher AUC of the starting model and higher disease risk.



The dashed and dotted line shows the net reclassification improvement per cutoff value for risk stratification when ΔAUC is 0.10. The dotted line shows the percentages of correct (non-cases moving down) minus incorrect (non-cases moving up) reclassification in non-cases for each cutoff value. The dashed line shows the percentages of correct (cases moving up) minus incorrect (cases moving down) reclassification in cases for each cutoff value. Each value is obtained from the mean of 100 simulations. Cutoff values are used to define two risk categories, one category included individuals with a risk higher or equal to the cutoff value and the other included those with a risk lower than the cutoff value.

Chapter 3

Predictive genetic testing for the identification of high-risk groups: a simulation study on the impact of predictive ability

Raluca Mihaescu, Ramal Moonesinghe,
Muin J. Khoury, A. Cecile J.W. Janssens

Genome Medicine. 2011;3(7):51

ABSTRACT

Background: Genetic risk models could potentially be useful in identifying high-risk groups for the prevention of complex diseases. We investigated the performance of this risk stratification strategy by examining epidemiological parameters that impact the predictive ability of risk models.

Methods: We assessed sensitivity, specificity, positive and negative predictive value for all possible risk thresholds that can define high-risk groups and investigated how these measures depend on the frequency of disease in the population, the frequency of the high-risk group, and the discriminative accuracy of the risk model, as assessed by the area under the receiver-operating characteristic curve (AUC). In a simulation study, we modeled genetic risk scores of 50 genes with equal odds ratios and genotype frequencies, and varied the odds ratios and the disease frequency across scenarios. We also performed a simulation of age-related macular degeneration risk prediction based on published odds ratios and frequencies for six genetic risk variants.

Results: We show that when the frequency of the high-risk group was lower than the disease frequency, positive predictive value increased with the AUC but sensitivity remained low. When the frequency of the high-risk group was higher than the disease frequency, sensitivity was high but positive predictive value remained low. When both frequencies were equal, both positive predictive value and sensitivity increased with increasing AUC, but higher AUC was needed to maximize both measures.

Conclusions: The performance of risk stratification is strongly determined by the frequency of the high-risk group relative to the frequency of disease in the population. The identification of high-risk groups with appreciable combinations of sensitivity and positive predictive value requires higher AUC.

BACKGROUND

There is increasing interest in the potential use of testing multiple genetic variants for the prediction of common complex diseases such as type 2 diabetes, osteoporosis and cardiovascular disease, particularly because this could help targeting preventive and therapeutic interventions to individuals and groups with high genetic risk. While to date most genetic risk models show only modest predictive performance (1-7), improved prediction is expected when many new genetic risk factors are discovered in the coming years, both common and rare variants with intermediate to large effects on disease risk. Notwithstanding these anticipated discoveries, the predictive ability of genetic risk models for complex diseases is likely to remain modest, because non-genetic risk factors have a substantial impact on disease risk as well (8, 9).

Despite the modest predictive ability, some argue that genetic risk models can still be useful in health care and disease prevention to identify individuals at very high risk (10). Preventive strategies can be targeted to individuals at very high risk even though this may only be a small subgroup (11, 12). The feasibility of this strategy will depend not solely on the predictive ability of the risk model, but also on the threshold level that is chosen. For certain diseases, well defined clinical cut-off values exist, such as the Framingham risk score for cardiovascular disease (13, 14), but in most instances the relevant thresholds have not been determined. Risk thresholds are chosen on a cost-benefit analysis of false negative and false positive findings across all thresholds, and generally are a trade-off. High threshold values are needed to identify individuals with a high probability to develop future disease, but this may only identify a fraction of the patients, whereas lower thresholds will identify most individuals who will develop the disease but also classify many individuals wrongly at increased risk. Therefore, apart from the discriminative accuracy of the risk model, the threshold chosen has a major impact on the sensitivity, specificity, positive and negative predictive value (PPV, NPV) when the risk model is used as a dichotomous test.

For single genetic tests, the relationship between the epidemiological assessment of the genetic association (e.g. genotype frequency and odds ratio) and the predictive accuracy of the test (e.g. sensitivity and PPV) have been described by simple arithmetic formulas (15). These formulas show that the frequency of the risk variant relative to the frequency of disease determines whether the test will have high sensitivity or high PPV, and that both can be maximized only when genotype and disease frequencies are approximately equal. For instance, screening for a common disease using rare variants can detect only a few individuals at very high risk. Conversely, screening for a rare disease using common variants detects most individuals that will ultimately develop the disease at the cost of many false positive findings. It would be of

interest to make use of the genomic era developments in this analysis by including multiple risk variants.

In this study, we examined the performance of risk stratification based on genetic risk models that include multiple variants simultaneously. We investigated sensitivity, specificity, positive and negative predictive values of genetic risk models along the range of threshold values that can be chosen to define high-risk groups. This detailed exploration of the interrelationships between sensitivity, PPV, prevalence of risk group and disease prevalence using genetic risk scores instead of single risk variants has not been reported before. We repeated the analyses for thresholds that define high-risk groups with a frequency lower, equal or higher than the disease frequency for increasing values of the area under the receiver operating characteristic curve (AUC). To address these objectives we used simulated data across a wide variety of odds ratios and frequencies for genetic variants. We also carried out an additional simulation based on published odds ratios and frequencies for six genetic polymorphisms predicting age-related macular degeneration (AMD) risk (16).

METHODS

Simulated data

For the construction of simulated data sets, we used a modelling procedure that has been described in detail elsewhere (8). In short, the procedure creates a dataset in such a way that the frequencies and odds ratios (OR) of the risk genotypes and the disease risk match prespecified values. For simplicity, we assumed that each individual polymorphism had only two genotypes, one of which was associated with an increased risk of disease and the other with the referent or baseline risk. We assumed that genetic variants are inherited independently and that their joint effects follow a multiplicative risk model. And finally, we did not include gene-gene and gene-environment interactions in our analyses, which may further improve the predictive ability of genetic risk models. While these assumptions do impact the exact estimate of the AUC, e.g., modelling interaction effects might give higher AUC, they do not affect the main aim of our paper, namely impact of a given AUC on the sensitivity, specificity, PPV and NPV for different thresholds of the genetic risk model. The population size was 10 000 individuals and the population disease risk was varied across scenarios (i.e., 10 and 30%, respectively). We simulated 50 genetic risk factors, each having a risk genotype with a frequency of 30% and an OR that varied across scenarios (i.e., 1.1, 1.5 and 2.0, respectively).

Simulation study of age-related macular degeneration

We constructed a dataset using the disease risk from prevalence estimates in adults 40 years of age or over (17), and genotypic parameters from a published risk prediction model for AMD (16). We used the same modelling procedure as in our main simulation study and a sample size of 10 000 individuals. The model included six genetic risk variants in the following genes or gene-regions: *CFH* (rs1061170, rs1410996), *LOC387715* (rs10490924), *C2* (rs9332739), *CFB* (rs641153) and *C3* (rs2230199). For each locus we considered the effect from the univariate logistic regression analysis with AMD as outcome variable and the genetic variants as predictor variables. For each locus the three genotypes were entered independently, with the exception of *C2* and *CFB* for which the genotypes were grouped in two categories, one conferring an increased risk of disease. An additional file shows genotype ORs and genotypic frequencies in controls (see Additional file 1). The prevalence of disease in the AMD simulation was 9% (17).

Statistical analyses

In the main simulation study, we constructed a genetic risk score which was a simple count of the number of risk genotypes. Note that this score has perfect correlation with predicted risk, because all variants have the same frequency of the risk genotype and the same OR. The disease risk increases with the number of risk genotypes in the genetic risk model. In the AMD simulation, we derived predicted risks using logistic regression analysis with genetic risk variants entered as categorical variables. High-risk groups were defined as all individuals with risk scores above a chosen threshold.

First, to evaluate the impact of genotype frequencies and ORs on the overall discriminative accuracy of genetic risk models, we assessed the AUC (18). Next, to assess the predictive performance of genetic risk models for defining the high-risk group, we calculated the sensitivity, specificity, PPV and NPV for each possible threshold. The sensitivity is the percentage of individuals classified at high-risk among affected individuals and specificity is the percentage of individuals classified as not being at high-risk among unaffected individuals. PPV is the probability that individuals classified at high-risk will develop the disease, and NPV is the probability that individuals classified as not being at high-risk will remain free of disease. All measures are presented against cut-off values and the percentage of individuals at high-risk to examine the impact of the frequency of the high-risk group on the relationship between the sensitivity, specificity, PPV and NPV. Note that the frequency of the high-risk group defined by a certain threshold is different from the frequency of the risk genotype of each single genetic marker. Finally, to replicate the comparison between epidemiological assessment and predictive accuracy of the test (15) we assessed sensitivity, specificity, PPV and NPV for increasing AUC, in

high-risk groups with a frequency lower, equal or higher than the disease risk. For this purpose, the threshold values were chosen such that the frequency of the high-risk groups was 5%, 30% or 50% as the disease risk was 30%. To achieve variation in AUC we modelled 5 to 600 variants with OR of 1.1 and risk genotype frequency of 30%. Results are presented as means from 100 simulations. All analyses were performed using the R programming language version 2.8.0 (19).

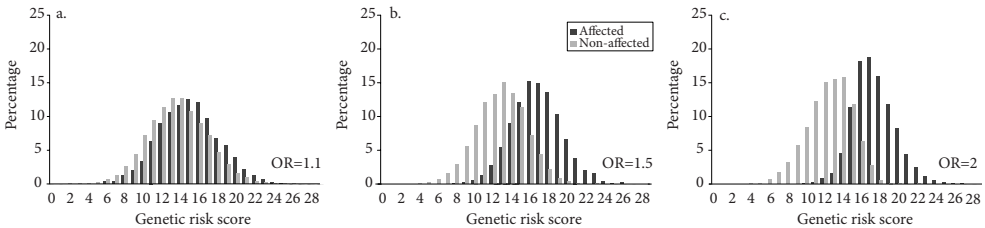
RESULTS

Figure 1 shows the distribution of genetic risk scores in affected and non-affected individuals for different ORs of the variants included. In our simulation study that included 50 genetic variants, the theoretical range of the risk score was 0 to 100, but the observed range was 2 to 32 with a median of 15 risk alleles. The AUC for the risk scores was 0.62 when the OR of each included variant was 1.1, 0.86 when OR was 1.5, and 0.94 when OR was 2.

Figure 2 shows the sensitivity and PPV for all possible thresholds of the genetic risk scores. When a higher threshold is used, the population at high-risk has a higher risk (higher PPV), but this will identify a smaller percentage of the affected individuals (lower sensitivity). Comparison of the graphs (**Figure 2a-c**) shows that for thresholds within the observed range of genetic risk scores, sensitivity and PPV were higher for higher ORs of the individual polymorphisms. When, for example, 15 was taken as the threshold risk score, the sensitivity was 67%, 91% and 97% and the PPV was 36%, 49% and 53% when the OR of each genetic variant was 1.1, 1.5 and 2, respectively. Using a higher threshold increased the specificity but decreased the NPV (see Additional file 2).

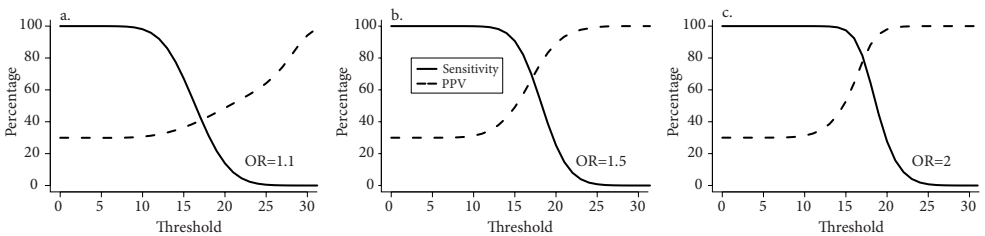
Figure 3 shows the relationship between the frequency of the high-risk group and the sensitivity, PPV, specificity and NPV. With increasing frequency of the population at high-risk sensitivity increased while PPV decreased; and specificity decreased while NPV increased. Note that because higher thresholds yield smaller high-risk categories, the lines depicting sensitivity and PPV show opposite trends in **Figures 2 and 3**. **Figure 3** shows that when e.g., the top 10% of the risk score distribution was considered the high-risk group sensitivity was 14% when the OR of each genetic variants was 1.1, indicating that most of the affected individuals were not detected. Sensitivity increased to 25% and 28% when OR was 1.5 and 2, showing that sensitivity did not markedly increase with increasing OR. The corresponding PPV values were 49%, 89% and 98%, indicating that PPV increased substantially with increasing OR. **Figure 3** shows that the lines cross when the frequency of the high-risk group is equal to 30%, i.e. the frequency of disease in the total population. The pattern remained the same when we repeated the analyses for a disease risk of 10% (see Additional file 3).

Figure 1. Distribution of genetic risk scores in affected and non-affected individuals.



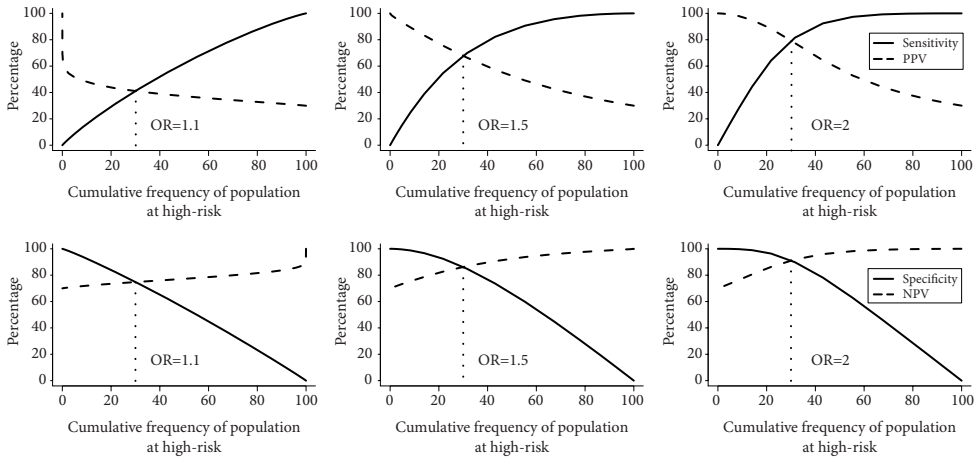
Genetic risk scores are based on 50 genetic risk variants. Each risk variant has an odds ratio of 1.1 (Figure 1a), 1.5 (Figure 1b) and 2 (Figure 1c). Disease risk is 30%.

Figure 2. Sensitivity and positive predictive value (ppv) for different thresholds.



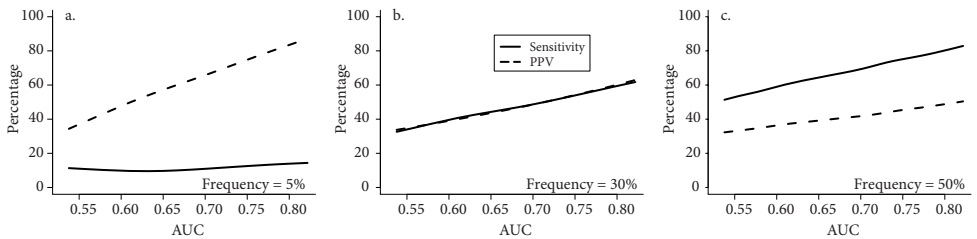
High-risk group is defined as all individuals with a genetic risk score equal or higher than the chosen threshold. Genetic risk scores are based on 50 genetic risk variants. The OR indicates the value of the odds ratio for each risk variant, which respectively is 1.1 (Figure 2a), 1.5 (Figure 2b) and 2 (Figure 2c). Disease risk is 30%.

Figure 3. Sensitivity, specificity, positive and negative predictive value (PPV, NPV) for different frequencies of the population at high-risk.



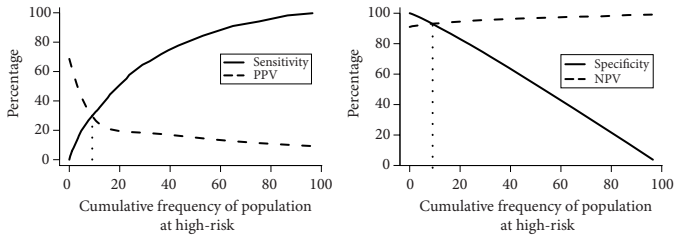
The frequency of the population at high-risk is defined as the proportion of individuals with a number of risk alleles equal or higher than the chosen threshold. The graphs in the upper row show the sensitivity and PPV for all possible risk thresholds, and the graphs in the lower row the specificity and NPV. Genetic risk scores are based on 50 genetic risk variants. The OR indicates the value of the odds ratio for each risk variant. Disease risk is 30%.

Figure 4. Sensitivity and positive predictive value (PPV) when the frequency of the high-risk group is lower, equal or higher than the disease risk.



The frequency of the high-risk group is defined as the proportion of individuals with a number of risk alleles equal or higher than the chosen threshold. High-risk groups have a frequency of 5% (Figure 4a), 30% (Figure 4b) and 50% (Figure 4c). Five to 600 variants are included in the genetic risk models to obtain an increase in the area under the ROC curve (AUC). Each risk variant has a frequency of 30% and odds ratio (OR) of 1.1. Disease risk is 30%.

Figure 5. Sensitivity, specificity, positive and negative predictive value (PPV, NPV), for age-related macular degeneration simulation.



Predicted risks of age-related macular degeneration are obtained using logistic regression analysis based on six genetic variants entered as categorical variables. The frequency of the population at high-risk is defined as the proportion of individuals with predicted risks equal or higher than the chosen risk threshold. The genotypic odds ratios and frequencies were obtained from the paper by Seddon et al. (16). Disease risk is 9%.

Increasing the OR of all variants included in the genetic risk score also increases the AUC of the risk score. **Figure 4** shows the impact of increasing AUC on sensitivity and PPV for high-risk groups that were of lower, equal or higher frequency than the disease frequency in the population. The AUC ranged from 0.51 to 0.82. When the frequency of the high-risk group was lower than disease frequency, PPV markedly increased with increasing AUC, but sensitivity remained low even for high AUC because, by definition, the high-risk group was rarer than the disease (**Figure 4a**). When the frequency of the high-risk group was higher than the disease risk, sensitivity reached around 80% but PPV remained below 50% when AUC was 0.82 (**Figure 4c**). Only when the size of the high-risk group was equal to the disease risk in the population, sensitivity and PPV were approximately equal and both increased with the increase in AUC (**Figure 4b**). However, when AUC was 0.82 both sensitivity and PPV were only slightly higher than 60%. Similarly, specificity and NPV were equal only when high-risk groups had a frequency equal to disease risk (data not shown).

Finally, we examined the same associations using simulated data based on published odds ratios and frequencies for six known AMD genetic risk factors. The range of predicted risks was 0.2% to 62% (see Additional file 4, which shows the distribution of predicted risks in individuals with and without AMD) and the AUC was 0.76 (95% CI 0.74 to 0.78). We observed the same impact of the relative magnitude of the size of the high-risk groups and disease risk on the sensitivity, specificity, PPV and NPV as in our main simulation study (see **Figure 5**).

DISCUSSION

This study investigated the relationships between sensitivity, PPV, prevalence of risk group and disease prevalence when genetic risk scores, as opposed to single risk variants, are used for risk stratification. A major finding from this analysis is that when the frequency of the high-risk group approximates the disease frequency, both sensitivity and PPV increase with higher AUC. At all other frequencies of the high-risk group, higher AUC will increase either sensitivity or PPV. Selecting the optimal cut-off threshold will consequently be a trade-off between higher sensitivity at the price of lower PPV, or vice versa.

While the relationship between the number of individuals carrying a certain genetic risk factor and the risk of disease in the population was shown to influence the screening performance for a single marker (15), we have proven this is also true for a genetic test comprised of multiple genetic risk factors. Furthermore, we extended the analyses to the context of the overall model performance, and looked at the influence of the discriminatory ability of a genetic model on screening parameters for risk groups with a frequency lower, equal or higher than the disease risk.

Genetic tests are usually assessed in term of their ability to distinguish risk groups with large differences in risk. Nevertheless, it has been shown that large relative risks are not sufficient to demonstrate the model's clinical validity and utility (20-22). Measures like sensitivity, specificity, PPV and NPV are needed to determine the clinical utility of the test (22). While sensitivity and specificity are not affected by the incidence of disease because they are characteristics of the test, PPV and NPV strongly depend on disease risk. However, even for rare diseases, risk groups with a high PPV may be selected. Kraft et al. used the example of prostate cancer 5-year risk prediction to illustrate this (22). They show that 60 year old men with 9 or more risk alleles and a positive family history for prostate cancer, which represent 1% of the population, have a risk of 30% to develop prostate cancer over the next 5 years. The incidence of disease in the population of 60 year old men is about 2%. Thus, the size of the group at high risk was smaller than disease risk. We show that in addition to a smaller size of the high risk group and high OR for the risk factors, a high AUC is needed to obtain a high PPV. In a recent study the AUC of a genetic score of 33 SNPs and family history of prostate cancer was estimated at 0.64 (23). A higher AUC is needed to select a risk group with bigger PPV, especially if the high risk group is targeted for invasive interventions.

The observation that the sensitivity and PPV are equal when the frequency of the high-risk group equals the frequency of disease in the population holds across different settings. First, this relationship holds irrespective of whether the disease risk refers to the lifetime risk, a cumulative incidence over certain time period or the disease prevalence. Evidently, if we

consider e.g., lifetime risks instead of 10-year risks, the frequency of the high-risk group for which the sensitivity and PPV are equal will be larger, because lifetime risks by definition are higher than 10-year risks. Then for the same AUC values, these larger high-risk groups will have higher sensitivity and PPV. However, prediction models that consider longer time periods generally have lower AUC, implying that combinations of higher sensitivity and PPV may not be observed. Put differently, lifetime risk models with lower AUC may yield the same sensitivity/PPV combination as 10-year risk models with higher AUC, but the value of using a model with low AUC may become questionable.

Second, the relationship also holds irrespective of how the risks are calculated. There are several ways in which genetic risks can be expressed. One is to use a simple genetic risk score based on the number of risk alleles carried. This approach, which we used in our analyses, assumes that each allele has the same effect on the risk of disease (24, 25). Another option is to calculate a weighted risk score, which is a genetic risk score where the risk alleles are weighted for their effect on disease risk (14). Besides constructing risk scores, one can also directly derive predicted risks from multivariate logistic regression analyses with genetic variants entered as continuous or categorical variables. Results presented in this study are applicable to simple count scores and more complex weighted risk scores such as predicted risks, as emphasized by the simulation of AMD risk prediction, since in this study we have evaluated cut-off values that simply dichotomise the risk. Nevertheless, it should be pointed out that different approaches will likely yield different AUC values.

Third, the relationship also holds for risk models in general, i.e., also other non-genetic risk models, such as the Framingham risk score for prediction of cardiovascular disease. Basically the relationship is valid for any continuous variable that is dichotomised to create risk groups, such as blood pressure, cholesterol or triglyceride level. This is also true for risk models that include together novel biomarkers and established risk factors, a topic that has attracted recently a lot of research (26, 27).

When risk models are used to target interventions to high-risk subgroups, these subgroups are defined by choosing cut-off values for the predicted risks. The cut-off corresponding to a frequency of the high-risk group equal to the disease frequency optimizes both the sensitivity and the PPV, but is not necessarily optimal. Cut-off values are chosen on the basis of cost-benefit analyses, balancing the harms and benefits of false positive and false negative classifications of risk. The cut-off defining a risk group with a frequency equal to disease frequency is optimal only when the harm and benefit have equal weights. Selection of optimal cut-off based on a decision-analytic approach is a complex process that requires detailed input information of measures like sensitivity, specificity, PPV, NPV and related costs. For example, a recent study reported the effect

of family history and 14 SNPs on the cost-effectiveness of chemoprevention with finasteride for prostate cancer (28). The results show that genetic testing may marginally improve the cost-effectiveness of chemoprevention in individuals with more risk alleles especially in men with a positive family history. However, no optimal cut-off number of risk alleles was determined and the cost-effectiveness varied significantly with small changes of the model parameters. Our analyses do show however that, when AUC is low to moderate, selecting a subgroup with a substantially increased risk (i.e., high PPV) will include only a small percentage of all people who will develop the disease (i.e., low sensitivity). Obviously, the predictive ability is the fundamental prerequisite of a test, but what level of predictive ability is needed varies between applications.

Our observations have implications for health care applications of genetic testing, but also for the direct-to-consumer offer of personal genome tests via the internet. For health care applications that need high PPV, such as targeting invasive interventions to people at the highest risk, a low AUC means that only a small proportion of this group will be identified. For applications that need high sensitivity, such as screening programs, the interventions will be given to a very large part of the population, mostly to people who will not develop the disease. And finally, low AUC means for personal genome testing that most people who will develop the disease will not be identified as having high risks.

CONCLUSIONS

Anticipating the advances in this field, it is essential to develop more rigorous approaches to evaluate the clinical usefulness of risk models (29, 30). We have shown that when a threshold for genetic risk is used for selection of individuals at high risk to develop disease in the future, sensitivity, specificity and PPV of the test are strongly influenced by the relative magnitude of the size of the high-risk group and the disease risk in the population. In addition, selection of high-risk groups with clinically useful combinations of sensitivity and positive predictive value is only possible when the AUC values are higher.

REFERENCES

1. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB, Sr., Cupples LA: **Genotype score in addition to common risk factors for prediction of type 2 diabetes.** *N Engl J Med* 2008;**359:**2208-2219.
2. van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, Hofman A, Sijbrands EJ, Janssens AC: **Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study.** *Diabetes* 2008;**57:**3122-3128.
3. Nguyen TV, Eisman JA: **Pharmacogenomics of osteoporosis: opportunities and challenges.** *J Musculoskelet Neuronal Interact* 2006;**6:**62-72.
4. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, Hirschhorn JN, Berglund G, Hedblad B, Groop L, Altshuler DM, Newton-Cheh C, Orho-Melander M: **Polymorphisms associated with cholesterol and risk of cardiovascular events.** *N Engl J Med* 2008;**358:**1240-1249.
5. Morrison AC, Bare LA, Chambless LE, Ellis SG, Malloy M, Kane JP, Pankow JS, Devlin JJ, Willerson JT, Boerwinkle E: **Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study.** *Am J Epidemiol* 2007;**166:**28-35.
6. van der Net JB, Janssens AC, Defesche JC, Kastelein JJ, Sijbrands EJ, Steyerberg EW: **Usefulness of genetic polymorphisms and conventional risk factors to predict coronary heart disease in patients with familial hypercholesterolemia.** *Am J Cardiol* 2009;**103:**375-380.
7. Richards JB, Rivadeneira F, Inouye M, Pastinen TM, Soranzo N, Wilson SG, Andrew T, Falchi M, Gwilliam R, Ahmadi KR, Valdes AM, Arp P, Whittaker P, Verlaan DJ, Jhamai M, Kumanduri V, Moorhouse M, van Meurs JB, Hofman A, Pols HA, Hart D, Zhai G, Kato BS, Mullin BH, Zhang F, Deloukas P, Uitterlinden AG, Spector TD: **Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study.** *Lancet* 2008;**371:**1505-1512.
8. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM: **Predictive testing for complex diseases using multiple genes: fact or fiction?** *Genet Med* 2006;**8:**395-400.
9. Wray NR, Yang J, Goddard ME, Visscher PM: **The genetic interpretation of area under the ROC curve in genomic profiling.** *PLoS Genet* 2010;**6:**e1000864.
10. Humphries SE, Yiannakouris N, Talmud PJ: **Cardiovascular disease risk prediction using genetic information (gene scores): is it really informative?** *Curr Opin Lipidol* 2008;**19:**128-132.
11. Jackson R, Lynch J, Harper S: **Preventing coronary heart disease.** *BMJ* 2006;**332:**617-618.
12. Manuel DG, Lim J, Tanuseputro P, Anderson GM, Alter DA, Laupacis A, Mustard CA: **Revisiting Rose: strategies for reducing coronary heart disease.** *BMJ* 2006;**332:**659-662.
13. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, Ridker PM: **Association between a literature-based genetic risk score and cardiovascular events in women.** *JAMA*, **303:**631-637.
14. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, Kivimaki M, Humphries SE: **Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study.** *BMJ* 2010;**340:**b4838.
15. Khoury MJ, Newill CA, Chase GA: **Epidemiologic evaluation of screening for risk factors: application to genetic screening.** *Am J Public Health* 1985;**75:**1204-1208.
16. Seddon JM, Reynolds R, Maller J, Fagerness JA, Daly MJ, Rosner B: **Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables.** *Invest Ophthalmol Vis Sci* 2009;**50:**2044-2053.

17. Klein R, Rowland ML, Harris MI: **Racial/ethnic differences in age-related maculopathy. Third National Health and Nutrition Examination Survey.** *Ophthalmology* 1995;**102:**371-381.

18. Hanley JA, McNeil BJ: **The meaning and use of the area under a receiver operating characteristic (ROC) curve.** *Radiology* 1982;**143:**29-36.

19. Vienna: R Foundation for Statistical Computing. R Development Core Team: **R: a language and environment for statistical computing.** Version 2.8.0. (http://www.R-project.org)

20. Janssens AC, van Duijn CM: **Genome-based prediction of common diseases: advances and prospects.** *Hum Mol Genet.* 2008;**17:**R166-173.

21. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P: **Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker.** *Am J Epidemiol.* 2004;**159:**882-890.

22. Kraft P, Wacholder S, Cornelis MC, Hu FB, Hayes RB, Thomas G, Hoover R, Hunter DJ, Chanock S: **Beyond odds ratios--communicating disease risk based on genetic profiles.** *Nat Rev Genet.* 2009; **10:**264-269.

23. Aly M, Wiklund F, Xu J, Isaacs WB, Eklund M, D'Amato M, Adolfsson J, Gronberg H: **Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm-1 cohort study.** *Eur Urol.* 2011;**60:**21-28.

24. Janssens AC, Moonesinghe R, Yang Q, Steyerberg EW, van Duijn CM, Khoury MJ: **The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases.** *Genet Med.* 2007;**9:**528-535.

25. Meigs JB: **Prediction of type 2 diabetes: the dawn of polygenetic testing for complex disease.** *Diabetologia* 2009;**52:**568-570.

26. Velagaleti RS, Gona P, Larson MG, Wang TJ, Levy D, Benjamin EJ, Selhub J, Jacques PF, Meigs JB, Tofler GH, Vasan RS: **Multimarker Approach for the Prediction of Heart Failure Incidence in the Community.** *Circulation* 2010;**122:**1700-1706.

27. Chao C, Song Y, Cook N, Tseng CH, Manson JE, Eaton C, Margolis KL, Rodriguez B, Phillips LS, Tinker LF, Liu S: **The lack of utility of circulating biomarkers of inflammation and endothelial dysfunction for type 2 diabetes risk prediction among postmenopausal women: the Women's Health Initiative Observational Study.** *Arch Intern Med.* 2010;**170:**1557-1565.

28. Reed SD, Scales CD, Jr., Stewart SB, Sun J, Moul JW, Schulman KA, Xu J: **Effects of family history and genetic polymorphism on the cost-effectiveness of chemoprevention with finasteride for prostate cancer.** *J Urol.* 2011;**185:**841-847.

29. Janssens AC, van Duijn CM: **Genome-based prediction of common diseases: methodological considerations for future research.** *Genome Med.* 2009;**1:**20.

30. Janssens AC, Gwinn M, Bradley LA, Oostra BA, van Duijn CM, Khoury MJ: **A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions.** *Am J Hum Genet.* 2008;**82:**593-599.

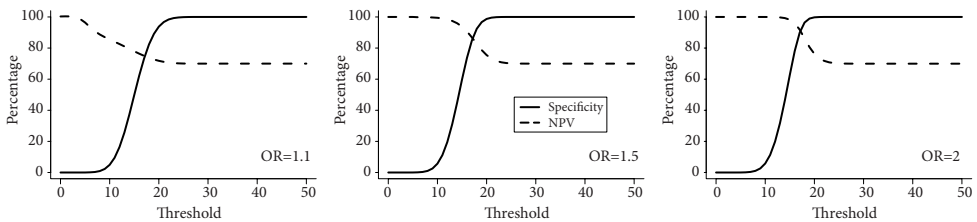
SUPPLEMENTARY MATERIAL

Table 1. Genotypic parameters for the age-related macular degeneration simulation.

Gene	SNP	Associated allele	OR[1]	Frequency, %[1]
CFH	rs1061170	C	2.1/4.1	44.6/24.0
CFH	rs1410996	C	3.1/7.2	40.4/46.0
LOC387715	rs10490924	T	2.8/6.2	38.2/9.3
C2	rs9332739	C	0.3	7.9
CFB	rs641153	T	0.6	12.3
C3	rs2230199	G	1.5/2.2	39.1/5.1

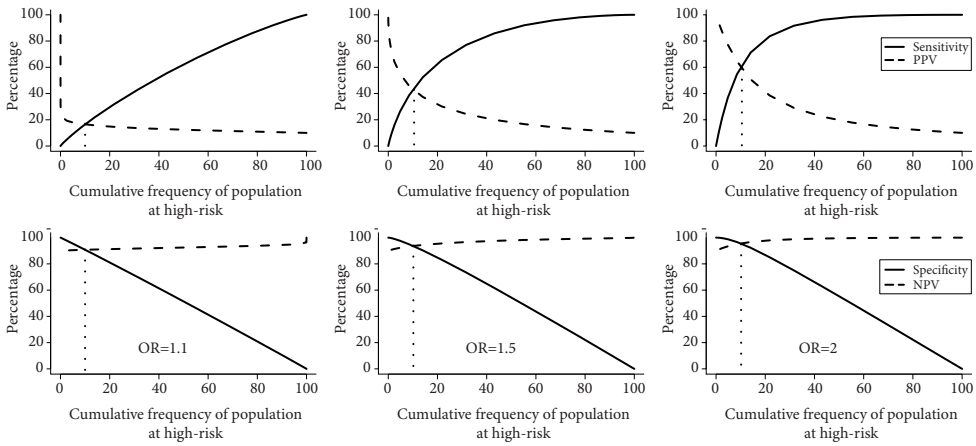
OR, odds ratio; SNP, single nucleotide polymorphisms. [1]Values are for heterozygous and homozygous carriers of the associated alleles or for the heterozygous and homozygous carriers combined.

Figure 1. Specificity and negative predictive value (NPV) for different thresholds.



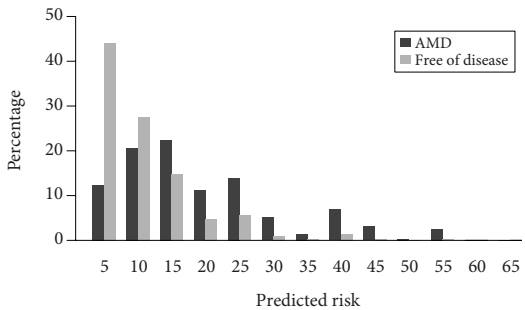
High-risk group is defined as all individuals with a genetic risk score equal or higher than the chosen threshold. Genetic risk scores are based on 50 genetic risk variants. The OR indicates the value of the odds ratio for each risk variant. Disease risk is 30%.

Figure 2. Sensitivity, Specificity, Positive and Negative Predictive Value (PPV, NPV) for Different Frequencies of the Population at High-risk.



The frequency of the population at high-risk is defined as the proportion of individuals with a number of risk alleles equal or higher than the chosen threshold. The graphs in the upper row show the sensitivity and PPV for all possible risk thresholds, and the graphs in the lower row the specificity and NPV. Genetic risk scores are based on 50 genetic risk variants. The OR indicates the value of the odds ratio for each risk variant. Disease risk is 10%.

Figure 3. Distribution of predicted risks in individuals with and without age-related macular degeneration.



Predicted risks are obtained from logistic regression analysis with age-related macular degeneration as outcome variable and 6 genetic risk variants entered as categorical predictor variables. We used the same genotypic odds ratios and frequencies as in the paper by Seddon et al.[1] Disease risk is 9%.

SUPPLEMENTAL REFERENCES

1. Seddon JM, Reynolds R, Maller J, Fagerness JA, Daly MJ, Rosner B: **Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables.** *Invest Ophthalmol Vis Sci* 2009;**50:**2044-2053.

Chapter 4

Estimating the predictive ability of genetic risk models using data from genome-wide association studies

Suman Kundu, Raluca Mihaescu, Catelijne Meijer,
Rachel Bakker, A. Cecile J.W. Janssens

In preparation

ABSTRACT

Objective: Genome-based risk stratification may improve the effectiveness of preventive and therapeutic medicine, but investigating such applications is premature when the expected predictive ability of genetic risk models is unknown. We assessed how accurately the predictive ability of genetic risk models can be estimated using odds ratios and frequencies of single-nucleotide polymorphisms (SNPs) obtained from published genome-wide association studies (GWAS).

Methods: We identified published empirical prediction studies that reported the area under the ROC curve (AUC) as a measure of predictive ability for genetic risk models and searched GWAS articles for all SNPs included in the models. We extracted odds ratios and risk allele frequencies, which were used to construct genotypes and disease status for a hypothetical population. Using these hypothetical data we reconstructed genetic risk models and compared their AUC values to those reported in the prediction studies.

Results: The accuracy of the AUC values estimated from simulated data varied with the method used for the construction of the risk models. When prediction studies used logistic regression analysis to construct the genetic risk model, AUC values estimated by the simulation method were similar to those published in the prediction studies with an average absolute difference of 0.02. This average absolute difference was 0.03 and 0.05 when the prediction studies used unweighted and weighted risk scores.

Conclusions: The predictive ability of genetic risk models can be estimated using simulated data based on GWAS results. Simulation methods can be useful to decide whether empirical investigation of genetic risk models is warranted.

INTRODUCTION

Genome-based risk stratification is one of the promises of the genomics research for the public health practice of screening and prevention of common diseases. In a stratified program, a risk model is used to distinguish risk groups that receive different interventions. This risk model can be based on clinical risk factors, genetic risk factors or a combination, and interventions can be different intensities of the same, e.g. varying screening interval or drug dose, or entirely different strategies, e.g. behavioral interventions, drug therapy and prophylactic surgery. In either case, the genetic risk model needs to have appreciable predictive ability to yield a meaningful categorization into risk groups.

Empirical studies on genetic risk models for multifactorial diseases so far show that the predictive ability is moderate at best (1, 2), with a few promising exceptions (3, 4). The predictive ability is expected to improve further with the identification of novel genetic variants, including common variants with smaller effects and rarer variants with larger effects (5), but this improvement is not evident. For example, genetic risk models with up to 40 variants predicted type 2 diabetes marginally better than models with fewer variants included (1), and rare variants only improve the predictive ability when they are not too rare (Chapter 5, this thesis).When newly discovered variants may not evidently improve the predictive ability, modeling strategies that quantify the predictive ability of genetic risk models can be an efficient approach to justify whether empirical investigation is warranted.

Several different modeling methods have been used to investigate the predictive ability of genetic risk models (6-10). All these methods assess the predictive ability as the degree to which the genetic risk model can discriminate between patients and nonpatients, quantified by the area under the receiver operating characteristic curve (AUC). Using epidemiological parameters such as the odds ratios (ORs) and frequencies of genetic variants, and a population-average risk of disease, these methods obtain the AUC values by simulating datasets for hypothetical populations (6, 7) or using analytical formulas (8-10). A comparison of the methods showed that the simulation methods could reproduce the AUC values of published prediction studies fairly accurately particularly when the odds ratios and allele frequencies were obtained from these same prediction studies (11). This observation demonstrates that the AUC value can be estimated using a simple model from a few basic parameters, but is of limited practical usefulness. When the data is available to obtain the ORs and frequencies, it can as well be used to assess the genetic risk model. Yet, calculation of the expected predictive ability prior to the collection of empirical data, based on published ORs and frequencies, could justify whether the study is worth conducting.

The question thus remains whether the AUC values can also be reproduced when the ORs and frequencies are obtained from other studies, e.g. from the GWAS that reported their discovery. One would expect that the estimated AUC values are different when ORs and frequencies differ and modeling strategies are used to combine their effects. Yet, the variation in ORs may not be large enough to produce substantially different AUCs, particularly since the AUC is argued to be an insensitive metric unable to detect minor improvements in predictive ability (12).

We investigated how accurately modeling methods can reproduce AUC values from genetic prediction studies when ORs and frequencies of the genetic variants are obtained from published GWAS. We additionally aimed to reproduce plots that are frequently used in prediction studies. As accuracy might be related to the computational method that was used to calculate individual risks, we selected published studies that had used unweighted or weighted risk scores or logistic regression analysis.

METHODS

We aimed to reproduce AUC values from published empirical studies using simulated data. For each prediction study, we constructed genotypes and disease status for a hypothetical population, estimated disease risks and assessed the AUC. To create a dataset for each hypothetical population, we used ORs and allele frequencies from published GWAS. The simulation method, study selection, data extraction and analyses are described next.

Simulation method

The simulation method has been described in detail elsewhere (7). In short, the method creates a dataset of individual genotypes and disease status for a hypothetical population based on ORs and frequencies of genetic variants, and a population-average risk of disease. The dataset is constructed in such a way that the population-average disease risk, allele frequencies and ORs estimated from the dataset match the prespecified input values. In this study, the input values were obtained from published GWAS (see below). Genotypes and disease status were constructed for 100,000 individuals. Finally, to estimate individual disease risks we used Bayes' theorem, which specifies that the posterior odds of disease are obtained by multiplying the prior odds by the likelihood ratio of the individual genotype profile.

Selection of genetic risk prediction studies

We aimed to identify studies that assessed genetic risk models based on single-nucleotide polymorphisms (SNPs). Risk models that additionally included nongenetic risk factors or genetic variants other than SNPs, such as haplotypes and copy number variations, were not considered. The risk models could be unweighted risk scores, weighted risk scores or logistic

regression models. Unweighted risk scores assume that all SNPs contribute equally to the risk of disease and are commonly calculated as the number of risk alleles across all SNPs. Weighted risk scores and logistic regression models assume different effects of SNPs, where the scores or linear predictors are calculated as the sum of the *log(OR)* of the risk alleles across all SNPs. In weighted risk scores the weights (ORs) are obtained from the literature, generally GWAS or meta-analyses, where in logistic regression models the weights are estimated in the prediction study.

Two different strategies were used to select genetic prediction studies. We searched PubMed for studies on diseases that are frequently investigated for genetic risk prediction, namely age-related macular degeneration, colorectal cancer, Crohn's disease, prostate cancer, type 1 diabetes, and type 2 diabetes, and additionally considered prediction studies from our recent review on type 2 diabetes (1). We selected studies that (i) reported the AUC value for a genetic risk model that was based on SNPs only, and (ii) explicitly stated which SNPs were included. For each disease we used the following search strategy in PubMed: "(genetic(title) OR genomic(title) OR genes(title) OR DNA(title) OR polymorphism(title) OR polygenic(title)) AND (risk(All Fields)) AND (score(All Fields) OR model(All Fields) OR prediction(All Fields))" (accessed August 2012). The PubMed search yielded in 515 publications, of which 20 met the inclusion criteria. From our review on type 2 diabetes we identified 10 prediction studies reporting about genetic risk models that were based on SNPs only. Five of these were already covered from the PubMed search, thus a total of 25 studies were included in our analyses.

Data extraction

From the selected prediction studies, we retrieved citations for the SNPs included in the risk models. Ideally, these citations were GWAS, meta-analyses or pooled analyses, from which we extracted unadjusted per allele ORs with the 95% confidence intervals (CIs) and the frequencies of the risk alleles in controls. We made the following decisions to handle multiple citations and missing data: if more than one citation was given for the same SNP, we selected the study with the largest sample size; if the cited study did not report per allele ORs, these were calculated from per genotype ORs; if CIs were not reported, these were calculated from an allele by disease status contingency table using the sample size and allele frequencies from the cited study; and if allele frequencies were not reported, frequencies were obtained from the 1000 Genomes Project (13). If the cited publications did not report original analyses of OR and allele frequencies, for example when citations were reviews or earlier prediction studies, we took ORs and frequencies from the largest GWAS or meta-analysis published up to 12 months before the prediction study. And finally, if no GWAS or meta-analysis was published, as was the case for two SNPs in our analysis, we used ORs and allele frequencies from the prediction study itself. Two investigators (SK, CM)

independently extracted data from the cited publications and discrepancies were discussed with a third investigator (AJ). **Supplementary Table 1** lists all SNPs, risk allele frequencies and per allele ORs that were used in the analyses.

Data analyses

The primary outcome measure was the AUC value of the genetic risk model. The AUC was obtained from the predicted risks and the disease status. When the risk model in the published prediction study was constructed as weighted risk scores or logistic regression model, the disease risks calculated using Bayes' theorem were used. When the risk model was constructed as unweighted risk scores, we similarly obtained unweighted allele scores in our simulated data.

To estimate the AUC values we examined two different scenarios: in the first, we estimated AUC using the point estimates of published ORs, and in the second, we randomly draw ORs from the published 95% CIs for each SNP. To obtain robust estimates of the AUC, all simulations were repeated 100 times. Results are presented as averages of 100 iterations.

As secondary outcome measures, we also aimed to reproduce plots that are frequently used in prediction studies. We selected four different plots: a histogram showing the distribution of the number of risk alleles among patients and nonpatients, a scatter plot showing the variation in predicted risks stratified by the number of risk alleles, a quintiles plot presenting the ORs with 95% CIs for quintiles of genetic risks, and a receiver operating characteristic (ROC) plot for genetic risk models showing the sensitivity versus 1-specificity across all possible threshold values of the predicted risks. For each plot we arbitrarily selected an example from the published prediction studies. We constructed the data for the hypothetical population in the same way as explained above, except that we used the sample size and population disease risk of the published prediction study because these impact the CIs in the quintiles plot and the absolute risks in the scatter plot. All analyses were performed using the PredictABEL package in R software, version 2.14.1 (www.r-project.org) (14).

RESULTS

Table 1 shows the published AUC values along with those estimated using the simulation method. When the prediction studies had used unweighted risk scores or weighted risk scores the absolute differences ranged from 0.01 to 0.06 (mean 0.03) and from 0.01 to 0.08 (mean 0.05), respectively. When prediction studies used logistic regression models to calculate individual risks the absolute difference in AUC ranged from 0.00 to 0.04 (mean 0.02). For each prediction study, the average AUC estimated from random values of the ORs obtained from the published 95% CIs was exactly the same as the AUC value estimated from the point estimates of published ORs (data not shown).

Table 1. Published and simulated AUC values for genetic risk models.

Published prediction study					Simulated data		Difference between published and simulated AUC
Disease	**Study design**	**Reference**	**Number of SNPs**	**AUC**	**AUC based on point estimates of ORs**	**AUC (95% CI) based on random values from 95% CI of ORs**	
Unweighted risk score							
Crohn's disease	Case-control	(18)	7	0.70	0.67	0.67 (0.66-0.67)	-0.03
Prostate cancer	Case-control	(20)	33	0.64[a]	0.67	0.67 (0.67-0.67)	0.03
Type 1 diabetes	Case-control	(21)	7	0.65[a]	0.64	0.64 (0.63-0.64)	-0.01
Type 2 diabetes	Cross-sectional	(22)	15	0.57	0.59	0.59 (0.59-0.60)	0.02
Type 2 diabetes	Prospective cohort	(23)	17	0.62	0.60	0.59 (0.59-0.60)	-0.02
Type 2 diabetes	Prospective cohort	(16)	18	0.56	0.60	0.60 (0.60-0.60)	0.04
Type 2 diabetes	Prospective cohort	(24)	18	0.58[b]	0.59	0.59 (0.58-0.59)	0.01
Type 2 diabetes	Cross-sectional	(25)	19	0.55	0.60	0.60 (0.60-0.61)	0.05
Type 2 diabetes	Prospective cohort	(26)	20	0.54	0.60	0.60 (0.60-0.60)	0.06
Weighted risk score							
Prostate cancer	Case-control	(27)	28	0.62	0.66	0.66 (0.66-0.66)	0.04
Prostate cancer	Case-control	(28)	33	0.59	0.67	0.67 (0.67-0.67)	0.08
Type 2 diabetes	Cross-sectional	(22)	15	0.59	0.60	0.60 (0.60-0.60)	0.01
Type 2 diabetes	Prospective cohort	(26)	20	0.55	0.61	0.61 (0.61-0.61)	0.06
Logistic regression model							
AMD	Case-control	(29)	3	0.73	0.69	0.69 (0.69-0.69)	-0.04
AMD	Case-control	(30)	4	0.77	0.76	0.76 (0.76-0.76)	-0.01

Published prediction study					Simulated data		Difference between published and simulated AUC
Disease	Study design	Reference	Number of SNPs	AUC	AUC based on point estimates of ORs	AUC (95% CI) based on random values from 95% CI of ORs	
Logistic regression model							
AMD	Case-control	(31)	13	0.82	0.78	0.78 (0.78-0.78)	-0.04
Colorectal cancer	Case-control	(32)	10	0.57	0.60	0.60 (0.60-0.60)	0.03
Colorectal cancer	Prospective cohort	(33)	14	0.58	0.60	0.60 (0.60-0.60)	0.02
Crohn's disease	Case-control	(18)	7	0.71	0.68	0.68 (0.67-0.68)	-0.03
Prostate cancer	Prospective cohort	(34)	36	0.67	0.69	0.69 (0.69-0.69)	0.02
Prostate cancer	Case-control	(19)	9	0.61	0.62	0.62 (0.61-0.62)	0.01
Type 2 diabetes	Case-control	(35)	3	0.58	0.59	0.59 (0.59-0.60)	0.01
Type 2 diabetes	Prospective cohort	(36)	3	0.56	0.58	0.58 (0.58-0.58)	0.02
Type 2 diabetes	Case-control	(37)	11	0.62	0.61	0.61 (0.61-0.61)	-0.01
Type 2 diabetes	Case-control	(38)	11	0.63	0.63	0.63 (0.63-0.63)	0.00
Type 2 diabetes	Cross-sectional	(39)	17	0.59	0.61	0.61 (0.61-0.61)	0.02
Type 2 diabetes	Prospective cohort	(16)	18	0.60	0.61	0.61 (0.61-0.61)	0.01
Type 2 diabetes	Case-control	(15)	18	0.60	0.61	0.61 (0.61-0.62)	0.01
Type 2 diabetes	Case-control	(40)	19	0.60	0.61	0.61 (0.61-0.62)	0.01

AUC, area under the receiver operating characteristic curve; CI, confidence interval; AMD, age-related macular degeneration; GWAS, genome-wide association studies; SNP, single nucleotide polymorphism.

[a]Adjusted for age; [b]Adjusted for sex

Table 2. Odds ratios of 18 single nucleotide polymorphisms in two prediction studies on type 2 diabetes and their corresponding values in the cited genome-wide association studies.

Gene	SNP	OR in prediction study		OR in cited GWAS*	
		GoDARTS study[15]	Rotterdam study[16]	GoDARTS study	Rotterdam study
ADAM30/NOTCH2	rs2641348†	1.15 (1.01, 1.30)	1.01 (0.88, 1.17)	1.10 (1.06, 1.15)	
ADAMTS9	rs4607103‡‡	1.05 (0.96, 1.16)	1.14 (1.03, 1.28)	1.09 (1.06, 1.12)	
CDC123	rs12779790¥	1.10 (0.99, 1.21)	1.05 (0.94, 1.19)	1.11 (1.07, 1.14)	
CDKAL1	rs10946398§	1.11 (1.02, 1.21)	1.11 (1.02, 1.22)	1.12 (1.08, 1.16)[1]	
CDKN2A/2B	rs10811661	1.21 (1.08, 1.35)	1.10 (0.98, 1.24)	1.20 (1.14, 1.25)	
CDKN2A/2B	rs564398‡	1.13 (1.04, 1.22)	1.04 (0.95, 1.14)	1.12 (1.07, 1.17)	
FTO	rs8050136	1.11 (1.02, 1.20)	1.09 (0.99, 1.19)	1.15 (1.09, 1.22)	
HHEX-IDE	rs1111875	1.02 (0.94, 1.11)	1.06 (0.97, 1.15)	1.13 (1.08, 1.17)	
IGF2BP2	rs4402960	1.12 (1.03, 1.22)	1.11 (1.01, 1.22)	1.17 (1.10, 1.25)	
JAZF1	rs864745§§	1.00 (0.93, 1.09)	1.09 (1.00, 1.19)	1.10 (1.07, 1.13)	
KCNJ11	rs5219	1.25 (1.15, 1.36)	1.03 (0.93, 1.13)	1.18 (1.04, 1.34)	1.14 (1.10, 1.19)
PPARG	rs1801282	1.21 (1.07, 1.36)	1.09 (0.95, 1.24)	1.14 (1.08, 1.20)	
SLC30A8	rs13266634	1.10 (1.01, 1.20)	1.13 (1.02, 1.24)	1.12 (1.07, 1.16)	
TCF2	rs757210††	1.07 (0.99, 1.16)	1.07 (0.98, 1.18)	1.12 (1.07, 1.18)	1.22 (1.15, 1.30)[1]
TCF7L2	rs7903146	1.36 (1.24, 1.48)	1.31 (1.19, 1.44)	1.47 (1.33, 1.62)	1.38 (1.31, 1.46)
THADA	rs7578597	1.04 (0.90, 1.19)	1.10 (0.96, 1.27)	1.15 (1.10, 1.20)	
TSPAN8/LGR5	rs7961581¶	1.09 (1.00, 1.19)	1.09 (0.99, 1.20)	1.09 (1.06, 1.12)	
WFS1	rs10010131**	1.07 (0.99, 1.16)	1.12 (1.05, 1.27)	1.11 (1.07, 1.16)	

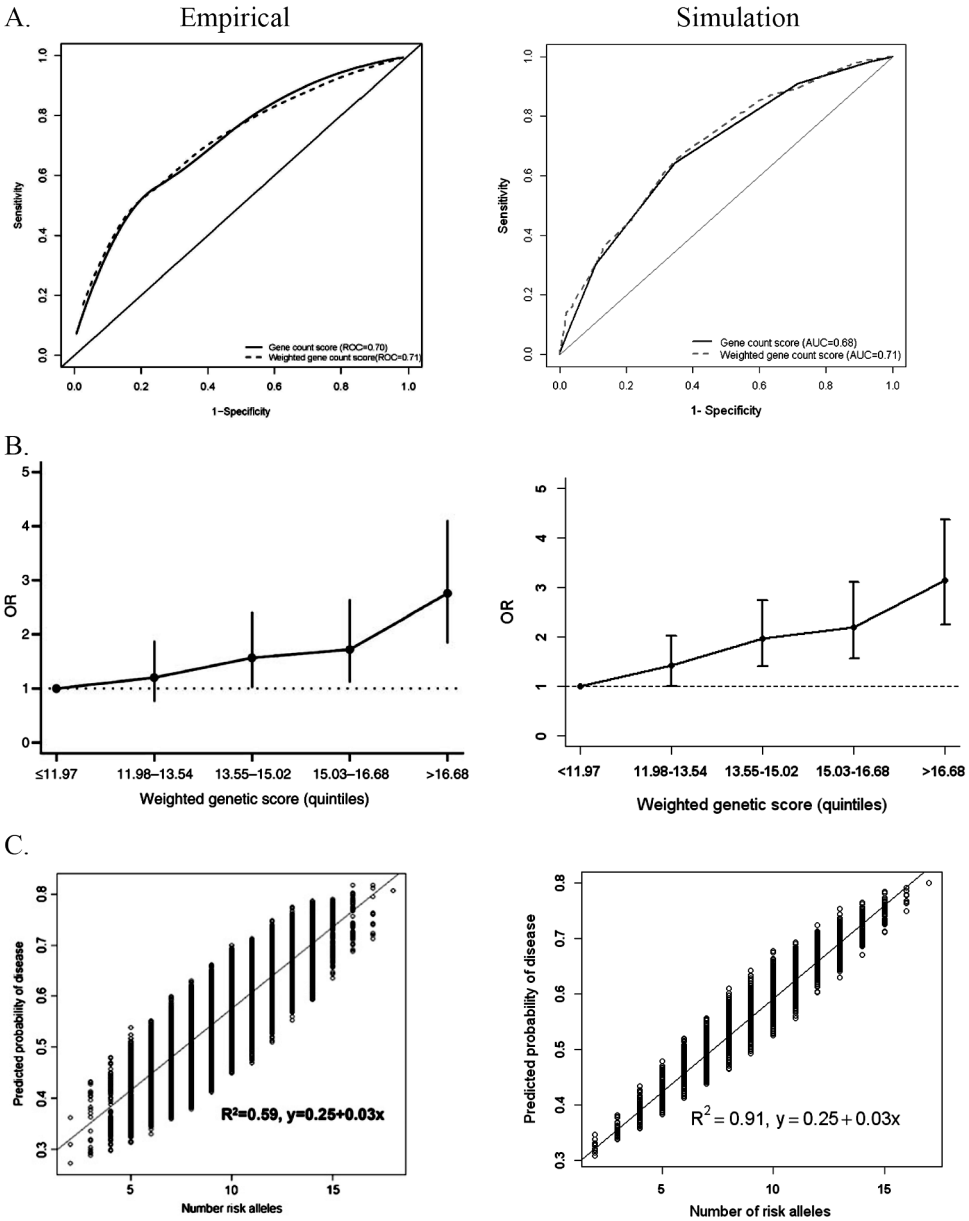
The table is adapted from (41). The risk models of the GoDARTS study and the Rotterdam Study included the same 18 genes and both had an AUC of 0.60. The AUC values from simulated data were 0.62 and 0.61, respectively. For several genes, the GoDARTS study and the Rotterdam Study used different SNPs: †rs1493694, r^2=0.74; ‡rs1412829, r^2=0.97; §rs7754840, r^2=1.00; ¶rs1353362, r^2=0.96; ¥rs11257622; r^2=0.83; **rs10012946, r^2=1.00; ††rs4430796, r^2=0.61; ‡‡rs4411878, r^2=0.95; §§rs1635852, r^2=0.97.

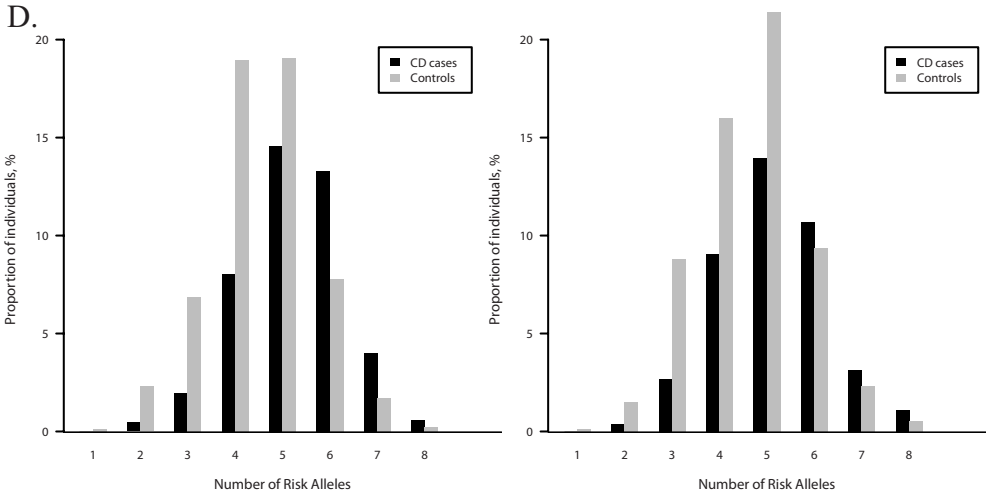
The SNPs used in the GWAS studies are proxy for the SNPs listed in the table: [1] The same proxy SNPs mentioned above are used; [2] The proxy SNP is: rs10923931, r^2=0.85; [3]The proxy SNP is: rs5015480, r^2=0.97.

* When only one value is presented, both prediction studies cited the same GWAS.

SNP, single nucleotide polymorphism; GWAS, genome-wide association study.

Figure 1. Plots published in the empirical prediction studies and those produced by the simulation method.





A. Receiver operating characteristic (ROC) curves for unweighted and weighted gene count scores (18); **B**. Quintiles plot presenting the odds ratios with 95% confidence intervals by quintiles of the weighted genetic scores (22)[1]; **C**. Scatter plot showing the variation in predicted risks stratified by the number of risk alleles (32)[2]; **D**. Histogram showing the distribution of the number of risk alleles among patients and nonpatients (18).

[1] Springer and *Diabetologia*, volume **52,** 2009, pages 600-8, Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score – the CoLaus Study, Lin, X. *et al.*, Figure 1b, original copyright notice is given to the publication in which the material was originally published; with kind permission from Springer Science and Business Media.

[2] Reproduced from Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42 103 individuals., Dunlop, M.G. et al, 62:871-881, copyright notice 2013, with permission from BMJ Publishing Group Ltd.

To illustrate how values for the input parameters can differ between two comparable studies on the same risk model, we compared the ORs for two individuals, prediction studies on type 2 diabetes (**Table 2**). These two risk models included the same genes and had the same observed AUC value (0.60), despite the fact that the observed ORs for several SNPs were different (15, 16). For example, the OR of *KCNJ11* was 1.25 in the GoDARTS study and 1.03 in the Rotterdam study, and the OR of *NOTCH2* was 1.15 and 1.01, respectively. The ORs from the GWAS or meta-analyses were similar for the two studies, but generally higher than those reported in the two prediction studies themselves. Risk allele frequencies for each variant did not differ between the cited studies and the prediction studies (**Supplementary Table S1**). The simulation method estimated an AUC of 0.62 for the GoDARTS study and 0.61 for the Rotterdam study.

Figure 1 shows four plots from published prediction studies and their equivalents produced by the simulation method. In general, the more detailed the information presented in the plots, the more differences were observed between the empirical and simulated studies. The ROC curve and the histogram were more similar between the empirical and simulation studies

(**Figure 1a and 1d**), than the quintiles and scatter plots (**Figure 1b and 1c**). The scatterplot showed a wider spread of predicted risks for each risk allele in the empirical data compared to the simulation study. For the quintiles plot, the accuracy of the reproduction appeared to be affected by the choice of cut-off values and the simulation sample size. Between iterations, the graphs differed most when the sample size and cut-off values of the published study were considered (**Supplementary Figure 1a**). The variation was less when the cut-off values were chosen based on the estimated risk scores, and hence varied, between the iterations (**Supplementary Figure 1b**), particularly when sample size was increased to 100,000 and cut-off values of the published study were considered (**Supplementary Figure 1c**).

DISCUSSION

We investigated how accurately modeling studies can estimate the AUC values from empirical prediction studies using ORs and frequencies from GWAS. The simulation method used in this study estimates AUC values fairly accurately, predominantly when prediction studies used logistic regression models. The simulation method also reproduced plots that are frequently reported in prediction studies.

Before discussing the implications of our findings, the assumptions of the simulation method need to be addressed. To estimate individual disease risks, the simulation method assumes that (i) the combined effect of genetic variants follows a multiplicative (log additive) risk model; (ii) genetic variants inherit independently, that is no linkage disequilibrium between the variants; (iii) genetic variants have independent effects on the disease risk, which indicates no interaction among variants; and (iv) effect sizes for genetic variants are considered as per allele ORs unadjusted for any other variants (marginal effects). These assumptions impact the predictive ability of risk models and therefore affect AUC values (17), but they seem to be valid for two reasons. First, with the exception of the marginal effect sizes, the assumptions are also assumed in empirical prediction studies. The simulation method assumes the same risk models as the empirical studies, as Bayes' theorem is computationally similar to logistic regression models and weighted risk scores. The only difference, the difference between marginal and adjusted effect sizes, seems not large enough to affect an aggregate measure like AUC. Second, we recently showed that the observed AUC values and those estimated from simulated data were similar when ORs and frequencies of the genetic variants were obtained from the empirical prediction study itself (11), suggesting that the modeling method itself produces accurate results. Therefore, we do not expect that the modeling assumptions and the simulation method as such influence the accuracy of AUC values.

While the AUC values from empirical prediction studies might be reproducible in simulation studies when the values for the parameters are obtained from the same study, this does not necessarily imply that the AUC can also be estimated a priori based on published values from earlier studies. Differences in study design and population characteristics lead to differences in odds ratios between empirical studies. Here we showed that the AUC estimates based on odds ratios and frequencies from large-scale studies and GWAS fairly reproduced the AUC values of empirical studies, suggesting that differences in the values of the parameters have only limited impact on the accuracy of the AUC.

The level of accuracy in reproducing the AUC values from simulated data varied between the risk models. The AUC values were more accurately reproduced for empirical studies that used logistic regression models than for studies that investigated weighted or unweighted risk scores. This finding might be explained by differences in calibration. In logistic regression, odds ratios are best estimates of effect sizes in the specific population under study, whereas weighted and unweighted risk scores assume external weights (in unweighted all weights are set equal to 1) which may not apply in the particular study population. The simulation method that we used assumes perfect calibration, which is more similar to logistic regression than to the risk scores.

Simulation methods that investigate the predictive ability of genetic risk models can be used to obtain the expected predictive performance of genetic variants before conducting an empirical study. This information is helpful to set research priorities and to allocate resources to the most promising applications of genome-based knowledge. Simulation studies can also be used to assess the predictive ability when empirical data collection is not an option. We recently investigated the predictive ability of personal genome tests offered by three companies to consumers directly via the internet. Empirical data do not exist and their collection is time consuming and expensive. Using simulation studies we showed for six diseases that the predictive ability on the population level was similar between the companies, but that for individual consumers differences in predicted risks were substantial (Chapter 9, this thesis).

In conclusion, the predictive ability of genetic risk models can be estimated fairly accurately using data about odds ratios and genotype frequencies from GWAS and meta-analyses. Simulation methods are a cost and time efficient tool to determine the expected predictive ability of genetic risk models. For a responsible use of research resources, discussion is needed on what level of predictive ability is required for application in different health care scenarios. Empirical assessment of prediction models is only warranted when that level of predictive ability is anticipated.

REFERENCES

1. Willems SM, Mihaescu R, Sijbrands EJ, van Duijn CM and Janssens AC. (2011) **A methodological perspective on genetic risk prediction studies in type 2 diabetes: recommendations for future research.** *Curr Diab Rep,* **11;**511-8.
2. Husing A, Canzian F, Beckmann L, Garcia-Closas M, Diver WR, Thun MJ, Berg CD, Hoover RN, Ziegler, RG, Figueroa JD, et al. (2012) **Prediction of breast cancer risk by genetic risk factors, overall and by hormone receptor status.** *J Med Genet,* **49;**601-8.
3. Romanos J, van Diemen CC, Nolte IM, Trynka G, Zhernakova A, Fu J, Bardella MT, Barisani D, McManus R, van Heel DA, et al. (2009) **Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease.** *Gastroenterology,* **137;**834-40, 840 e1-3.
4. Maller J, George S, Purcell S, Fagerness J, Altshuler D, Daly MJ and Seddon JM, (2006) **Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration.** *Nat Genet,* **38;**1055-9.
5. Wu C, Walsh KM, Dewan AT, Hoh J and Wang Z, (2011) **Disease risk prediction with rare and common variants.** *BMC Proc,* **5 Suppl 9,** S61.
6. Pepe MS, Gu JW and Morris DE. (2010) **The potential of genes and other markers to inform about risk.** *Cancer Epidemiol Biomarkers Prev,* **19;**655-65.
7. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW and van Duijn CM. (2006) **Predictive testing for complex diseases using multiple genes: fact or fiction?** *Genet Med,* **8;**395-400.
8. Lu Q and Elston RC. (2008) **Using the optimal receiver operating characteristic curve to design a predictive genetic test, exemplified with type 2 diabetes.** *Am J Hum Genet,* **82;**641-51.
9. Moonesinghe R, Liu T and Khoury MJ. (2010) **Evaluation of the discriminative accuracy of genomic profiling in the prediction of common complex diseases.** *Eur J Hum Genet,* **18;**485-9.
10. Gail MH. (2008) **Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk.** *J Natl Cancer Inst,* **100;**1037-41.
11. Kundu S, Karssen LC and Janssens AC. (2012) **Analytical and simulation methods for estimating the potential predictive ability of genetic profiling: a comparison of methods and results.** *Eur J Hum Genet*, **20**:1270-4.
12. Cook NR. (2007) **Use and misuse of the receiver operating characteristic curve in risk prediction.** *Circulation,* **115;**928-35.
13. Genomes Project, C. (2010) **A map of human genome variation from population-scale sequencing.** *Nature,* **467;**1061-73.
14. Kundu S, Aulchenko YS, van Duijn CM and Janssens AC. (2011) **PredictABEL: an R package for the assessment of risk prediction models.** *Eur J Epidemiol,* **26;**261-4.
15. Lango H, Consortium UKTDG, Palmer CN, Morris AD, Zeggini E, Hattersley AT, McCarthy MI, Frayling TM and Weedon MN. (2008) **Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk.** *Diabetes,* **57;**3129-35.
16. van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, Hofman A, Sijbrands EJ and Janssens AC. (2008) **Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study.** *Diabetes,* **57;**3122-8.
17. Moonesinghe R, Khoury MJ, Liu T and Janssens AC. (2011) **Discriminative accuracy of genomic profiling comparing multiplicative and additive risk models.** *Eur J Hum Genet,* **19;**180-5.

18. Peter I, Mitchell AA, Ozelius L, Erazo M, Hu J, Doheny D, Abreu MT, Present DH, Ullman T, Benkov K. et al. (2011) **Evaluation of 22 genetic variants with Crohn's disease risk in the Ashkenazi Jewish population: a case-control study.** *BMC Med Genet,* **12;**63, doi:10.1186/1471-2350-12-63.

19. Helfand BT, Fought AJ, Loeb S, Meeks JJ, Kan D and Catalona WJ. (2010) **Genetic prostate cancer risk assessment: common variants in 9 genomic regions are associated with cumulative risk.** *J Urol,* **184;**501-5.

20. Johansson M, Holmstrom B, Hinchliffe SR, Bergh A, Stenman UH, Hallmans G, Wiklund F and Stattin, P. (2012) **Combining 33 genetic variants with prostate-specific antigen for prediction of prostate cancer: longitudinal study.** *Int J Cancer,* **130;**129-37.

21. Yamashita H, Awata T, Kawasaki E, Ikegami H, Tanaka S, Maruyama T, Shimada A, Nakanishi K, Takahashi K, Kobayashi T. et al. (2011) **Analysis of the HLA and non-HLA susceptibility loci in Japanese type 1 diabetes.** *Diabetes Metab Res Rev,* **27;**844-848.

22. Lin X, Song K, Lim N, Yuan X, Johnson T, Abderrahmani A, Vollenweider P, Stirnadel H, Sundseth SS, Lai E. et al. (2009) **Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score--the CoLaus Study.** *Diabetologia,* **52;**600-8.

23. Qi Q, Li H, Wu Y, Liu C, Wu H, Yu Z, Qi L, Hu FB, Loos RJ and Lin X. (2010) **Combined effects of 17 common genetic variants on type 2 diabetes risk in a Han Chinese population.** *Diabetologia,* **53;**2163-6.

24. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB Sr. et al. (2008) **Genotype score in addition to common risk factors for prediction of type 2 diabetes.** *N Engl J Med,* **359;**2208-19.

25. Wang J, Stancakova A, Kuusisto J and Laakso M. (2010) **Identification of undiagnosed type 2 diabetic individuals by the Finnish diabetes risk score and biochemical and genetic markers: a population-based study of 7232 Finnish men.** *J Clin Endocrinol Metab,* **95;**3858-62.

26. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, Kivimaki M and Humphries SE. (2010) **Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study.** *BMJ,* **340;**b4838.

27. Sun J, Kader AK, Hsu FC, Kim ST, Zhu Y, Turner AR, Jin T, Zhang Z, Adolfsson J, Wiklund F. et al. (2011) **Inherited genetic markers discovered to date are able to identify a significant number of men at considerably elevated risk for prostate cancer.** *Prostate,* **71;**421-30.

28. Kader AK, Sun J, Reck BH, Newcombe PJ, Kim ST, Hsu FC, D'Agostino RB Jr, Tao S, Zhang Z, Turner AR. et al. (2012) **Potential Impact of Adding Genetic Markers to Clinical Parameters in Predicting Prostate Biopsy Outcomes in Men Following an Initial Negative Biopsy: Findings from the REDUCE Trial.** *Eur Urol*, **62;**953-61.

29. Scholl HP, Charbel Issa P, Walier M, Janzer S, Pollok-Kopp B, Borncke F, Fritsche LG, Chong NV, Fimmers R, Wienker T et al. (2008) **Systemic complement activation in age-related macular degeneration.** *PLoS One,* **3;**e2593.

30. Hecker LA, Edwards AO, Ryu E, Tosakulwong N, Baratz KH, Brown WL, Charbel Issa P, Scholl HP, Pollok-Kopp B, Schmid-Kubista KE et al. (2010) **Genetic control of the alternative pathway of complement in humans and age-related macular degeneration.** *Hum Mol Genet,* **19;**209-15.

31. Grassmann F, Fritsche LG, Keilhauer CN, Heid IM and Weber BH. (2012) **Modelling the genetic risk in age-related macular degeneration.** *PLoS One,* **7;**e37979.

32. Dunlop MG, Tenesa A, Farrington SM, Ballereau S, Brewster DH, Kossler T, Pharoah P, Schafmayer C, Hampe J, Volzke H et al. (2012) **Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42 103 individuals.** *Gut*, **62;**871-81.

33. Lubbe SJ, Di Bernardo MC, Broderick P, Chandler I and Houlston RS. (2012) **Comprehensive evaluation of the impact of 14 genetic variants on colorectal cancer phenotype and risk.** *Am J Epidemiol,* **175;**1-10.

34. Aly M, Wiklund F, Xu J, Isaacs WB, Eklund M, D'Amato M, Adolfsson J and Gronberg,H. (2011) **Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm-1 cohort study.** *Eur Urol,* **60;** 21-8.

35. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, Rayner NW, Shields B, Owen KR, Hattersley AT et al. (2006) **Combining information from common type 2 diabetes risk polymorphisms improves disease prediction.** *PLoS Med,* **3;**e374.

36. Vaxillaire M, Veslot J, Dina C, Proenca C, Cauchi S, Charpentier G, Tichet J, Fumeron F, Marre M, Meyre D et al. (2008) **Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study.** *Diabetes,* **57;**244-54.

37. Hu C, Zhang R, Wang C, Wang J, Ma X, Lu J, Qin W, Hou X, Wang C, Bao Y et al. (2009) **PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population.** *PLoS One,* **4;**e7643.

38. Miyake K, Yang W, Hara K, Yasuda K, Horikawa Y, Osawa H, Furuta H, Ng MC, Hirota Y, Mori H et al. (2009) **Construction of a prediction model for type 2 diabetes mellitus in the Japanese population based on 11 genes with strong evidence of the association.** *J Hum Genet,* **54;**236-41.

39. Fontaine-Bisson B, Renstrom F, Rolandsson O, Magic Payne F, Hallmans G, Barroso I and Franks PW. (2010) **Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population.** *Diabetologia,* **53;**2155-62.

40. Sparso T, Grarup N, Andreasen C, Albrechtsen A, Holmkvist J, Andersen G, Jorgensen T, Borch-Johnsen K, Sandbaek A, Lauritzen T et al. (2009) **Combined analysis of 19 common validated type 2 diabetes susceptibility gene variants shows moderate discriminative value and no evidence of gene-gene interaction.** *Diabetologia,* **52;**1308-14.

41. Janssens AC and van Duijn CM. (2009) **Genome-based prediction of common diseases: methodological considerations for future research.** *Genome Med,* **1;**20.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Risk allele frequencies and odds ratios of genetic variants considered in the analyses.

A. Type 2 diabetes

SNP	Locus	Gene	Risk allele	Risk allele frequency	OR	Lower CI	Upper CI	First author publication year
rs10923931	1p12	NOTCH2	T	0.11	1.13	1.08	1.17	Zeggini 2008
rs2641348	1p31.1	ADAM30	G	0.107	1.1	1.06	1.15	Zeggini 2008
rs7578597	2p21	THADA	T	0.90	1.15	1.10	1.20	Zeggini 2008
rs10490072	2p16.1	Intergenic (BCL11A)	T	0.72	1.05	1.03	1.08	Zeggini 2008
rs3792267	2q37.3	CAPN10	G	0.23	1.19	1.07	1.33	Song 2004
rs780094	2p23.3	GCKR	G	0.44	1.11	1.01	1.22	Horikawa 2008
rs780094	2p23.3	GCKR	G	0.651	1.09	1.02	1.16	Sparso 2008
rs4607103	3p14.1	ADAMTS9	C	0.76	1.09	1.06	1.12	Zeggini 2008
rs1470579	3q27.2	IGF2BP2	NR	0.33	1.18	1.07	1.31	Horikawa 2008
rs17036101	3p25.2	Intergenic (PPARG)	G	0.93	1.15	1.1	1.21	Zeggini 2008
rs1801282	3p25.2	PPARG	C	0.86	1.14	1.08	1.20	Saxena 2007
rs1801282	3p25.2	PPARG	C	0.84	1.28	0.95	1.69	Altshuler 2000
rs4402960	3q27.2	IGF2BP2	T	0.29*	1.14	1.11	1.17	Ng 2008
rs4402960	3q27.2	IGF2BP2	T	0.29	1.14	1.11	1.18	Saxena 2007
rs4402960	3q27.2	IGF2BP2	T	0.29*	1.17	1.10	1.25	Zeggini 2008
rs10010131	4p16.1	WFS1	G	0.63	1.11	1.08	1.16	Sandhu 2007
rs10010131	4p16.1	WFS1	G	0.63*	1.12	1.09	1.16	Wasson 2008
rs1801214	4p16.1	WFS1	T	0.63*	1.13	1.08	1.18	Voight 2010
rs9472138	6p21.1	VEGFA	T	0.28	1.06	1.04	1.09	Zeggini 2008
rs10946398	6p22.3	CDKAL1	C	0.32*	1.12	1.08	1.16	Zeggini 2007
rs7754840	6p22.3	CDKAL1	G	0.41	1.28	1.17	1.41	Horikawa 2008
rs7754840	6p22.3	CDKAL1	C	0.31	1.12	1.08	1.16	Saxena 2007
rs7756992	6p22.3	CDKAL1	G	0.27*	1.16	1.13	1.19	Ng 2008
rs7756992	6p22.3	CDKAL1	G	0.26	1.20	1.13	1.27	Steinthorsdottir 2007
rs1799884	7p13	GCK	A	0.15	1.28	1.06	1.53	Shaat 2006
rs864745	7p15.1	JAZF1	T	0.50	1.10	1.07	1.13	Zeggini 2008
rs1800795	7p15.3	IL6	G	0.82*	1.10	1.01	1.20	Huth 2006
rs13266634	8q24.11	SLC30A8	C	0.71*	1.16	1.13	1.19	Ng 2008
rs13266634	8q24.11	SLC30A8	C	0.65	1.12	1.07	1.16	Saxena 2007

SNP	Locus	Gene	Risk allele	Risk allele frequency	OR	Lower CI	Upper CI	First author publication year
rs3802177	8q24.11	SLC30A8	NR	0.56	1.16	1.05	1.27	Horikawa 2008
rs10811161	9p21.3	Intergenic (CDKN2A/B)	T	0.84*	1.22	1.18	1.26	Ng 2008
rs10811661	9p21.3	Intergenic (CDKN2A/B)	T	0.55	1.27	1.15	1.40	Horikawa 2008
rs10811661	9p21.3	Intergenic (CDKN2A/B)	T	0.83	1.20	1.14	1.25	Saxena 2007
rs10811661	9p21.3	Intergenic (CDKN2A/B)	T	0.84*	1.2	1.14	1.25	Zeggini 2007
rs564398	9p21.3	Intergenic (CDKN2A/B)	T	0.59*	1.12	1.07	1.17	Zeggini 2007
rs12779790	10p13	Intergenic (CDC123)	G	0.18	1.11	1.07	1.14	Zeggini 2008
rs7901695	10q25.2	TCF7L2	C	0.32*	1.37	1.31	1.43	Zeggini 2007
rs7903146	10q25.2	TCF7L2	T	0.31*	1.46	1.42	1.51	Cauchi 2007
rs7903146	10q25.2	TCF7L2	T	0.31*	1.54	1.39	1.70	Grant 2006
rs7903146	10q25.2	TCF7L2	T	0.30	1.47	1.33	1.62	Helgason 2007
rs7903146	10q25.2	TCF7L2	T	0.04	1.48	1.20	1.84	Miyake 2008
rs7903146	10q25.2	TCF7L2	T	0.26	1.37	1.31	1.43	Saxena 2007
rs7903146	10q25.2	TCF7L2	T	0.31*	1.37	1.28	1.47	Zeggini 2008
rs1111875	10q23.33	HHEX	C	0.53	1.13	1.08	1.17	Saxena 2007
rs5015480	10q23.33	Intergenic (HHEX)	C	0.55*	1.17	1.11	1.24	Zeggini 2008
rs7923837	10q23.33	Intergenic (HHEX)	G	0.19	1.27	1.13	1.43	Horikawa 2008
rs7923837	10q23.33	Intergenic (HHEX)	G	0.61*	1.23	1.18	1.29	Ng 2008
rs7480010	11p12	Intergenic (LOC387761)	G	0.30	1.17	1.08	1.27	Sladek 2007
rs5215	11p15.1	KCNJ11	C	0.34*	1.16	1.09	1.23	Zeggini 2008
rs5219	11p15.1	KCNJ11	T	0.36	1.18	1.04	1.34	Gloyn 2003
rs5219	11p15.1	KCNJ11	T	0.47	1.14	1.10	1.19	Saxena 2007
rs11037909	11p11.2	EXT2	T	0.74	1.18	1.08	1.29	Sladek 2007
rs10830963	11q14.3	MTNR1B	G	0.28	1.09	1.05	1.12	Prokopenko 2008
rs1387153	11q14.3	Intergenic (MTNR1B)	T	0.28	1.09	1.06	1.11	Voight 2010
rs2237892	11p15.4	KCNJ11	C	0.65	1.53	1.28	1.70	Hu 2009
rs2237892	11p15.4	KCNJ11	C	0.61	1.42	1.34	1.49	Yasuda 2008
rs2237895	11p15.4	KCNJ11	C	0.40	1.19	1.14	1.25	Unoki 2008
rs2283228	11p15.4	KCNJ11	A	0.58	1.26	1.18	1.34	Unoki 2008
rs689	11p15.5	INS	T	0.72*	1.16	0.95	1.41	Meigs 2008
rs7961581	12q21.1	TSPAN8 (TSPAN8, LGR5)	C	0.27	1.09	1.06	1.12	Zeggini 2008

SNP	Locus	Gene	Risk allele	Risk allele frequency	OR	Lower CI	Upper CI	First author publication year
rs1153188	12q13.2	Intergenic (DCD)	A	0.73	1.08	1.05	1.11	Zeggini 2008
rs1800574	12q24.31	HNF1A	T	0.03	1.31	1.08	1.57	Weedon 2005
rs8050136	16q12.2	FTO	A	0.42*	1.11	1.08	1.15	Ng 2008
rs8050136	16q12.2	FTO	A	0.38	1.17	1.12	1.22	Scott 2007
rs8050136	16q12.2	FTO	A	0.42*	1.17	1.12	1.22	Zeggini 2007
rs8050136	16q12.2	FTO	A	0.42*	1.15	1.09	1.22	Zeggini 2008
rs8050136	16q12.2	FTO	A	0.38	1.17	1.12	1.22	Saxena 2007
rs4430796	17q12	HNF1B	A	0.49	1.22	1.15	1.30	Gudmundsson 2007
rs4430796	17q12	HNF1B	G	0.47*	1.14	1.08	1.20	Voight 2008
rs4430796	17q12	HNF1B	G	0.28	1.16	1.05	1.29	Wang 2009
rs7501939	17q12	HNF1B (TCF2)	C	0.58	1.19	1.12	1.26	Gudmundsson 2007
rs757210	17q12	HNF1B (TCF2)	A	0.43	1.12	1.07	1.18	Winckler 2007
rs1056137	20p12.3	FERMTL (KCNJ11)	NR	0.36	1.18	1.04	1.34	Gloyn 2003

*Allele frequency is estimated from 1000 genome project

B. Age-related macular degeneration

SNP	Locus	Gene	Risk allele	Allele frequency	OR	Lower CI	Upper CI	First author year
rs1329424	1q31.3	CFH	T	0.35	1.88	1.68	2.10	Chen 2010
rs2274700	1q31.3	CFH (HF1)	G	0.68	2.39	1.98	2.89	Hageman 2005
rs380390	1q31.3	CFH	C	0.70	4.60†	2.00	11.00	Klein 2005
rs6677604	1q31.3	CFH	G	0.86	2.85	1.77	4.58	Hughes 2007
rs800292	1q31.3	CFH (HF1)	G	0.37	1.13	0.96	1.34	Hageman 2005
rs2285714	4q25	PLA2G12A (CFI)	T	0.40	1.31	1.18	1.45	Chen 2010
rs4151667	6p21.33	CFB	T	0.95*	2.78	1.79	4.35	Gold 2006
rs547154	6p21.33	C2	G	0.91*	2.27	1.67	3.03	Gold 2006
rs10490924	10q26.13	ARMS2 (ARMS2–HTRA1)	T	0.21	3.00	2.64	3.40	Cipriani 2012
rs10490924	10q26.13	ARMS2 (LOC387715)	T	0.33	2.97	2.58	3.43	Rivera 2005
rs10468017	15q21.3	ALDH1A2 (LIPC)	C	0.70	1.22	1.14	1.30	Neale 2010
rs493258	15q21.3	ALDH1A2 (LIPC)	C	0.56	1.14	1.09	1.20	Chen 2010
rs2230199	19p13.3	C3	C	0.16	1.74	1.47	2.06	Chen 2010
rs2230199	19p13.3	C3	G	0.26	1.66	1.44	1.92	Maller 2007

SNP	Locus	Gene	Risk allele	Allele frequency	OR	Lower CI	Upper CI	First author year
rs429358	19q13.32	APOE	C	0.89	1.28	1.01	1.64	Fritsche 2009
rs7412	19q13.32	APOE	C	0.09	1.55	1.17	2.07	Fritsche 2009
rs9621532	22q12.3	SYN3/TIMP3	A	0.95	1.41	1.27	1.57	Chen 2010

*Allele frequency is estimated from 1000 genome project; †Dominant effects of the risk allele, which is used in the simulation method.

C. Colorectal cancer

SNP	Locus	Gene	Risk allele	Allele frequency	OR	Lower CI	Upper CI	First author year
rs6691170	1q41	Intergenic (DUSP10)	T	0.40*	1.06	1.03	1.09	Houlston 2010
rs10936599	3q26.2	MYNN	C	0.75*	1.08	1.04	1.10	Houlston 2010
rs6983267	8q24.21	Intergenic (-)	G	0.48*	1.21	1.15	1.27	Tomlinson 2007
rs16892766	8q23.3	Intergenic (EIF3H)	A	0.07	1.25	1.19	1.32	Tomlinson 2008
rs10795668	10p14	Intergenic (-)	A	0.67	1.12	1.10	1.16	Tomlinson 2008
rs3802842	11q23.1	C11orf93 (-)	C	0.43	1.11	1.08	1.15	Tenesa 2008
rs11169552	12q13.12	ATF1 (DIP2B, ATF1)	C	0.75*	1.09	1.05	1.11	Houlston 2010
rs4444235	14q22.2	BMP4	C	0.46	1.11	1.08	1.15	Houlston 2008
rs4779584	15q13.3	Intergenic (GREM1/ SCG5)	T	0.76	1.54	1.46	1.62	Jaeger 2008
rs9929218	16q22.1	CDH1	G	0.71	1.10	1.06	1.12	Houlston 2008
rs4939827	18q21.1	SMAD7	T	0.48	1.18	1.12	1.23	Broderick 2007
rs4939827	18q21.1	SMAD7	T	0.52	1.20	1.16	1.24	Tenesa 2008
rs10411210	19q13.11	RHPN2	C	0.90	1.15	1.10	1.20	Houlston 2008
rs961253	20p12.3	Intergenic (-)	A	0.36	1.12	1.08	1.16	Houlston 2008
rs4925386	20q13.33	LAMA5	C	0.68*	1.08	1.05	1.10	Houlston 2010

* Allele frequency is estimated from 1000 genome project.

D. Crohn's disease

SNP	Locus	Gene	Risk allele	Risk allele frequency	OR	Lower CI	Upper CI	First author year
rs11209026	1p31.3	IL23R	G	0.93	2.22	1.37	3.70	Duerr 2006
rs7517847	1p31.3	IL23R	T	0.65	1.72	1.39	2.13	Duerr 2006
rs1373692	5p13.1	Intergenic (-)	?	0.64	1.45	1.26	1.68	Libioulle 2007
rs13361189	5q33.1	IRGM	C	0.08	1.38	1.15	1.66	Parkes 2007
rs2066847	16q12.1	NOD2	C	0.03	2.45	1.51	3.98	Economou 2004
rs2066844	16q12.1	NOD2	T	0.04	1.74	0.88	3.42	Economou 2004
rs2066845	16q12.1	NOD2	C	0.05	1.93	1.23	3.00	Economou 2004

E. Prostate cancer

SNP	Locus	Gene	Risk allele	Allele frequency	OR	Lower CI	Upper CI	First author year
rs2710646	2p15	EHBP1	A	0.20	1.16	1.06	1.26	Gudmundsson 2008
rs2710647	2p15	EHBP1	C	0.56	1.11	0.95	1.30	Johansson 2012
rs721048	2p15	EHBP1	A	0.19	1.15	1.10	1.21	Gudmundsson 2008
rs721048	2p15	EHBP1	A	0.14	1.18	1.10	1.28	Kim 2010
rs1465618	2p21	THADA	A	0.23	1.08	1.03	1.12	Eeles 2009
rs1465618	2p21	THADA	A	0.21	1.15	1.04	1.26	Kim 2010
rs12621278	2q31.1	ITGA6	A	0.94	1.33	1.25	1.43	Eeles 2009
rs12621278	2q31.1	ITGA6	A	0.96	1.35	1.27	1.44	Kim 2010
rs2660753	3p12.1	Intergenic (-)	T	0.11	1.19	1.09	1.31	Eeles 2008
rs2660753	3p12.1	Intergenic (-)	T	0.10	1.24	1.04	1.48	Kim 2010
rs10934853	3q21.3	EEFSEC (-)	A	0.28	1.12	1.08	1.16	Gudmundsson 2009
rs10934853	3q21.3	EEFSEC (-)	A	0.24	1.12	1.06	1.18	Kim 2010
rs4857841	3q21.3	EEFSEC (-)	A	0.28	1.12	1.08	1.16	Gudmundsson 2009
rs7679673	4q24	TET2	C	0.55	1.10	1.06	1.14	Eeles 2009
rs7679673	4q24	TET2	C	0.62	1.14	1.09	1.20	Kim 2010
rs12500426	4q22.3	PDLIM5	A	0.46	1.08	1.05	1.12	Eeles 2009
rs17021918	4q22.3	PDLIM5	C	0.66	1.11	1.08	1.15	Eeles 2009
rs17021918	4q22.3	PDLIM5	C	0.65	1.14	1.10	1.18	Kim 2010
rs9364554	6q25.3	SLC22A3	T	0.31	1.21	1.14	1.28	Eeles 2008
rs9364554	6q25.3	SLC22A3	T	0.27	1.17	1.06	1.29	Kim 2010
rs10486567	7p15.2	JAZF1	G	0.77	1.11	1.09	1.13	Thomas 2008
rs12155172	7p15.3	Intergenic (-)	A	0.20	1.05	1.00	1.10	Eeles 2009
rs6465657	7q21.3	LMTK2	C	0.48	1.18	1.12	1.25	Eeles 2008

SNP	Locus	Gene	Risk allele	Allele frequency	OR	Lower CI	Upper CI	First author year
rs6465657	7q21.3	LMTK2	C	0.51	1.14	1.05	1.23	Kim 2010
rs1512268	8p21.2	Intergenic (NKX3.1)	A	0.45	1.18	1.14	1.22	Eeles 2009
rs1512268	8p21.2	Intergenic (NKX3.1)	T	0.42	1.17	1.12	1.23	Kim 2010
rs2928679	8P21.2	Intergenic (NKX3.1)	T	0.42	1.05	1.01	1.09	Eeles 2009
rs2928679	8P21.2	Intergenic (NKX3.1)	A	0.46	1.13	1.02	1.25	Kim 2010
rs10086908	8q24.21	Intergenic (-)	C	0.72	1.15	1.08	1.22	Al Olama 2009
rs10086908	8q24.21	Intergenic (-)	T	0.63	1.13	1.08	1.19	Kim 2010
rs1016343	8q24.21	Intergenic (-)	T	0.22	1.26	1.18	1.34	Al Olama 2009
rs1016343	8q24.21	Intergenic (-)	T	0.21	1.39	1.24	1.55	Eeles 2008
rs12543663	8q24.21	Intergenic (-)	C	0.32	1.13	1.07	1.20	Al Olama 2009
rs13252298	8q24.21	Intergenic (-)	G	0.72	1.19	1.13	1.26	Al Olama 2009
rs1447295	8q24.21	Intergenic (-)	A	0.09	1.60	1.43	1.77	Gudmundsson 2007
rs1447295	8q24.21	Intergenic (-)	A	0.11	1.58	1.43	1.74	Gudmundsson 2009
rs1447295	8q24.21	Intergenic (-)	A	0.07	1.47	1.33	1.62	Kim 2010
rs16901979	8q24.21	Intergenic (-)	A	0.03	1.79	1.53	2.11	Gudmundsson 2007
rs16901979	8q24.21	Intergenic (-)	A	0.04	1.80	1.55	2.09	Gudmundsson 2009
rs16901979	8q24.21	Intergenic (-)	A	0.03	1.82	1.44	2.30	Kim 2010
rs16902094	8q24.21	Intergenic (-)	G	0.15	1.21	1.15	1.26	Gudmundsson 2009
rs16902094	8q24.21	Intergenic (-)	G	0.27	1.20	1.12	1.30	Kim 2010
rs4242382	8q24.21	Intergenic (-)	A	0.12	1.46	1.42	1.50	Thomas 2008
rs445114	8q24.21	Intergenic (-)	T	0.64	1.14	1.10	1.19	Gudmundsson 2009
rs620861	8q24.21	Intergenic (-)	T	0.64	1.17	1.11	1.23	Al Olama 2009
rs620861	8q24.21	Intergenic (-)	G	0.62	1.16	1.11	1.20	Kim 2010
rs6983267	8q24.21	Intergenic (-)	G	0.54	1.31	1.24	1.38	Al Olama 2009
rs6983267	8q24.21	Intergenic (-)	G	0.53	1.41	1.29	1.54	Eeles 2008
rs6983267	8q24.21	Intergenic (-)	G	0.49	1.20	1.14	1.26	Kim 2010
rs6983267	8q24.21	Intergenic (-)	G	0.53	1.24	1.22	1.27	Thomas 2008
rs6983267	8q24.21	Intergenic (-)	G	0.53	1.26	1.19	1.34	Yaeger 2007
rs6983561	8q24.21	Intergenic (-)	C	0.04	1.66	1.43	1.92	Al Olama 2009
rs4961199	8q21.3	CNGB3 (CPNE3)	A	0.16	1.17	1.15	1.20	Thomas 2008
rs1571801	9q33.2	DAB2IP	A	0.25	1.28	1.17	1.40	Duggan 2007
rs4962416	10q26.13	CTBP2	C	0.26	1.15	1.04	1.27	Kim 2010
rs4962416	10q26.13	CTBP2	C	0.27	1.20	1.17	1.22	Thomas 2008
rs10993994	10q11.23	MSMB	T	0.42	1.37	1.30	1.45	Eeles 2008

SNP	Locus	Gene	Risk allele	Allele frequency	OR	Lower CI	Upper CI	First author year
rs10993994	10q11.23	MSMB	T	0.34	1.25	1.12	1.40	Kim 2010
rs10993994	10q11.23	MSMB	T	0.41	1.28	1.25	1.31	Lou 2009
rs10993994	10q11.23	MSMB	T	0.40	1.26	1.24	1.28	Thomas 2008
rs10896449	11q13.3	Intergenic (-)	G	0.53	1.28	1.16	1.40	Eeles 2008
rs10896449	11q13.3	Intergenic (-)	G	0.53	1.16	1.11	1.22	Kim 2010
rs10896449	11q13.3	Intergenic (-)	G	0.52	1.22	1.19	1.24	Thomas 2008
rs10896450	11q13.3	Intergenic (-)	G	0.47	1.13	1.06	1.21	Gudmundsson 2009
rs11228565	11q13.3	Intergenic (-)	A	0.20	1.23	1.16	1.31	Gudmundsson 2009
rs12418451	11q13.3	Intergenic (-)	A	0.30	1.16	1.09	1.23	Zheng 2009
rs7127900	11p15.5	Intergenic (IGF2, IGF2AS,INS,TH)	A	0.24	1.25	1.20	1.30	Kim 2010
rs7127900	11p15.5	Intergenic (-)	A	0.20	1.22	1.17	1.27	Eeles 2009
rs11649743	17q12	HNF1B	G	0.76	1.16	1.11	1.22	Kim 2010
rs11649743	17q12	HNF1B	G	0.82	1.18	1.11	1.24	Sun 2008
rs4430796	17q12	HNF1B	A	0.49	1.22	1.15	1.30	Gudmundsson 2007 (2)
rs4430796	17q12	HNF1B	A	0.49	1.22	1.17	1.26	Kim 2010
rs7501939	17q12	HNF1B	T	0.61	1.41	1.29	1.55	Eeles 2008
rs1859962	17q24.3	Intergenic (-)	G	0.49	1.27	1.16	1.40	Eeles 2008
rs1859962	17q24.3	Intergenic (-)	G	0.46	1.20	1.14	1.27	Gudmundsson 2007 (2)
rs1859962	17q24.3	Intergenic (-)	G	0.53	1.21	1.12	1.30	Kim 2010
rs887391	19a13.2	Intergenic (-)	T	0.76	1.14	1.08	1.20	Kim 2010
rs8102476	19q13.2	Intergenic (-)	C	0.54	1.12	1.08	1.15	Gudmundsson 2009
rs8102476	19q13.2	Intergenic (-)	C	0.50	1.12	1.08	1.15	Kim 2010
rs266849	19q13.33	Intergenic (-)	G	0.80	1.24	1.16	1.33	Eeles 2008
rs2735839	19q13.33	Intergenic (KLK3)	A	0.85	1.42	1.31	1.53	Eeles 2008
rs2735839	19q13.33	Intergenic (KLK3)	G	0.86	1.30	1.11	1.51	Kim 2010
rs9623117	22q13.1	TNRC6B	C	0.22	1.13	1.05	1.22	Kim 2010
rs9623117	22q13.1	TNRC6B	C	0.21	1.18	1.11	1.26	Sun 2009
rs5759167	22q13.2	Intergenic (-)	G	0.53	1.16	1.14	1.20	Eeles 2009
rs5759167	22q13.2	Intergenic (-)	G	0.55	1.18	1.14	1.21	Kim 2010
rs5945572	Xp11.22	Intergenic (-)	A	0.35	1.23	1.16	1.30	Gudmundsson 2008
rs5945619	Xp11.22	NUDT11 (NUDT10, NUDT11)	C	0.39	1.27	1.12	1.43	Kim 2010
rs5945619	Xp11.22	NUDT11 (-)	C	0.38	1.28	1.21	1.35	Eeles 2008

F. Type 1 diabetes

SNP	Locus	Gene	Risk allele	Allele frequency	OR	Lower CI	Upper CI	First author year
rs1990760	2q24.2	IFIH1	A	0.61	1.18	1.06	1.19	Todd 2007
rs3087243	2q33.2	CTLA4	G	0.75	1.04	0.88	1.23	Ikegami 2006
rs6897932	5p13.2	IL7R	G	0.73	1.12	1.06	1.19	Todd 2007
rs706778	10p15.1	IL2RA	A	0.54	1.23	1.06	1.43	Kawasaki 2009
rs689	11p15.5	INS	A	0.95*	3.29	1.68	6.43	Awata 2009
rs2292239	12q13.2	ERBB3	A	0.27	1.38	1.16	1.65	Awata 2009
rs2903692	16p13.13	CLEC16A	G	0.85	1.28	1.05	1.57	Awata 2009

*Allele frequency is estimated from 1000 genome project.

Locus and gene names were obtained from the Ensembl database, release 68. When gene names were different in the paper, the latter are mentioned between square brackets. A dash between the square brackets means that the gene name was not reported in the cited publication. All risk allele frequencies and odds ratios are from the cited study, unless otherwise indicated. *Allele frequency obtained from 1000 genomes project.

Supplementary Table 2. Risk allele frequencies of 18 single nucleotide polymorphisms in two prediction studies on type 2 diabetes and their corresponding values in the cited genome-wide association studies.

Gene	SNP	Frequency in prediction study		Frequency in cited GWAS*	
		GoDARTS study[15]	Rotterdam study[16]	GoDARTS study	Rotterdam study
ADAM30/NOTCH2	rs2641348†	0.11	0.12	0.11	
ADAMTS9	rs4607103‡‡	0.77	0.76	0.76	
CDC123	rs12779790¥	0.20	0.18	0.18	
CDKAL1	rs10946398§	0.34	0.31	0.31	
CDKN2A/2B	rs10811661	0.85	0.81	0.83	
CDKN2A/2B	rs564398‡	0.59	0.57	0.59	
FTO	rs8050136	0.41	0.37	0.42	
HHEX-IDE	rs1111875	0.62	0.59	0.53	
IGF2BP2	rs4402960	0.33	0.30	0.29	
JAZF1	rs864745§§	0.50	0.52	0.50	
KCNJ11	rs5219	0.36	0.37	0.36	0.47
PPARG	rs1801282	0.87	0.88	0.86	
SLC30A8	rs13266634	0.70	0.70	0.65	
TCF2	rs757210††	0.37	0.49	0.43	0.49
TCF7L2	rs7903146	0.30	0.29	0.30	0.31
THADA	rs7578597	0.91	0.88	0.90	
TSPAN8/LGR5	rs7961581¶	0.29	0.29	0.27	
WFS1	rs10010131**	0.60	0.60	0.63	

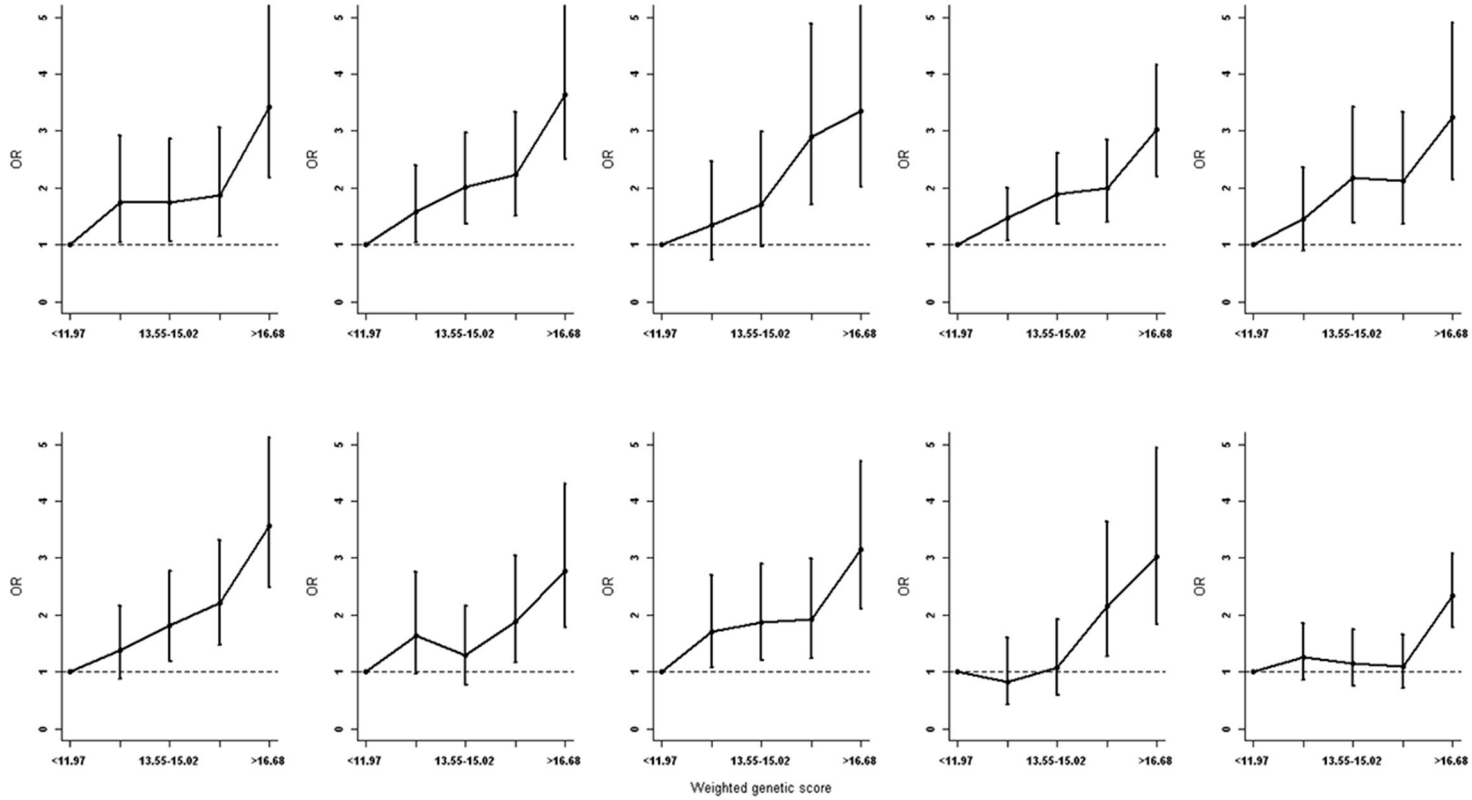
Table is adapted from (41). The risk models of the GoDARTS study and the Rotterdam Study included the same 18 genes and both had an AUC of 0.60. The AUC values from simulated data were 0.62 and 0.61 respectively. For several genes, the GoDARTS study and the Rotterdam Study used different SNPs: †rs1493694, r^2=0.74; ‡rs1412829, r^2=0.97; §rs7754840, r^2=1.00; ¶rs1353362, r^2=0.96; ¥rs11257622; r^2=0.83; **rs10012946, r^2=1.00; ††rs4430796, r^2=0.61; ‡‡rs4411878, r^2=0.95; §§rs1635852, r^2=0.97.

* When only one value is presented, both prediction studies cited the same GWAS.

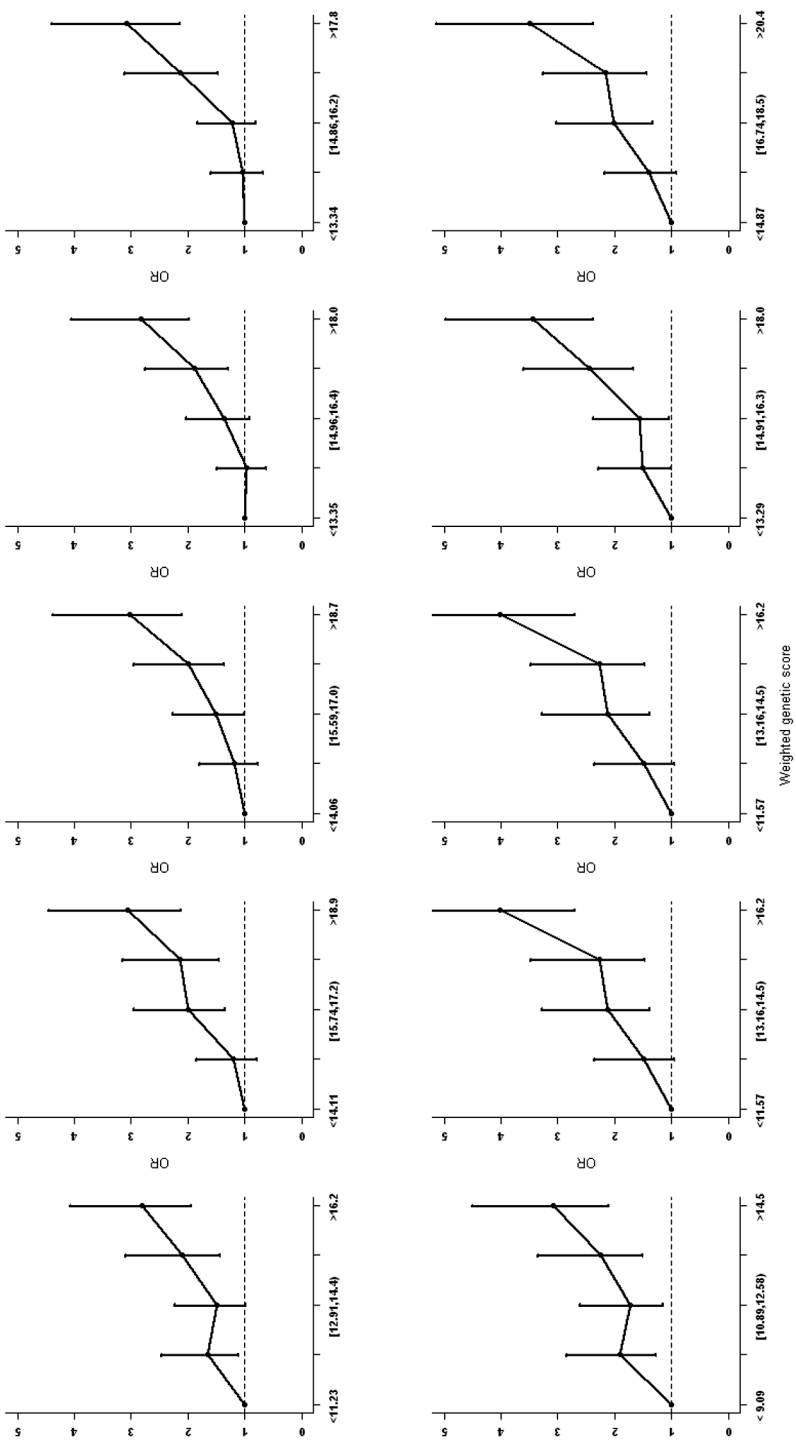
SNP, single nucleotide polymorphism; GWAS, genome-wide association study.

Supplementary Figure 1. Quintiles plots of genetic risk scores in different iterations.

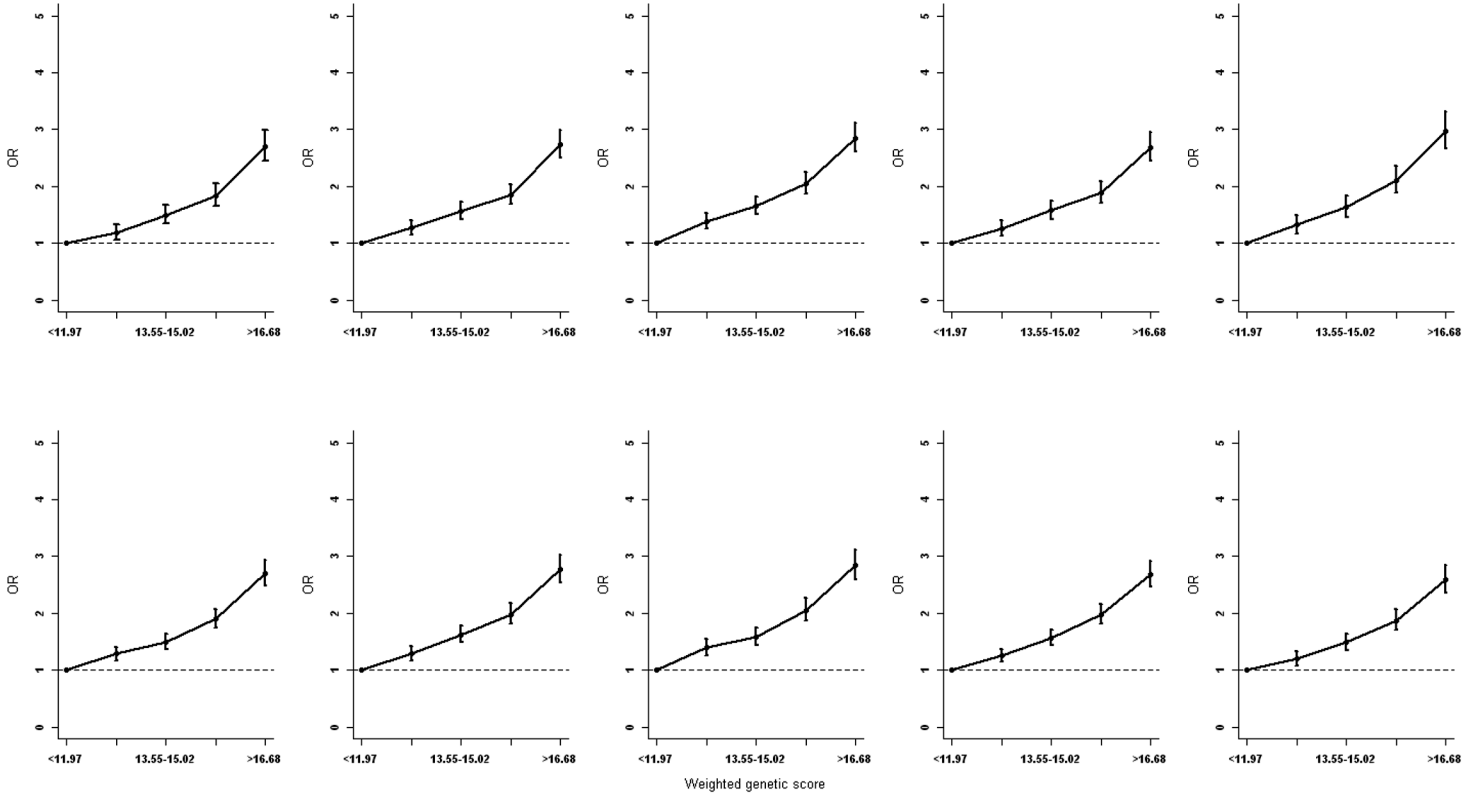
A. Considering the same sample size and risk scores to define the quintiles as reported in the prediction study.



B. Considering the same sample size as reported in the prediction study but quintiles are defined based on estimated risk scores.



C. Considering the same risk scores to define the quintiles as reported in the prediction study, but using a sample size of 100,000.



Quintiles plots presenting the odds ratios (OR) with 95% confidence intervals by quintiles of the weighted genetic scores.

Chapter 5

Incremental value of rare genetic variants for the prediction of multifactorial diseases

Raluca Mihaescu, Michael J. Pencina, Alvaro Alonso, Kathryn L. Lunetta, Susan R. Heckbert, Emelia J. Benjamin, A. Cecile J.W. Janssens

Genome Medicine. 2013;5(8):76.

ABSTRACT

Background: It is often assumed that rare genetic variants will improve available risk prediction scores. We aimed to estimate the added predictive ability of rare variants for risk prediction of common diseases in hypothetical scenarios.

Methods: In simulated data, we constructed risk models with an area under the ROC curve (AUC) ranging between 0.50 and 0.95, to which we added a single variant representing the cumulative frequency and effect (odds ratio, OR) of multiple rare variants. The frequency of the rare variant ranged between 0.0001 and 0.01 and the OR between 2 and 10. We assessed the resulting AUC, increment in AUC, integrated discrimination improvement (IDI), net reclassification improvement (NRI(>0.01)) and categorical NRI. The analyses were illustrated by a simulation of atrial fibrillation risk prediction based on a published clinical risk model.

Results: We observed minimal improvement in AUC with the addition of rare variants. All measures increased with the frequency and OR of the variant, but maximum increment in AUC remained below 0.05. Increment in AUC and NRI(>0.01) decreased with higher AUC of the baseline model, whereas IDI remained constant. In the atrial fibrillation example, the maximum increment in AUC was 0.02 for a variant with frequency=0.01 and OR=10. IDI and NRI showed at most minimal increase for variants with frequency greater than or equal to 0.005 and OR greater than or equal to 5.

Conclusions: Since rare variants are present in only a minority of affected individuals, their predictive ability is generally low at the population level. To improve the predictive ability of clinical risk models for complex diseases, genetic variants must be common and have substantial effect on disease risk.

BACKGROUND

Genome-wide association studies (GWASs) have uncovered an incredible number of common susceptibility variants, but they explain only a small part of the heritability of complex diseases (1). In search for the missing heritability, genetic research is investigating common variants with weak effects on disease risk, gene-gene interactions, structural variations and rare variants (2). With the introduction of next generation sequencing, much effort is currently directed towards rare variants. Expected to have a predominant effect on protein structure, rare variants are more likely to be functional and to display strong effects on disease risk (3-6). Sequencing of coding regions of the genome already has proved successful in identifying rare polymorphisms associated with common traits and complex diseases (7-10).

The predictive ability of rare variants and their potential to improve clinical risk models are uncertain for the population at large, as they are present in only a minority of the affected individuals. The predictive ability of rare variants in common diseases is understudied. Two methodological papers investigated the increment in area under the receiver operating characteristic curve (AUC) when rare variants were added to models based on common variants using simulated data (11, 12). They showed that the maximum increment in AUC was 0.06, but they did not provide the effect sizes of the rare variants, which makes it difficult to interpret the significance of their results. Additionally, AUC is considered an insensitive measure to detect potentially clinically important improvement in prediction (13-15). Two new metrics were developed and rapidly gained popularity: the integrated discrimination improvement (IDI) and the net reclassification improvement (NRI) (16). These metrics may be able to detect clinically significant improvement in prediction due to rare variants that the AUC fails to uncover.

We investigated the value of rare genetic variants for risk prediction of complex diseases. We examined the relation between the predictive ability of rare variants and their frequency, strength of effect (OR) and the predictive ability of the baseline risk model. We assessed the improvement in model performance by delta AUC (ΔAUC), IDI and NRI. To this end, we simulated a large dataset and constructed risk models based on common variants for increasing values of the baseline AUC. In separate scenarios, we added rare genetic variants with varying odds ratios (OR) and frequencies. We further used hypothetical data that replicated the empirical populations used to derive a recently published clinical model for atrial fibrillation (AF) (17). This common cardiac arrhythmia is associated with increased morbidity, mortality and significant health-care costs (18). Numerous common genetic variants associated with atrial fibrillation risk have been identified (19-22) and rare genetic variants are expected to improve the detection of at-risk individuals (23, 24).

METHODS

Simulation of data

First, we used a simulation procedure to investigate the effect of the predictive accuracy of the baseline model on the discrimination of the model updated with rare variants. The modeling procedure has been described in detail by Janssens et al. (25). In short, this procedure creates a dataset of genotypes for a hypothetical population. Genotypes, coded as 0, 1, or 2 based on the number of risk alleles, are assigned in such a way that the allele frequencies of the genetic variants match specified values and are in Hardy-Weinberg equilibrium. By changing the number, frequency and ORs of simulated variants we created baseline models with an AUC ranging between 0.50 and 0.95. We added rare genetic variants to the simulated dataset of common variants. Rare variants were simulated as a single variant representing multiple rarer variants. That is, for example, 20 independent rare variants each with a frequency of 1 in 2000 individuals can collectively be viewed as a single variant with a frequency of 0.01. The variant was coded as 1 or 0 if the individual carried any or none of the risk alleles. We simulated rare variants with a frequency of 0.0001, 0.001, 0.005 or 0.01, and an OR of 2, 5 or 10. We used arbitrary values for the parameters of the rare genes, but based our choice on the literature (3, 26-28).

To compare the added value of rare and common variants for risk prediction, we also simulated 10 to 100 common variants each with a risk allele frequency of 0.05 or 0.30 and an OR of 1.10 or 1.05. We have used these parameters because most of the approximately 400 single nucleotide polymorphisms (SNPs) discovered in 100 recent GWAS had an OR of approximately 1.10 and future GWAS efforts are expected to uncover SNPs with even lower effect sizes (28). Disease risk was 4% as in the AF example, or 0.10 to examine the impact of higher disease risks on the measures of predictive ability. In the simulations, disease risk can be interpreted as a disease incidence, e.g. a disease incidence over 5 years. For both the main simulations and the AF example, population size was 200,000 for scenarios in which rare variants were added and 20,000 for scenarios in which common variants were added. Predicted risks for each individual were obtained from logistic regression analyses, were calculated in the range 0 to 1 and were rounded to two decimals.

Simulation study of atrial fibrillation

Background for choosing the example of atrial fibrillation

Complex diseases can be multifactorial, i.e. caused by an intricate effect of multiple environmental and genetic risk factors, but can also include monogenic forms. One such example is atrial fibrillation, which consist of a rare familial form that is a monogenetic disease and a common non-familial form (29). Targeted use of prevention strategies is warranted to reduce the burden

of AF, which requires accurate detection of individuals at high risk. Algorithms for detection of individuals at risk, based on routinely collected clinical risk factors, have already been developed and validated in various populations (17, 30, 31). The predictive accuracy of these clinical scores leaves ample opportunity for improvement, and so fuels the research for finding new biomarkers, including genetic variants (23, 32, 33). Several susceptibility variants for AF have been found (19-21) but their combined predictive ability is low as they explain only a fraction of the heritability (1). While sequencing efforts are ongoing for AF, research focused on the potential use of rare variants for risk prediction of AF becomes very relevant (23, 24). We assessed the incremental value of rare genetic variants over an existing clinical risk score for AF.

Methods for constructing the dataset of clinical and genetic risk factors

To assess the value of rare variants for AF risk prediction we simulated a hypothetical population that reflected the characteristics of the community-based cohort in which the clinical risk score was developed (i.e., the combination of Atherosclerosis Risk in Communities Study, Cardiovascular Health Study and Framingham Heart Study; see Supplementary Data for details on study design) (34-36). We simulated the distribution of clinical and genetic risk factors separately in individuals with and without the outcome by random sampling from a multivariate normal distribution. To derive categorical clinical variables and genetic variants, we transformed the corresponding continuous variables into categorical variables. We simulated clinical variables to be correlated, as observed in the empirical population. We assumed that genetic variants were uncorrelated with one another and were uncorrelated with clinical risk variables. Detailed information about the simulation strategy is provided in the Supplementary Methods.

Description of the clinical model

Variables included in the clinical risk score were: age, race, smoking status, weight, height, systolic blood pressure (SBP), diastolic blood pressure (DBP), diabetes, medication for hypertension, history of congestive heart failure, and history of myocardial infarction (see **Table 1 in Supplementary Data** for the distribution of clinical variables). In the empirical dataset, i.e. the combination of Atherosclerosis Risk in Communities Study, Cardiovascular Health Study and Framingham Heart Study, the outcome was defined as AF during 5 years of follow-up. Individuals were free of AF at the beginning of follow-up. The disease incidence was 0.04. Simulations accurately replicated the empirical data (see Supplementary Methods).

Description of genetic variables

We used the same parameters for the rare variants as in the simulation scenarios where we varied the baseline AUC. To estimate the added value of recently identified susceptibility single nucleotide polymorphisms (SNPs) for AF, we added to the clinical variables ten genetic variants with the same frequency and OR as the top ten (i.e. in terms of p-value) uncorrelated SNPs from a recent meta-analysis performed in the CHARGE AF consortium (22) (see **Table 2 in Supplementary Data**).

Metrics

We assessed *discrimination* of the baseline, genetic and combined models; *improvement in discrimination*; and *clinical usefulness* of updating the baseline model with genetic variants. We used AUC as a *global measure of discrimination*. AUC indicates the degree to which the predicted risks can discriminate between individuals who will and will not develop the disease. AUC generally ranges from 0.50 (equal to tossing a coin) to 1.00 (perfect discrimination). We used the increment in AUC (ΔAUC), IDI and continuous NRI as measures of *global improvement in discrimination*. IDI was calculated as the difference in mean predicted probabilities between cases and controls between the two models (16). NRI is an overall measure of correct reclassification of cases to higher risk categories and of controls to lower risk categories (16). The continuous NRI (NRI(>0)) does not use categories but takes into account any increase or decrease in predicted risk produced by the model update (37). Since we rounded risks to 0.01 (i.e. 1%) the NRI without categories used here is denoted as NRI(>0.01). In other words the minimal change in risk is 0.01. We used categorical NRI to assess *clinical usefulness*. Clinical usefulness concerns the reclassification of individuals in risk categories that leads to changes in preventive or therapeutic interventions. We defined three risk categories by using the risk cutoffs of 2.5 and 5%, similar to those used in the evaluation of the clinical risk score for AF (17). We also report the NRI in cases and controls separately, as this may provide additional insight into the impact of model update (16, 37). For scenarios with various baseline AUC, we calculated ΔAUC, NRI(>0.01) and IDI. For scenarios with the clinical risk score we calculated ΔAUC, IDI, NRI(>0.01) and categorical NRI.

Reported measures are median results from 200 simulations unless stated otherwise. All analyses were performed using the R programming language, version 2.11.1 (38).

RESULTS

Simulation analyses

Figure 1 shows that, for a disease risk of 4%, the median AUC and NRI(>0.01) only improved when variants were not very rare and had higher ORs, and only when baseline AUC values were in the lower range. Across higher baseline AUC values, the median NRI(>0.01) became negative, suggesting that rare variants produced more incorrect than correct risk reclassifications. The median IDI was close to zero for very rare variants and minimally increased with higher frequency and OR of the rare variants. The median IDI was constant across most baseline AUC values. When disease risk was higher (i.e. 10%), most performance measures slightly increased compared to the scenarios with lower disease risk (see **Supplementary Figure 1**). For rare variants with OR=10 and frequency ≥0.005, the median increment in AUC varied between 0.01 and 0.05 depending on the value of the baseline AUC. The median NRI(>0.01) varied between 0.18 and 0.55 and, in contrast to the scenario with the lower disease risk, increased with a higher baseline AUC.

As a comparison, we investigated the addition of 10 to 100 common variants, each with a frequency of 0.05 and an OR of 1.10 or a frequency of 0.30 and an OR of 1.05. We found a higher increase in AUC compared to the addition of rare variants (see **Figure 2**). NRI(>0.01) was mostly positive and increased with the number of variants added. In contrast, IDI was minimal even with the addition of 100 variants. Surprisingly, although the increment in AUC was higher, the IDI was in some instances lower for common variants compared to rare variants. As such, across low baseline AUC values rare variants with OR=10 and frequency ≥0.005 showed higher IDI than 100 common variants. This trend was seen also when disease risk was higher (i.e. 10%) (see **Supplementary Figure 2**). Furthermore, rare variants with OR=10 and frequency ≥0.005 showed also higher NRI(>0.01) across higher baseline AUC values compared to 100 common variants (see **Supplementary Figure 2**).

To investigate the added value of rare and common variants at the individual level, we additionally assessed the magnitude of change in absolute risk at reclassification. Having a rare variant substantially increased the risk in less than 1% of both cases and controls when disease risk was 4% (median increase in absolute risk: 0.35 in cases and 0.24 in controls; see **Figure 3a**). When disease risk was 10% the risk in 10% of cases largely increased while the risk in less than 1% of controls increased only marginally (median increase in absolute risk: 0.78 in cases and 0.02 in controls; see **Figure 3b**). The median decrease was negligible in both cases and controls that did not carry the risk variant (i.e. -0.01). In contrast, when 100 common variants were added to the model and the disease risk was 4%, the risk minimally increased or decreased in individuals that were reclassified to higher or lower risk categories (median increase in absolute risk: 0.03

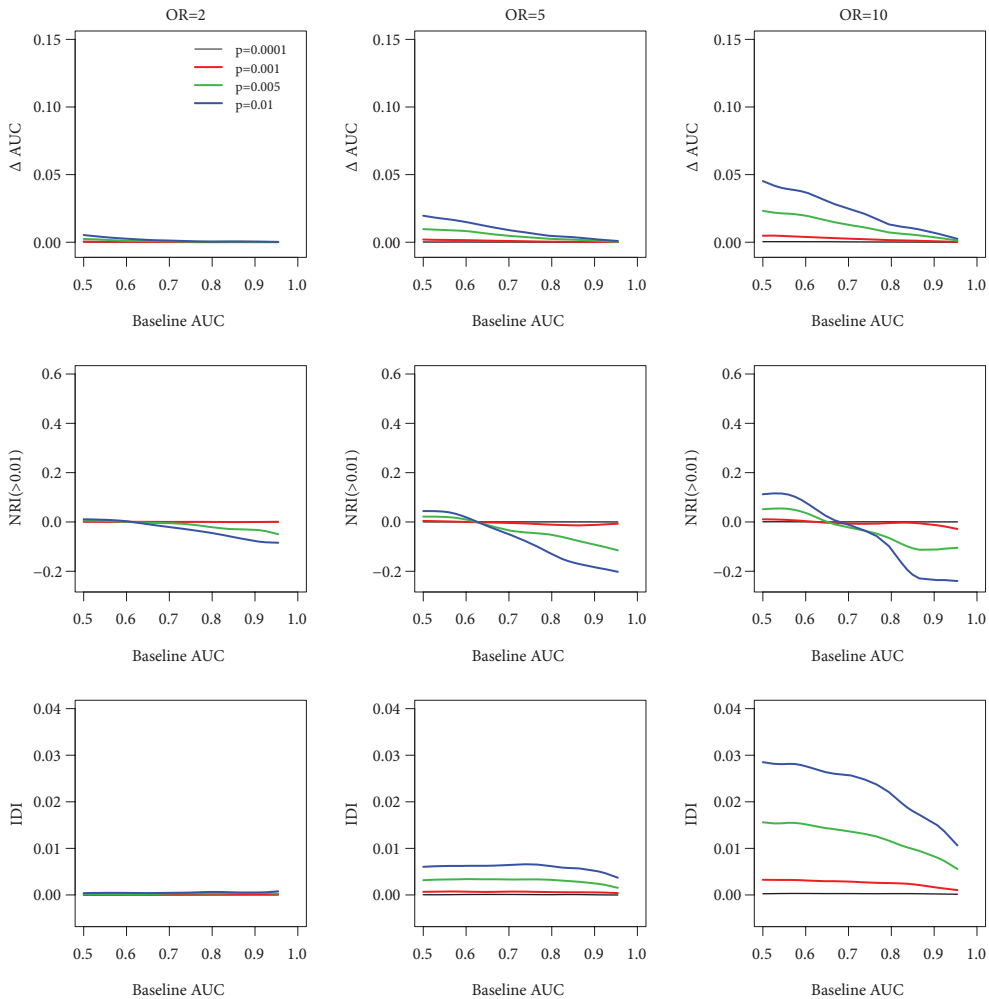
in cases and 0.02 in controls, median decrease: -0.02 and -0.01; see **Figure 3c**). When common variants were added, about a half of the cases and controls moved in the right direction while around a quarter moved in the wrong direction. Similar results were observed when disease risk was 10% (see **Figure 3d**). Besides the individuals that carried the risk variant, an increase in risk was observed also in some individuals that did not carry a rare variant. This risk increase was minimal and was due to a difference in beta coefficients between the two regression models (data not shown).

Table 1. Performance of genetic and combined (clinical and genetic) risk models for AF using rare and common variants.

OR	Frequency	# variants	AUC			IDI	NRI(>0.01)	NRI categorical		
			genetic	combined	Δ			total	cases	controls
	Rare variants									
2	**0.0001**	1	0.50	0.76	0	0	0	0	0	0
	0.001	1	0.50	0.76	0	0	0	0	0	0
	0.005	1	0.50	0.76	0	0	-0.01	0	0	0
	0.01	1	0.51	0.76	0	0	-0.03	0	0	0
5	**0.0001**	1	0.50	0.76	0	0	0	0	0	0
	0.001	1	0.50	0.76	0	0	-0.01	0	0	0
	0.005	1	0.51	0.76	0	0	-0.05	0.01	0.01	0
	0.01	1	0.52	0.77	0.01	0.01	-0.10	0.02	-0.01	0.03
10	**0.0001**	1	0.50	0.76	0	0	0	0	0	0
	0.001	1	0.50	0.76	0	0	-0.03	0.00	0	0.01
	0.005	1	0.52	0.77	0.01	0.02	-0.11	0.02	-0.01	0.03
	0.01	1	0.54	0.78	0.02	0.03	-0.13	0.04	-0.02	0.06
	Common variants									
1.14–1.45*	**0.03–0.84***	10	0.59	0.77	0.01	0.01	0.20	0.04	0.01	0.04

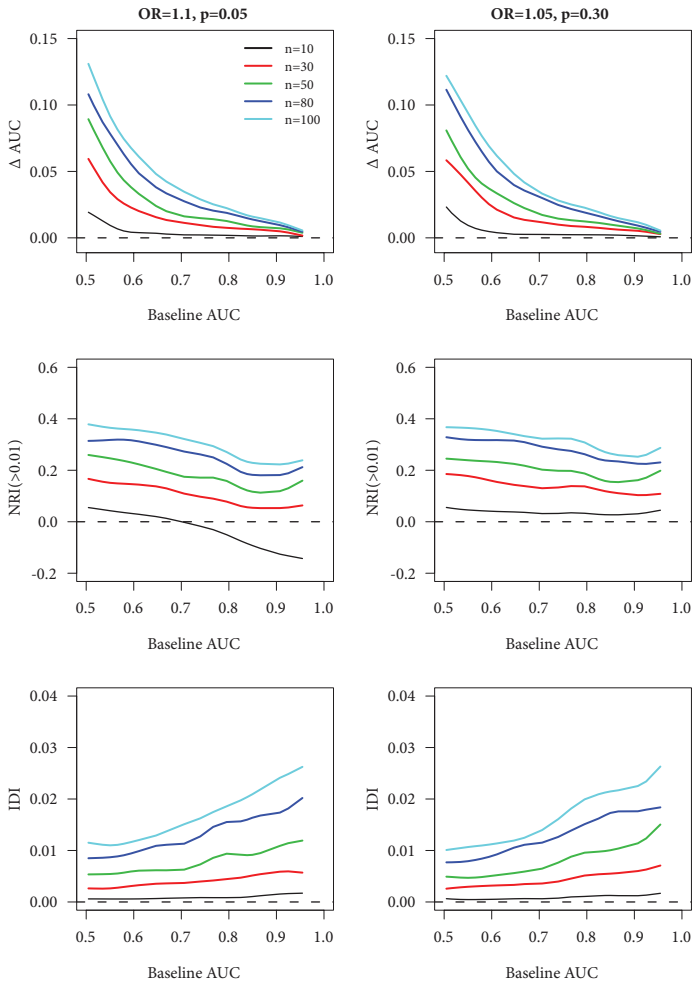
AUC, area under the receiver operating characteristic curve; ΔAUC, change in AUC between the model with and without genetic variants; IDI, integrated discrimination improvement; NRI, net reclassification improvement (cut-offs 2.5% and 5%); OR, odds ratio. *Using parameters from the top ten uncorrelated SNPs in the CHARGE AF meta-analysis; in the table are listed the range of OR and allele frequency.[21] Variables included in the clinical risk score were: age, weight, height, systolic blood pressure (SBP), diastolic blood pressure (DBP), diabetes, medication for hypertension, history of congestive heart failure, history of myocardial infarction, smoking status and race. Disease risk is 0.04 and population size is 200,000 for rare variants scenarios and 20,000 for common variants scenarios. Results are median values from 200 simulations.

Figure 1. Change in AUC, NRI(>0.01) and IDI per different values of the baseline AUC when rare genetic variants were added to prediction baseline model.



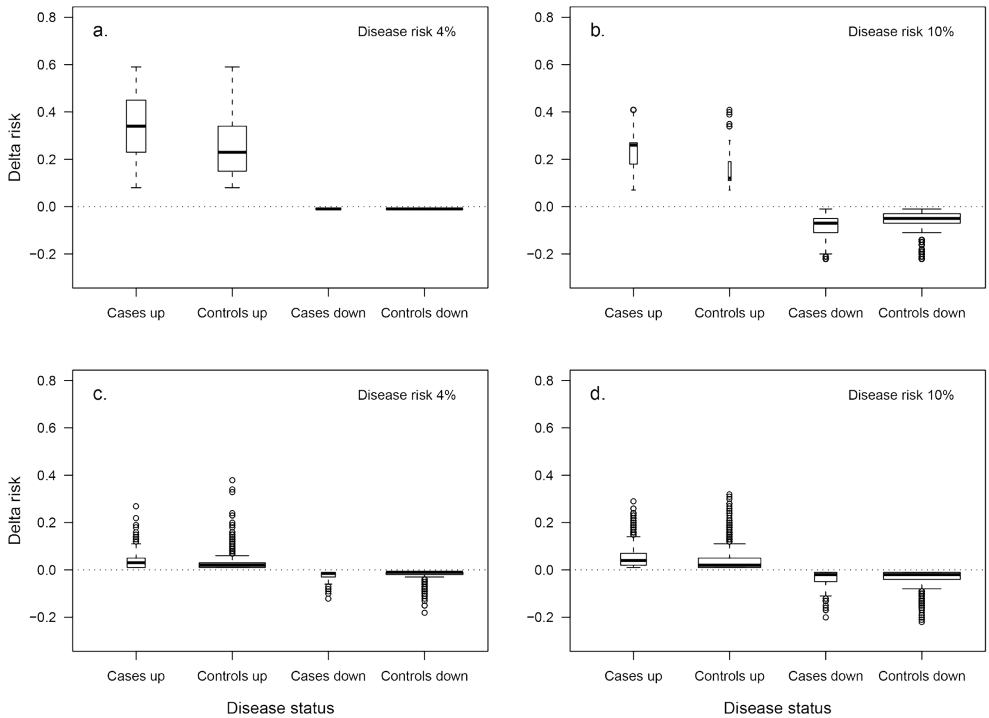
ΔAUC, change in AUC between the model with and without the rare genetic variants; AUC, area under the receiver operating characteristic curve; IDI, integrated discrimination improvement; n, number of common variants added; NRI, net reclassification improvement; OR, odds ratio; p, frequency of the risk allele; AUC 1, AUC of the baseline model. Disease risk is 4%. Population size is 200,000. Results are median values from 10 simulations.

Figure 2. Change in AUC, NRI(>0.01) and IDI per different values of the baseline AUC when common genetic variants were added to prediction baseline model.



ΔAUC, change in AUC between the model with and without the common genetic variants; AUC, area under the receiver operating characteristic curve; IDI, integrated discrimination improvement; n, number of common variants added; NRI, net reclassification improvement; OR, odds ratio; p, frequency of the risk allele; AUC 1, AUC of the baseline model. Disease risk is 0.4%. Population size is 20,000. Results are median values from 10 simulations.

Figure 3. Change in absolute risk at model update with rare and common genetic variants.



On the x axis is shown the correct reclassification of cases and controls (i.e. Cases up; Controls down) and incorrect reclassification (i.e. Cases down; Controls up) when rare variants with a cumulative OR of 10 and frequency of 0.01 (Figures 3a-b) or 100 common variants each with a OR of 1.05 and a frequency of 0.30 (Figures 3c-d) are added to a baseline model with an AUC=0.70. The bold line shows the median, the boxes indicate the interquartile ranges (25%-75% range), and the whiskers present 1.5 times the interquartile range. Box widths are proportional to the square-root of the number of individuals in the groups. Disease risk is 4% in figures 3a and 3c, and 10% in Figures 3b and 3d.The plot is obtained from one simulation using 200,000 individuals for figures 3a-b, and 20,000 individuals for figures 3c-d.

Clinical example: genetic prediction of atrial fibrillation

The median AUC of the clinical model was 0.76 (95% confidence interval, 0.75 to 0.78). Rare variants improved the AUC of the clinical model only when they were more frequent and had very high OR (see **Table 1)**. Adding a rare variant with a frequency of 0.01 increased AUC by 0.02 when OR was 10, but did not improve AUC when OR was 2. IDI and NRI were zero for very rare variants, i.e. when frequency was 0.0001 or 0.001. IDI minimally increased with a higher frequency and OR of the rare variant (see **Table 1**). A variant with frequency of 0.01 and an OR of 10 yielded an IDI of 0.03. NRI(>0.01) was negative in most scenarios. The higher the

frequency and OR of the rare variant, the larger the negative value of NRI(>0.01). In contrast to the NRI(>0.01), categorical NRI and NRI in controls, but not in cases, were positive and minimally increased with the frequency and OR of the rare variants. Adding ten variants with empirical ORs and frequencies showed a minimal improvement in all model performance measures.

DISCUSSION

Using a hypothetical population, we have shown that rare variants only minimally improved AUC and did not yield clinically relevant positive NRI(>0.01) and IDI when disease risk was low. Rare variants produced larger increments in AUC when the baseline model had lower AUC, but in these scenarios NRI(>0.01) and IDI remained close to zero. Addition of rare variants to the baseline model largely increased predicted risks for the few individuals carrying the risk variant, whereas predicted risks were only slightly decreased for those who did not carry the variant. For a higher disease risk, rare variants with strong effects showed improvement in AUC across a wider range of baseline AUC values and significant positive NRI(>0.01) and IDI. Addition of rare variants to the baseline model largely increased predicted risks only in cases, as expected.

Before addressing the implications of these results for future research, we discuss several methodological aspects of our study that might have impacted the results. First, we modeled rare variants as a single variant. This is a common procedure used to investigate the association of multiple extremely rare variants with disease risk and does not affect the results presented here (9, 39). Second, we assumed that each genetic variant was uncorrelated with other variants and clinical risk factors. While linkage equilibrium between rare variants is a very realistic assumption, rare variants may be in linkage disequilibrium with common variants. In fact, it has been suggested that common variants may share a haplotype with the true rare causal variants (6). Third, genetic variants may be associated with intermediate risk factors for disease, which are often the variables included in traditional clinical risk scores. Such correlations would likely decrease the impact of the variants and hence show less improvement in performance than reported in this paper.

We have shown that, from a population perspective, rare variants are only useful for risk prediction of complex diseases when they have strong effects on disease risk, when they are not too rare and when the risk of disease is high. **Figure 1** shows that when disease risk was 4% the addition of rare variants resulted in an improvement in AUC only when the baseline AUC was low. As shown, this trend was more pronounced when rare variants had higher OR. NRI(>0.01) showed a minimal added value of rare variants only when baseline AUC was lower (≤0.70) and variants had very strong effects (OR=10). When baseline AUC was ≥0.80, the NRI(>0.01)

indicated that overall more wrong reclassifications of risk were done by addition of rare variants. To summarize, rare variants only improved discrimination when baseline AUC values were low, but even then the improvement was minimal. Only when the OR of the rare variant was very large, its frequency higher and disease risk high, were the AUC and NRI(>0.01) improved across a wider range of baseline AUC. If the expected effect sizes might be smaller than previously thought, the predictive ability of rare variants will be even lower than our results indicate. The NRI(>0.01) and IDI values were higher than those observed with the addition of 100 common variants with a frequency and OR as used in this study. Thus, despite a lower improvement in AUC, rare variants may result in larger improvements in NRI(>0.01) and IDI compared to common variants. This apparent discrepancy in observations is explained by the fact that AUC only considers the rank in predicted risks, not actual values, whereas NRI and IDI do depend on actual magnitude of changes in predicted risks before and after updating the model. Rare variants with strong effect by definition have a substantial impact on disease risk, albeit for a small group of individuals.

Interestingly, the degree of precision had a large impact on the global improvement of discrimination as measured by continuous NRI. The discrepancy in results was most striking for the rare variants. As such, we observed a NRI(>0) of 0.17 when a rare variant with OR of 10 and frequency of 0.01 was added to the clinical AF model and predicted risks were not rounded, compared to an NRI(>0.01) of -0.13 when risks were rounded to two decimals. This is likely explained by the fact that, by definition, most individuals did not carry the rare risk variant and this resulted in a very small decrease in risk for most individuals, a change that was not captured when risks were rounded. In contrast, the difference in AUC and categorical NRI between non-rounded and rounded risks was minimal, i.e. at most 0.01 in a few scenarios from the AF example. This raises the question what is the amount of precision to be reported for risk predictions and what is the most appropriate continuous NRI measure.

CONCLUSIONS

In conclusion, we have shown that addition of rare variants to baseline risk models that include clinical or genetic risk factors resulted in model improvement only when the rare variants had strong effects on disease risk. This improvement was larger with a higher disease risk because the odds ratios lead to different likelihood ratios when the disease is more common. We have also shown that rare variants largely increased the risk in some individuals, while most individuals were reclassified to a slightly decreased risk. Very rare variants, by definition, occur in only few individuals that ultimately develop the disease and therefore have poor sensitivity and a limited predictive ability. This means that most individuals will either not be reclassified into another

risk category or will be reclassified on the basis of clinically irrelevant changes in predicted risks. Counterintuitively, most individuals who will develop the disease will see their risk slightly decreased after testing for rare variants. Although this decrease in disease risk is minimal, individuals with predicted risks just above the threshold may be moved to a lower risk category. In the case of AF, a disease associated with stroke and increased mortality (40), this would mean that many individuals would not benefit from the potentially lifesaving preventive measures.

While rare variants are unlikely to improve the prediction of common diseases in the population, they do have substantial impact on disease risk in carriers of the rare variants. When rare variants have very strong effects on disease risk, they are probably more aggregated within certain families and resemble a Mendelian transmission. It would be of high interest to compare family history information (41) with tests including rare variants and, further, to investigate if such variants can be more predictive in families with positive family history. Apart from such exceptions, it should be anticipated that the study of rare variants will have its largest contribution in advancing our understanding of disease pathophysiology.

REFERENCES

1. Jostins L, Barrett JC: **Genetic risk prediction in complex disease.** *Hum Mol Genet* 2011;**20:**R182-188.
2. Makowsky R, Pajewski NM, Klimentidis YC, Vazquez AI, Duarte CW, Allison DB, de los Campos G: **Beyond missing heritability: prediction of complex traits.** *PLoS Genet* 2011;**7:**e1002051.
3. Bodmer W, Bonilla C: **Common and rare variants in multifactorial susceptibility to common diseases.** *Nat Genet* 2008;**40:**695-701.
4. Gorlov IP, Gorlova OY, Frazier ML, Spitz MR, Amos CI: **Evolutionary evidence of the effect of rare variants on disease etiology.** *Clin Genet* 2011;**79:**199-206.
5. Visscher PM, Goddard ME, Derks EM, Wray NR: **Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses.** *Mol Psychiatry* 2011.
6. Wang K, Dickson SP, Stolle CA, Krantz ID, Goldstein DB, Hakonarson H: **Interpretation of association signals and identification of causal variants from genome-wide association studies.** *Am J Hum Genet* 2010;**86:**730-742.
7. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, Shan Y, Li S, Yang Q, Asan, Ni P, Tian G, Xu J, Liu X, Jiang T, Wu R, et al: **Sequencing of 50 human exomes reveals adaptation to high altitude.** *Science* 2010;**329:**75-78.
8. Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR: **Pooled association tests for rare variants in exon-resequencing studies.** *Am J Hum Genet* 2010;**86:**832-838.
9. Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP: **Rare independent mutations in renal salt handling genes contribute to blood pressure variation.** *Nat Genet* 2008;**40:**592-599.
10. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA: **Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes.** *Science* 2009;**324:**387-389.
11. Wei C, Lu Q: **Collapsing ROC approach for risk prediction research on both common and rare variants.** *BMC Proceedings* 2011;**5(Suppl 9):S42**.
12. Wu C, Walsh KM, A.T. D, J. H, Z. W: **Disease risk prediction with rare and common variants.** *BMC Proceedings* 2011;**5(Suppl 9):S61**.
13. Cook NR: **Use and misuse of the receiver operating characteristic curve in risk prediction.** *Circulation* 2007;**115:**928-935.
14. Pepe MS, Janes HE: **Gauging the performance of SNPs, biomarkers, and clinical factors for predicting risk of breast cancer.** *J Natl Cancer Inst* 2008;**100:**978-979.
15. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, Zheng Y: **Integrating the predictiveness of a marker with its performance as a classifier.** *Am J Epidemiol* 2008;**167:**362-368.
16. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS: **Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond.** *Stat Med* 2008;**27:**157-172; discussion 207-112.
17. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB, Sinner MF, Sotoodehnia N, Fontes JD, Janssens AC, Kronmal RA, Magnani JW, Witteman JC, Chamberlain AM, Lubitz SA, Schnabel RB, Agarwal SK, McManus DD, Ellinor PT, Larson MG, Burke GL, Launer LJ, Hofman A, Levy D, Gottdiener JS, Kaab S, Couper D, Harris TB, Soliman EZ, Stricker BH, et al: **Simple Risk Model Predicts Incidence of Atrial Fibrillation in a Racially and Geographically Diverse Population: the CHARGE-AF Consortium.** *J Am Heart Assoc* 2013;**2:**e000102.

18. Wolowacz SE, Samuel M, Brennan VK, Jasso-Mosqueda JG, Van Gelder IC: **The cost of illness of atrial fibrillation: a systematic review of the recent literature.** *Europace* 2011;**13:**1375-1385.

19. Benjamin EJ, Rice KM, Arking DE, Pfeufer A, van Noord C, Smith AV, Schnabel RB, Bis JC, Boerwinkle E, Sinner MF, Dehghan A, Lubitz SA, D'Agostino RB, Sr., Lumley T, Ehret GB, Heeringa J, Aspelund T, Newton-Cheh C, Larson MG, Marciante KD, Soliman EZ, Rivadeneira F, Wang TJ, Eiriksdottir G, Levy D, Psaty BM, Li M, Chamberlain AM, Hofman A, Vasan RS, et al: **Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry.** *Nat Genet* 2009;**41:**879-881.

20. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, Palsson A, Blondal T, Sulem P, Backman VM, Hardarson GA, Palsdottir E, Helgason A, Sigurjonsdottir R, Sverrisson JT, Kostulas K, Ng MC, Baum L, So WY, Wong KS, Chan JC, Furie KL, Greenberg SM, Sale M, Kelly P, MacRae CA, et al: **Variants conferring risk of atrial fibrillation on chromosome 4q25.** *Nature* 2007;**448:**353-357.

21. Ellinor PT, Lunetta KL, Glazer NL, Pfeufer A, Alonso A, Chung MK, Sinner MF, de Bakker PI, Mueller M, Lubitz SA, Fox E, Darbar D, Smith NL, Smith JD, Schnabel RB, Soliman EZ, Rice KM, Van Wagoner DR, Beckmann BM, van Noord C, Wang K, Ehret GB, Rotter JI, Hazen SL, Steinbeck G, Smith AV, Launer LJ, Harris TB, Makino S, Nelis M, et al: **Common variants in KCNN3 are associated with lone atrial fibrillation.** *Nat Genet* 2010;**42:**240-244.

22. Ellinor PT, Lunetta KL, Albert CM, Glazer NL, Ritchie MD, Smith AV, Arking DE, Muller-Nurasyid M, Krijthe BP, Lubitz SA, Bis JC, Chung MK, Dorr M, Ozaki K, Roberts JD, Smith JG, Pfeufer A, Sinner MF, Lohman K, Ding J, Smith NL, Smith JD, Rienstra M, Rice KM, Van Wagoner DR, Magnani JW, Wakili R, Clauss S, Rotter JI, Steinbeck G, et al: **Meta-analysis identifies six new susceptibility loci for atrial fibrillation.** *Nat Genet* 2012;**44:**670-675.

23. O'Donnell CJ, Nabel EG: **Genomics of cardiovascular disease.** *N Engl J Med* 2011;**365:**2098-2109.

24. Parvez B, Darbar D: **The "missing" link in atrial fibrillation heritability.** *J Electrocardiol* 2011;**44:**641-644.

25. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM: **Predictive testing for complex diseases using multiple genes: fact or fiction?** *Genet Med* 2006;**8:**395-400.

26. Oexle K: **A remark on rare variants.** *J Hum Genet* 2010;**55:**219-226.

27. Ionita-Laza I, Ottman R: **Study designs for identification of rare disease variants in complex diseases: the utility of family-based designs.** *Genetics* 2011;**189:**1061-1068.

28. Pawitan Y, Seng KC, Magnusson PK: **How many genetic variants remain to be discovered?** *PLoS One* 2009;**4:e**7969.

29. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB, Sr., Newton-Cheh C, Yamamoto JF, Magnani JW, Tadros TM, Kannel WB, Wang TJ, Ellinor PT, Wolf PA, Vasan RS, Benjamin EJ: **Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study.** *Lancet* 2009;**373:**739-745.

30. Chamberlain AM, Agarwal SK, Folsom AR, Soliman EZ, Chambless LE, Crow R, Ambrose M, Alonso A: **A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk in Communities (ARIC) study).** *Am J Cardiol* 2011;**107:**85-91.

31. Schnabel RB, Larson MG, Yamamoto JF, Sullivan LM, Pencina MJ, Meigs JB, Tofler GH, Selhub J, Jacques PF, Wolf PA, Magnani JW, Ellinor PT, Wang TJ, Levy D, Vasan RS, Benjamin EJ: **Relations of biomarkers of distinct pathophysiological pathways and atrial fibrillation incidence in the community.** *Circulation* 2010;**121:**200-207.

32. Smith JG, Newton-Cheh C, Almgren P, Struck J, Morgenthaler NG, Bergmann A, Platonov PG, Hedblad B, Engstrom G, Wang TJ, Melander O: **Assessment of conventional cardiovascular risk factors and multiple biomarkers for the prediction of incident heart failure and atrial fibrillation.** *J Am Coll Cardiol* 2010;**56:**1712-1719.

33. **The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators.** *Am J Epidemiol* 1989;**129:**687-702.

34. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A, et al.: **The Cardiovascular Health Study: design and rationale.** *Ann Epidemiol* 1991;**1:**263-276.

35. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP: **The Framingham Offspring Study. Design and preliminary data.** *Prev Med* 1975;**4:**518-525.

36. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW: **Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers.** *Stat Med* 2011;**30:**11-21.

37. Vienna: R Foundation for Statistical Computing. R Development Core Team: **R: a language and environment for statistical computing.** Version 2.8.0. (http://www.R-project.org)

38. Guey LT, Kravic J, Melander O, Burtt NP, Laramie JM, Lyssenko V, Jonsson A, Lindholm E, Tuomi T, Isomaa B, Nilsson P, Almgren P, Kathiresan S, Groop L, Seymour AB, Altshuler D, Voight BF: **Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants.** *Genet Epidemiol* 2011. doi: 10.1002/gepi.20572.

SUPPLEMENTARY MATERIAL

Study design

Atherosclerosis Risk in Communities Study

The ARIC study recruited 15,792 men and women, aged 45-64 years, from 4 communities in the United States (Forsyth County, NC; Washington County, MD; Jackson, MS; and suburbs of Minneapolis, MN) in 1987-89. Participants were mostly white in the Minnesota and Washington County field centers, white and African-American in Forsyth County, and exclusively African-Americans in the Jackson field center. After baseline, study participants had three follow-up examinations approximately three years apart each (last one in 1996-98).

Cardiovascular Health Study

In 1989-90, CHS recruited 5201 mostly white men and women 65 years or older from 4 communities (Forsyth County, NC; Washington County, MD; Sacramento County, CA; and Pittsburgh, PA). In 1992-93, 687 African-Americans were recruited in 3 of the 4 communities to increase minority representation. CHS participants had annual exams through 1999, with ongoing surveillance of cardiovascular events.

Framingham Heart Study

The FHS Offspring cohort recruited 5124 predominantly white men and women, offspring (and their spouses) from the Original FHS cohort, in 1971-1975, with follow-up exams every 4-8 years.

The study procedures were approved by the Institutional Review Boards at the participating institutions.

Supplementary Table 1. Baseline characteristics in Atherosclerosis Risk in Communities Study, Cardiovascular Health Study and Framingham Heart Study combined dataset.

Baseline characteristics	Cases Mean (SD) or percentage, %	Controls Mean (SD) or percentage, %
Age	71.2 (7.7)	64.8 (7.7)
Weight, kg	79 (17)	78 (17)
Height, cm	168 (10)	167 (9)
SBP, mm Hg	137 (23)	130 (20)
DBP, mm Hg	70 (11)	72 (11)
Diabetes	23	15
Hypertension treatment	58	41
Heart failure	10	4
Myocardial infarction	15	6
Smoking	11	13
Race, white	87	83

CHF, history of congestive heart failure; DBP, diastolic blood pressure; HRX, medication for hypertension; MI, history of myocardial infarction; SBP, systolic blood pressure. Correlations were obtained using the 'hetcor' function from the "Polycor" R package (see main text for further details).

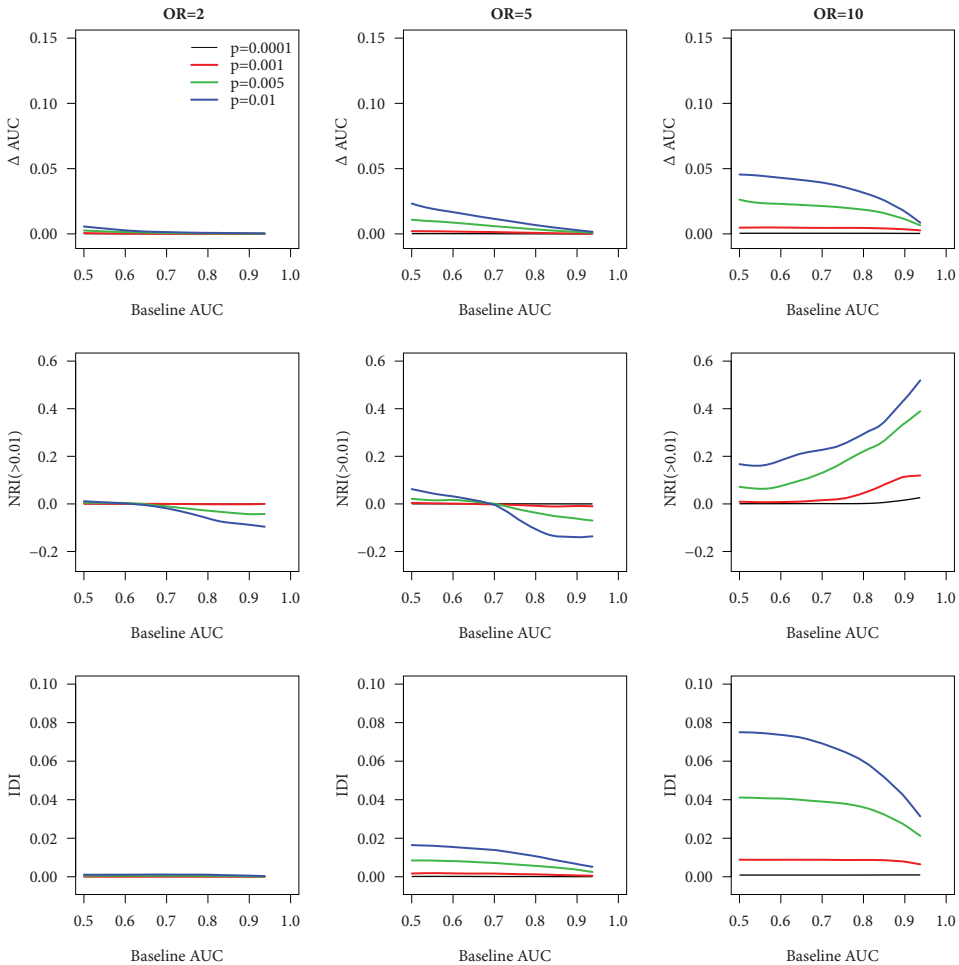
Supplementary Table 2. Description of top ten independent SNPs from the CHARGE AF meta-analysis (1).

RS number	Chromosome	Risk allele	Risk allele frequency	Risk allele OR
rs17042171	4	C	0.12	1.45
rs17375901	1	C	0.05	1.33
rs2106261	16	C	0.17	1.19
rs1256065	14	T	0.44	1.14
rs10934122	3	C	0.55	1.14
rs8092700	18	C	0.66	1.15
rs16829334	3	G	0.06	1.27
rs11963663	6	C	0.03	1.41
rs10824026	10	G	0.84	1.18
rs9624206	22	C	0.28	1.19

SUPPLEMENTAL REFERENCES

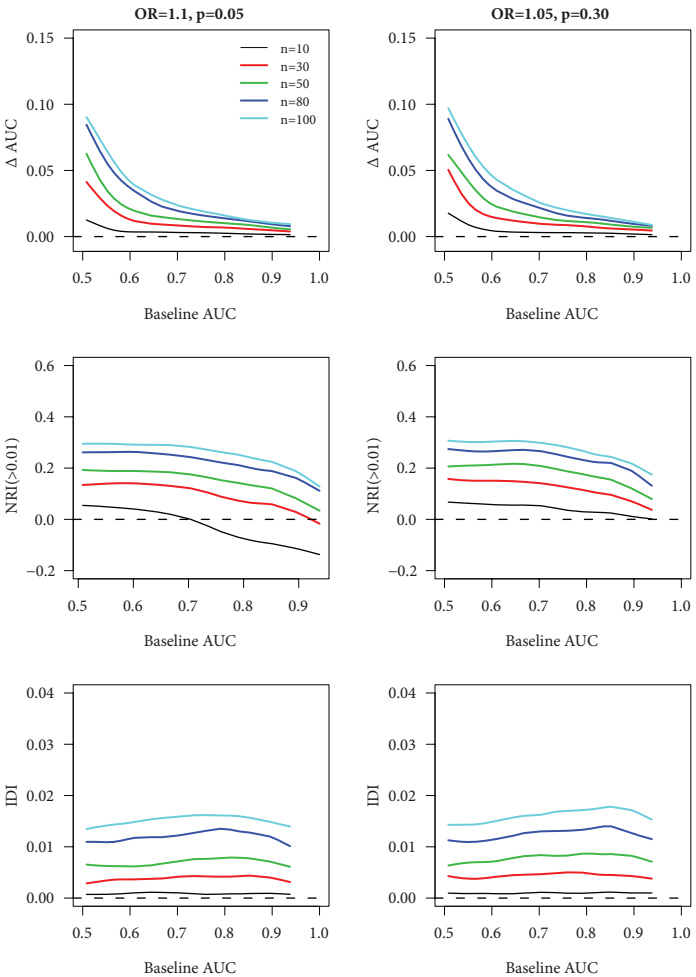
1. Ellinor PT, Lunetta KL, Albert CM, Glazer NL, Ritchie MD, Smith AV, et al. **Meta-analysis identifies six new susceptibility loci for atrial fibrillation.** Nat. Genet. 2012;**44**:670-675.

Supplementary Figure 1. Change in AUC, NRI(>0.01) and IDI per different values of the baseline AUC when rare genetic variants are added to prediction baseline model.



ΔAUC, change in AUC between the model with and without the rare genetic variants; AUC, area under the receiver operating characteristic curve; IDI, integrated discrimination improvement; NRI, net reclassification improvement; OR, odds ratio; p, frequency of the risk allele; AUC 1, AUC of the baseline model. Disease risk is 0.10 and population size is 200,000. Results are median values from 10 simulations.

Supplementary Figure 2 Change in AUC, NRI(>0.01) and IDI per different values of the baseline AUC when common genetic variants are added to prediction baseline model.



ΔAUC, change in AUC between the model with and without the common genetic variants; AUC, area under the receiver operating characteristic curve; IDI, integrated discrimination improvement; n, number of common variants added; NRI, net reclassification improvement; OR, odds ratio; p, frequency of the risk allele; AUC 1, AUC of the baseline model. Disease risk is 0.10 and population size is 20,000. Results are median values from 10 simulations.

SUPPLEMENTARY METHODS

In the atrial fibrillation example we simulated the distribution of risk factors separately in individuals with and without the outcome by random sampling from a multivariate normal distribution, with the help of the function 'mvrnorm' implemented in the "MASS" package in R. We generated numbers of individuals with and without disease equal to those observed in the empirical sample. We used the polyserial correlation to estimate the correlation between a continuous and a categorical variable and the polychoric correlation for a correlation between two categorical variables, as available from the function 'hetcor' from the R Package "Polycor". We used the correlation matrices obtained with the 'hetcor' function and the standard deviations (the standard deviation for the underlying normal variables was set to 1) to calculate the covariance matrices. We supplied these matrices and the vector of mean values (the mean for the underlying normal variables was set to 0) to the 'mvrnorm' function. To derive binary variables from the generated continuous ones, we cut the distribution of the generated continuous variables at the probit (i.e., the value of the standard normal distribution corresponding to the probability of the binary variable; e.g., when the value generated from the multivariate normal distribution was higher than the probit corresponding to the percentage of non-smokers it was recoded as 1, equivalent to positive smoking status).

We simulated the genetic variants from multivariate normal distributions and transformed the generated continuous variables by cutting them at the probit for the distribution of the genotypes. As such, in the case of rare variants, when the value from the continuous variable was lower than the probit for the proportion of individuals who did not carry a risk allele it was coded as 0 and when it was higher it was coded as 1. In the case of common variants, when the value from the continuous variable was lower than the probit for the proportion of homozygous carriers of the non-risk allele it was coded as 0; when the value was higher than the probit for the proportion of homozygous carriers of the non-risk allele but lower than the probit for the proportion of homozygous carriers of the non-risk allele plus the proportion of heterozygous carriers of the risk allele it was coded as 1; and when the value was higher than the probit for the proportion of homozygous carriers of the non-risk allele plus the proportion of heterozygous carriers of the risk allele it was coded as 2.

Supplementary Table 3. Correlations between clinical variables in Atherosclerosis Risk in Communities Study, Cardiovascular Health Study and Framingham Heart Study combined dataset.

Cases											
	Age	**Weight**	**Height**	**Systolic BP**	**Diastolic BP**	**Diabetes**	**Hypertension treatment**	**Heart failure**	**Myocardial Infarction**	**Smoking**	**Race (white)**
Age	1.00										
Weight	-0.35	1.00									
Height	-0.21	0.49	1								
Systolic BP	0.25	-0.09	-0.18	1							
Diastolic BP	-0.08	0.15	0.15	0.48	1						
Diabetes	-0.12	0.36	0.10	0.13	0.01	1					
Hypertension treatment	0	0.15	-0.15	0.17	-0.01	0.29	1				
Heart failure	0.14	-0.02	-0.17	0.03	-0.15	0.19	0.49	1			
Myocardial infarction	0.09	-0.02	0.16	-0.01	-0.04	0.27	0.29	0.27	1		
Smoking	-0.27	-0.10	0.02	-0.14	-0.06	-0.05	-0.15	-0.12	-0.13	1	
Race (white)	0.11	-0.19	0.10	-0.09	-0.25	-0.18	-0.18	-0.28	0.10	-0.02	1

Controls											
	Age	**Weight**	**Height**	**Systolic BP**	**Diastolic BP**	**Diabetes**	**Hypertension treatment**	**Heart failure**	**Myocardial Infarction**	**Smoking**	**Race (white)**
Age	1										
Weight	-0.22	1									
Height	-0.17	0.51	1								
Systolic BP	0.28	0.04	-0.11	1							
Diastolic BP	-0.14	0.18	0.12	0.52	1						
Diabetes	0.08	0.28	0.04	0.16	-0.05	1					
Hypertension treatment	0.19	0.19	-0.06	0.27	0.09	0.35	1				
Heart failure	0.09	0.16	-0.06	0.06	-0.08	0.28	0.52	1			
Myocardial infarction	0.21	0.08	0.10	0.01	-0.12	0.21	0.41	0.48	1		
Smoking	-0.19	-0.13	0.03	-0.09	-0.08	-0.04	-0.10	0.01	0	1	
Race (white)	0.06	-0.21	-0.01	-0.18	-0.26	-0.26	-0.30	-0.20	0.02	-0.10	1

BP, blood pressure. Correlations were obtained using the 'hetcor' function from the "Polycor" R package (see main text for further details).

Supplementary Table 4. Comparison of beta coefficients from Atherosclerosis Risk in Communities Study, Cardiovascular Health Study and Framingham Heart Study combined dataset and median values from 200 simulations.

Coefficients:												
	Intercept	**Age**	**Weight (kg)**	**Height (cm)**	**Systolic BP**	**Diastolic BP**	**Diabetes**	**Hypertension treatment**	**Heart failure**	**Myocardial Infarction**	**Smoking**	**Race (white)**
Empirical	-16.568	0.104	0.007	0.029	0.009	-0.012	0.254	0.367	0.660	0.426	0.254	0.346
Simulated	-16.427	0.099	0.003	0.032	0.011	-0.015	0.241	0.315	0.681	0.427	0.158	0.384

CHF, history of congestive heart failure; DBP, diastolic blood pressure; MI, history of myocardial infarction; SBP, systolic blood pressure.

Chapter 6

Completeness of reporting in genetic risk prediction studies: A review of published articles based on the GRIPS statement

Adriana I. Iglesias, Raluca Mihaescu, John P.A. Ioannidis, Muin J. Khoury, Julian Little, Cornelia M. van Duijn, A. Cecile J. W. Janssens

Submitted

ABSTRACT

Context: Genetic risk prediction studies are becoming increasingly prominent in the literature. Complete and transparent reporting is essential to maximize the utility of this literature. We aimed to assess the completeness of reporting of such studies.

Methods: Studies were eligible if they developed or validated a prediction model based on multiple DNA variants, using empirical data and were published in 2010. A data-extraction form with 63 elements and sub-elements based on the 25 items of the Genetic Risk Prediction Studies (GRIPS) statement was created and piloted.

Results: Forty-two studies met our inclusion criteria. Overall, more than half of the evaluated items (34/62) were reported in at least 85% of included articles. Seventy-seven per cent of the articles were identified as genetic risk prediction studies through title assessment, but only 31% used the keywords recommended by GRIPS in the title or abstract. All but one of the articles described how genetic variants were handled in the analyses, and 74% mentioned which allele was the risk variant. Of the articles that used regression analyses to construct genetic risk models, only 17% reported the model intercept. Overall, only 10% of the articles reported all essential items needed to perform external validation of the risk model.

Conclusions: Completeness of reporting in genetic risk prediction studies is adequate for general elements of study design but is suboptimal for several aspects that characterize genetic risk prediction studies, such as uniform keywords in title and abstract and the description of the model construction. Improvements in the transparency of reporting of these aspects would facilitate the identification, replication, and application of genetic risk prediction models.

INTRODUCTION

Rapid advances in genome wide association studies and the promise of personalized medicine (1-3) have increased interest in the potential utility of genetic susceptibility variants for individualized risk prediction of complex diseases. The number of scientific publications investigating the contribution of genetic variants to disease risk prediction is rapidly increasing.

Genetic risk prediction studies vary considerably in terms of study populations, study design, methods used and analyses performed (4, 5). Therefore, clear and complete reporting is essential to understand, apply and replicate the results of these studies. However, scientific papers may lack key information that is needed to evaluate the validity and clinical applicability of the study - for example, how genetic variants were selected, definition of the risk allele(s), or specification of the risk period evaluated (4). This important information about the risk model is crucial for replication of the risk model by other researchers.

To maximize the transparency, quality and completeness of reporting on research methodology and findings, guidelines have been published for reporting various research designs, (6) such as for: genetic association studies (STREGA), observational studies (STROBE), diagnostic accuracy studies (STARD), clinical trials (CONSORT), and those for tumor marker prognostic studies (REMARK) (7-10). Because none of these guidelines adequately address genetic risk prediction studies, the Genetic RIsk Prediction Studies (GRIPS) statement was developed by the Human Genome Epidemiology Network (HuGENet) (5, 11) and published concurrently in 10 journals in 2011.

The GRIPS statement contains a list of 25 items recommended for consideration in the reporting of genetic risk prediction studies. They include a clear description of the key elements of the study design, eligibility criteria, relevant dates and characteristics of the study population. Strong emphasis is put on information relevant to the interpretation and validation of genetic risk models. Examples include specification of the genetic variant that is considered as the risk variant, handling of genetic variants in the construction of the risk model, metrics used to assess model performance, and if pertinent, discussion of the clinical applicability of the risk model (5).

Several studies have evaluated whether the implementation of a reporting guidelines improves the quality of reporting (12-16). These showed, for example, that the quality of reporting of clinical trials improved in the journals that adopted the CONSORT statement (14). The aim of this article is to assess the reporting of genetic risk prediction studies that were published prior to the GRIPS publication. Our main objective is to raise awareness of the areas that need improvements in the reporting of genetic risk predictions papers for future publications.

MATERIALS AND METHODS

Literature Search

Since there is no validated search strategy for identifying genetic risk prediction studies, we devised a strategy with multiple search terms covering the keywords "genetic", "risk" and "prediction", considered as the three defining elements of this type of study. This strategy aims to include a broad scope of genetic risk prediction studies. The combination of terms used to search PubMed was: (genetic (ti) OR genomic (ti) OR genes (ti) OR DNA (ti) OR polymorphism (ti) OR polygenic (ti)) AND (risk (All Fields)) AND (score (All Fields) OR model (All Fields) OR prediction (All Fields)). We limited our search to studies published in English in 2010. Two reviewers (AI and RM) independently made a first selection based on the title and abstract of all retrieved articles, and then the full-text of all potentially eligible articles was assessed. Additional material such as supplementary data was also obtained, if available. Disagreements between the reviewers were discussed with a third evaluator (ACJWJ) to reach final consensus.

Article Selection

We included articles which developed or validated a prediction model that contained multiple DNA variants alone or in combination with non-genetic risk factors. DNA variants could include single nucleotide polymorphisms (SNPs) or haplotypes. Articles were excluded when they performed solely: a) univariate or multivariate genetic association studies i.e., articles that described the association between one or multiple genetic variants and the outcome, and did not aim to use the information for risk prediction; b) meta-analyses of single genetic associations; c) statistical, methodological or simulation studies, for example, articles that developed genetic association, family or sibling tests, risk prediction methodology or studies based on simulation data; d) gene-gene interaction association studies; and e) reviews.

Data Extraction and Analyses

A data extraction form based on the 25 items of the GRIPS statement was developed. Most items were mapped to two or more elements (see **Supplementary Table 1)**. The definition of each element was discussed between three evaluators (AI, RM and ACJWJ) to ensure that the collected information was clear and unambiguous. Subsequently, the data-extraction form was piloted on five randomly selected articles. Differences in extraction were discussed, clarified, and if necessary, elements were redefined. The final data-extraction form was applied to all articles included in the study (see **Table 1**).

In total, 62 elements were evaluated in each article, and were coded as "yes", "no", and if pertinent, "not applicable". The first two codes were used to calculate the percentage of

completeness of reporting for each element, as reference, the total number of articles for which the element was applicable was taken. Decisions whether elements could be "not applicable" were considered for each element separately. For example, the diagnostic criteria for cases and controls (element 5c), was applicable only for case-control studies and not applicable for other research designs. Not applicable was also assigned when a particular analysis was not performed in the study. For example, although it might be expected that internal validation and assessment of model fit would be undertaken in prediction studies, many studies did not perform internal validation and hence did not report this in the methods section. When they reported in the results section a particular analysis but did not report it in the methods section were assessed as "no", when they did not perform that particular analysis were assessed as "not applicable". This rationale was applied for elements assessing the reporting of internal and external validation, measures of model fit, predictive ability and reclassification. A detailed description of the elements based on GRIPS is available in **Supplementary Table 2**.

All procedures were carried out by two researchers independently: the pilot of the data-extraction by AI and ACJWJ, and the data extraction process and analysis by AI and RM.

RESULTS

Literature Search

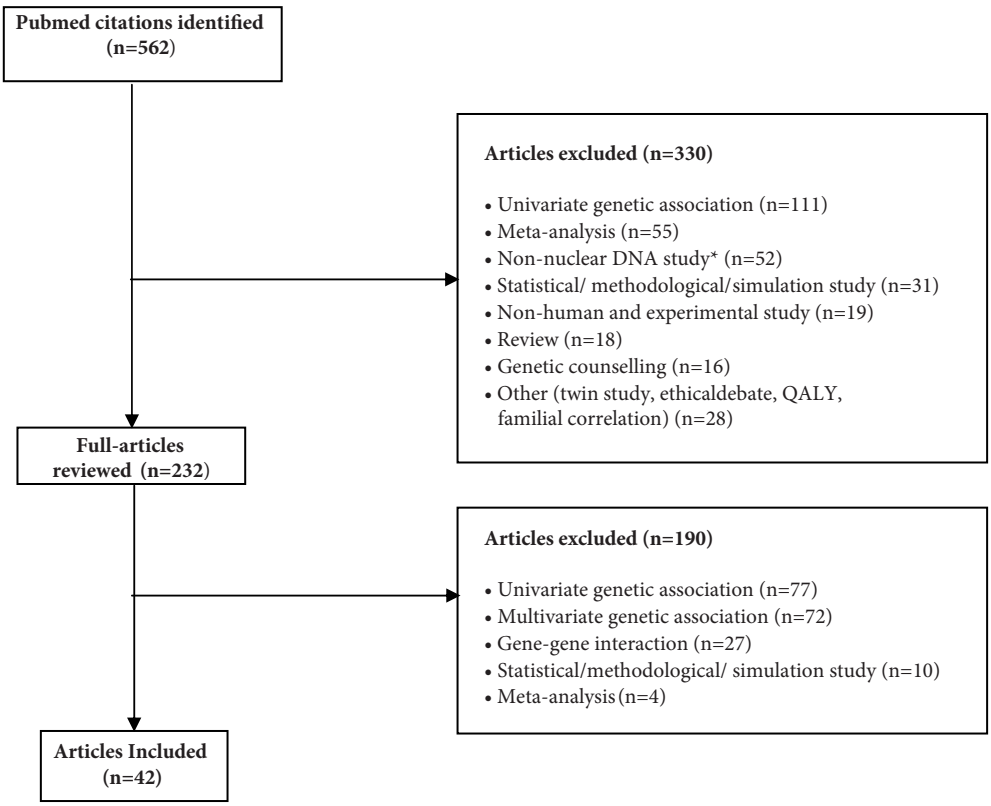
A total of 562 articles were identified through our search strategy of which 42 articles met the inclusion criteria (16-55). About half of the excluded articles were univariate and multivariate genetic association studies (n=260; 46.3%) (**Figure 1**).

Table 1 summarizes the completeness of reporting for each of the examined elements. Sixty-seven percent of articles stated in the title terms such as "risk assessment", "genetic risk score", "discriminatory accuracy" or "combined effects of genetic variants on disease risk" (see **Table 1)**, but only 31% used the GRIPS recommended keywords "genetic or genomic", "risk" and "prediction" in the title or abstract. **Table 2** shows that the included articles used the combination of the recommended keywords three times more often in the title and abstract compared to exclude articles (31% vs. 11 %). Included articles were found to use the keyword "prediction" four times more often in title (14% vs. 3%), and two times more often in title or abstract (31% vs. 15%).

All 42 articles mentioned the study design in the methods section and reported whether data were novel or previously collected. Sixty percent of the articles reported the periods of recruitment and/or follow-up, and 76% reported the eligibility criteria for the selection of study participants. Among the case-control studies (n=18), 56% stated the diagnostic criteria used for both cases and controls. Definition of the outcome was reported in all but one. Similarly, 93% of the articles reported the genetic variant included in the model with its respective rs number.

Reporting in Genetic Risk Prediction Studies

Figure1. Flowchart of article selection



*Studies that performed protein characterization, gene expression levels, mtDNA, RNA or miRNA analysis.

Table 1. Reporting of methodological characteristics for risk prediction studies in 2010 based on GRIPS checklist (n=42)

Section	Item	Sub-Item number and description		Reported		
				#	total applicable	%
Title and abstract		1a	Identifies article as a risk prediction study	28	42	67
		1b	Use recommended key words in the abstract: (genetic OR genomic) AND (risk) AND (prediction)	13	42	31
Introduction	Background and rationale	2a	States the role and number of variants discovered	39	42	93
		2b	Informs about models available or investigated	40	42	95
	Objectives	3a	Specifies the study objectives	42	42	100
		3b	States the specific model(s) that is/are investigated in the study	37	42	88
		3c	Specifies if the study concerns the development of the model(s), a validation effort, or both	39	42	93
Methods	Study design and settings	4a	Describes the study design and whether the data was *de novo* or previously collected	42	42	100
		4b	States the dates of data collection and/or follow up, when applicable	25	42	60
		4c	Describes setting and location	36	42	86
	Participants	5a	Describes eligibility criteria for participants	32	42	76
		5b	Describes sources of participants and/or methods of participants selection	34	42	81
		5c	For case-control studies, describes diagnostic criteria for both	10	18	56
			For patient-cohort studies, describes diagnostic criteria for cases	10	10	100
	Variables definition	6a	Defines non-genetic risk factors	35	36	97
		6b	**Defines outcome(s)**	41	42	98
		6c	**Defines genetic variants**	39	42	93
	Variables assessment	7a	Describes methods of assessment for each variable	24	36	67
		7b	Describes genotyping and laboratory methods	33	42	79
	Variables coding	**8a**	**Describes how genetic variants were handled in the analyses**	41	42	98
		8b	**Explains which genetic variant is considered as the risk variant**	31	42	74
		8c	Describes how quantitative variables were handled in the analyses	32	32	100

Section	Item	Sub-Item number and description		Reported		
				#	total applicable	%
		8d	If groupings were chosen describes whether these were bases on commonly used cutoffs	28	28	100
	Risk model construction	9a	Specifies the variables that were initially considered	41	42	98
		9b	Specifies the procedures followed for variables selection and specifies inclusion and deletion criteria, if applicable	26	28	93
		9c	Specifies model derivation for non-genetic variables, if applicable	36	36	100
		9d	Specifies model derivation for genetic variables	42	42	100
		9e	**Explains the horizon of the risk prediction, if applicable**	17	18	94
	Validation	10a	Describes internal validation	15	16	94
		10b	Describes external validation	8	9	89
	Missing data	11a	Specifies the percentage of missing values in their data	22	42	52
		11b	Specifies how missing data were handled in the analyses	27	42	64
	Statistical methods	12a	Describes measures of model fit	14	19	74
		12b	Describes measures of predictive ability	37	37	100±
		12c	Describes measures of reclassification	10	11	91
	Other	13a	Describes all subgroups that were examined	22	28	79
		13b	Describes why these subgroups were examined and whether they were based on a priori information or if analyses were exploratory, if done	11	28	39
		13b	If interaction effects were considered mentions which and why	9	17	53
Results	Participants	14a	Reports the number of individuals at each stage of the study	32	42	76
		14b	Reports the mains reasons for non-participation	19	42	45
		14c	Reports numbers of participants not genotyped	25	42	60
		14d	Reports the reasons for not genotyping	21	42	50
	Population	15a	Reports clinical and demographic characteristics of study population	37	42	88
		15b	Reports risk factors of study population included in the model	33	36	92

Section	Item	Sub-Item number and description		Reported		
				#	total applicable	%
	Model estimates	16a	Reports unadjusted associations between the variables and the outcome	28	42	67
		16b	**Reports adjusted estimates from the full risk model(s)**	31	42	74
		16c	**Reports the intercept (α) of the model, if applicable**	7	42	17
	Risk distributions	17	Reports distributions of predicted risks and/or risk scores	26	42	62
	Assessment	18a	Reports measures of model fit	16	19	84
		18b	Reports measures of predictive ability	37	37	100±
		18c	Reports measures of reclassification	11	11	100±
	Validation	19a	Reports internal validation	16	16	100±
		19b	Reports external validation	9	9	100±
	Other analyses	20a	Reports results of subgroups, interaction or exploratory analyses	37	37	100±
		20b	States the difference between exploratory and non-exploratory data	13	37	35
Discussion	Limitations	21	Discusses limitations of the study and their impact on the results of the study	36	42	86
	Interpretation	22	Gives an overall interpretation of results	42	42	100
	Generalizability	23a	Discusses the representativeness of study population within the target population	25	42	60
		23b	Discusses clinical relevance	28	42	67
		23c	Discusses what information is needed to adopt the proposed model	38	42	90
Other	Supplementary information	24	Describes how other authors can access the model protocols	7	42	17
	Funding	25a	Gives the source of funding and the role of the funders for the present study	35	42	83
		25b	States whether there are any conflicts of interest	36	42	86

Percentages are calculated based on the total number of articles for which the item was considered applicable. ± High percentages are due to the facts that calculations were based on the number articles that performed the analysis. Relevant elements for external validation of prediction studies are highlighted in **bold type**.

Table 2. Use of recommended keywords in title or abstract in articles retrieved with our search strategy, published in 2010

Keyword	Included articles n=42		Excluded articles n=520	
	Title n (%)	Title or abstract n (%)	Title n (%)	Title or abstract n (%)
Genetic	31 (74)	38 (90)	218 (42)	362 (70)
Genomic	5 (12)	10 (24)	20 (4)	50 (10)
Risk*	24 (57)	42 (100)	165 (32)	520 (100)
Prediction	6 (14)	13 (31)	14 (3)	76 (15)
Genetic or genomic	36 (86)	41 (98)	238 (46)	382 (73)
Articles that used the recommended combination of keywords: (genetic OR genomic) AND (risk) AND (prediction)	3 (7)	13 (31)	5 (1)	59 (11)

* Higher values were expected taking into account that in our search strategy "risk" (all fields) was used.

Table 3. Quality of reporting for general and specific items of other research designs

Guideline		STARD	STARD		STARD				CONSORT	CONSORT	CONSORT	CONSORT
	Article/Item	Siddiqui et al 2005	Coppus et al 2006*		Selman et al 2011*				Balasubramanian et al 2006	Agha et al 2007	Smith et al 2008	Rios et al 2008
Number of studies evaluated		**n=16**	**n=24 (1999)**	**n=27 (2004)**	**n= 85 pre 2004 gynecology**	**n=17 post 2004 gynecology**	**n= 85 pre 2004 obstetrics**	**n=17 post 2004 obstetrics**	**n=69**	**n=90**	**n=96**	**n=89**
General items	Identifies study type in title ¤	50%	29%	19%	65%	65%	19%	63%	NA	NA	24%	97%
	Use of recommended keywords in the abstract ¤				NA	NA	NA	NA	–	–	–	–
	Scientific background	NA	NA	NA	NA	NA	NA	NA	100%	100%	99%	100%
	Objectives	100%	88%	81%	64%	65%	94%	100%	100%	14%	99%	81%
	Describes inclusion and exclusion criteria of participants	100%†±	71%±	63%±	77.3%±	65%±	73%±	83%±	100%	34%	86%(incl) 56% (excl)§	96%
	Method of recruitment (how were participants enrolled)	81%	83%	96%	35%	94%	84%	91%	97%	34%	NA	NA
	Dates of data collection	94%†	79%	85%	86%	94%	63%	74%	97%	57%	36%	30%
	Periods of follow-up	–	–	–	–	–	–	–	98%	57%	72%	30%
	Definition of primary outcome	–	–	–	–	–	–	–	100%	24%	93%	35%
	Method of measurements	–	–	–	–	–	–	–	–	–	–	–
	Discusses the clinical applicability	100%	92%	96%	100%	100%	100%	97%	–	–	–	–
	States possible sources of bias¥	31%	4%	15%	22%	36%	63%	66%	68%	17%	82%	NA

Guideline	Article/Item	STARD	STARD		STARD				CONSORT	CONSORT	CONSORT	CONSORT
		Siddiqui et al 2005	Coppus et al 2006*		Selman et al 2011*				Balasubramanian et al 2006	Agha et al 2007	Smith et al 2008	Rios et al 2008
Number of studies evaluated		**n=16**	**n=24 (1999)**	**n=27 (2004)**	**n= 85 pre 2004 gynecology**	**n=17 post 2004 gynecology**	**n= 85 pre 2004 obstetrics**	**n=17 post 2004 obstetrics**	**n=69**	**n=90**	**n=96**	**n=89**
	Gives an overall interpretation of results	–	–	–	–	–	–	–	100%	96%	NA	NA
Specific items for STARD guideline	Definition of and/or rationale for the units, cut-off points, or categories of the results of the index test	75%†	67%	93%	97%	71%	96%	100%	–	–	–	–
	Reports any adverse event from performing the index test	¶	0%	4%	16%	18%	NA	NA	–	–	–	–
	Reports estimates of variability of diagnostic accuracy between readers, centers or subgroups of participants, if done	25%	17%	22%	0%	0%	51%	83%	–	–	–	–
Specific items for CONSORT guideline	Method of random sequence generation	–	–	–	–	–	–	–	42%	28%	49%	36%
	Method of allocation concealment	–	–	–	–	–	–	–	32%	29%	26%	24%
	Describes how blinding was achieved	–	–	–	–	–	–	–	1%	26%	1%	22%
	Gives sample size calculations	–	–	–	–	–	–	–	57%	29%	49%	36%
	Includes participant flow diagram	–	–	–	–	–	–	–	25%	51%	24%	29%

Guideline	Article/Item	STARD	STARD		STARD				CONSORT	CONSORT	CONSORT	CONSORT
		Siddiqui et al 2005	Coppus et al 2006*		Selman et al 2011*				Balasubramanian et al 2006	Agha et al 2007	Smith et al 2008	Rios et al 2008
Number of studies evaluated		**n=16**	**n=24 (1999)**	**n=27 (2004)**	**n= 85 pre 2004 gynecology**	**n=17 post 2004 gynecology**	**n= 85 pre 2004 obstetrics**	**n=17 post 2004 obstetrics**	**n=69**	**n=90**	**n=96**	**n=89**
Specific items for STROBE guideline	Describes any efforts to address potentially sources of bias	–	–	–	–	–	–	–	–	–	–	–
	Describes any methods used to examine subgroups and interaction	–	–	–	–	–	–	–	–	–	–	–
	Considers use of a flow diagram	–	–	–	–	–	–	–	–	–	–	–
	Describes any sensitivity analysis	–	–	–	–	–	–	–	–	–	–	–
Specific items for GRIPS guideline	Explains which genetic variant is considered as the risk variant	–	–	–	–	–	–	–	–	–	–	–
	Reports adjusted estimates from the full risk model	–	–	–	–	–	–	–	–	–	–	–
	Reports distributions of predicted risks and/or risks scores	–	–	–	–	–	–	–	–	–	–	–
	Reports the intercept (α) of the model	–	–	–	–	–	–	–	–	–	–	–

Table 3. Quality of reporting for general and specific items for other research designs.

Guideline		CONSORT		CONSORT	CONSORT	CONSORT	CONSORT	STROBE	STROBE	STROBE	STROBE	GRIPS
	Article/Item	**Han C et al 2009***		**Pandis et al 2010 †**	**Hopewell et al 2010**	**Kiehna et al 2011 †**	**Parsons et al 2011**		**Langan et al 2010**	**Papatha-nasiou et al 2010**	**Pooro-lajal et al 2011**	**This Study**
Number of studies evaluated		**n= 166 pre CONSORT**	**n=276 post CONSORT**	**n=95**	**n=616**	**n=27**	**n= 27**	**n= 73**	**n=138**	**n=244**	**n=60**	**n=42**
General items	Identifies study type¤	NA	NA	NA	33%	NA	78%	53%	87%	NA	75%	67%
	Use of recommended keywords in the abstract	–	–	–	–	–	–	–	93%	NA	–	29%
	Scientific background	96%	100%	99%	NA	100%	96%	97%	97%	NA	100%	95%
	Objectives	93%	100%	100%	NA	100%	78%	78%	89%	NA	98%	100%
	Describes inclusion and exclusion criteria of participants	98%	99%	98%	NA	100%	100%	93%	85%	97%	93%	76%
	Method of recruitment (how were participants enrolled)	98%	99%	94%	NA	NA	NA	63%	68%	NA	NA	81%
	Dates of data collection	14%	66%	41%	NA	93%	58%	63%	64%	NA	NA	60%
	Periods of follow-up	14%	66%	86%	NA	93%	58%	63%	64%	74%	93%	
	Definition of primary outcome	34%	75%	95%	53%	96%	85%	63%	91%	98%	100%	98%
	Method of measurements	–	–	–	–	–	–	NA	77%	82%	98%	67%
	Discusses the clinical applicability	–	–	–	–	–	–	–	–	–	–	67%

Guideline	Article/Item	CONSORT		CONSORT	CONSORT	CONSORT	CONSORT	STROBE	STROBE	STROBE	STROBE	GRIPS
		Han C et al 2009*		Pandis et al 2010 †	Hopewell et al 2010	Kiehna et al 2011 †	Parsons et al 2011		Langan et al 2010	Papatha-nasiou et al 2010	Pooro-lajal et al 2011	This Study
Number of studies evaluated		**n= 166 pre CONSORT**	**n=276 post CONSORT**	**n=95**	**n=616**	**n=27**	**n= 27**	**n= 73**	**n=138**	**n=244**	**n=60**	**n=42**
	States possible sources of bias¥	99%	100%	30%	NA	96%	63%	67%	55%	NA	100%	86%
	Gives an overall interpretation of results	69%	100%	18%	NA	97%	96%	82%	68%	NA	100%	100%
Specific items for STARD guideline	Definition of and/ or rationale for the units, cut-off points, or categories of the results of the index test	–	–	–	–	–	–	–	–	–	–	–
	Reports any adverse event from performing the index test	–	–	–	–	–	–	–	–	–	–	–
	Reports estimates of variability of diagnostic accuracy between readers, centers or subgroups of participants, if done	–	–	–	–	–	–	–	–	–	–	–
Specific items for CONSORT guideline	Method of random sequence generation	25%	66%	67%	34%	44%	50%	–	–	–	–	–
	Method of allocation concealment	15%	45%	48%	25%	37%	46%	–	–	–	–	–
	Describes how blinding was achieved	65%	36%	1%	41%	38%	44%	–	–	–	–	–

Guideline		CONSORT		CONSORT	CONSORT	CONSORT	CONSORT	STROBE	STROBE	STROBE	STROBE	GRIPS
	Article/Item	Han C et al 2009*		Pandis et al 2010 †	Hopewell et al 2010	Kiehna et al 2011 †	Parsons et al 2011		Langan et al 2010	Papatha-nasiou et al 2010	Pooro-lajal et al 2011	This Study
Number of studies evaluated		n= 166 pre CONSORT	n=276 post CONSORT	n=95	n=616	n=27	n= 27	n= 73	n=138	n=244	n=60	n=42
	Gives sample size calculations	40%	28%	66%	45%	34%	50%	–	–	–	–	–
	Includes participant flow diagram	56%	74%	47%	28%	96%	NA	–	–	–	–	–
Specific items for STROBE guideline	Describes any efforts to address potentially sources of bias	–	–	–	–	–	–	25%	32%	63%	15%	–
	Describes any methods used to examine subgroups and interaction	–	–	–	–	–	–	30%	22%	63%	30%	–
	Considers use of a flow diagram	–	–	–	–	–	–	8%	21%	2%	15%	–
	Describes any sensitivity analysis	–	–	–	–	–	–	14%	9%	55%	0%	–
Specific items for GRIPS guideline	Explains which genetic variant is considered as the risk variant	–	–	–	–	–	–	–	–	–	–	74%
	Reports adjusted estimates from the full risk model	–	–	–	–	–	–	–	–	–	–	74%

Guideline	CONSORT	CONSORT	CONSORT	CONSORT	CONSORT	STROBE	STROBE	STROBE	STROBE	GRIPS
Article/Item	Han C et al 2009*		Pandis et al 2010 †	Hopewell et al 2010	Kiehna et al 2011 †	Parsons et al 2011		Langan et al 2010	Papatha-nasiou et al 2010	Pooro-lajal et al 2011	This Study
Number of studies evaluated	**n= 166 pre CONSORT**	**n=276 post CONSORT**	**n=95**	**n=616**	**n=27**	**n= 27**	**n= 73**	**n=138**	**n=244**	**n=60**	**n=42**
Reports distributions of predicted risks and/or risks scores	–	–	–	–	–	–	–	–	–	–	62%
Reports the intercept (α) of the model	–	–	–	–	–	–	–	–	–	–	17%

Dash (–) indicates that the item is not stated in the guideline. NA, indicates Not Applicable when there is not information about that particular item in the study

¤ In case of RCT assessment of Title and Abstract was done based on the last update of the CONSORT statement 2010, as a consequence most of articles published before 2010 are fill as NA in this table. However, Smith et al and Rios et al explicitly evaluated if the title identify the study as a RCT, thus numbers are filled for those two articles.

¥ Possible sources of bias for STARD guideline includes how indeterminate and missing results were handled in the study.

* These articles aimed to compare the quality of reporting before and after publication of the corresponding guideline

† Includes articles reported as YES + Partial or Adequate + Inadequate

± For STARD guideline this item evaluates: inclusion and exclusion criteria, setting and location where the data were collected. Here setting makes reference to whether the participants were recruited from primary, secondary or tertiary setting

§ In this study inclusion (incl) and exclusion (excl) criteria of participants were evaluated separately

¶ For this particular item in this study all 16 studies were scored as NA

Search strategy: (CONSORT) (Ti/Ab) OR STROBE(Ti/Ab) OR STARD(Ti/Ab) OR REMARK(Ti/Ab)) AND (quality(Ti/Ab) OR transparency(Ti/Ab) OR completeness(Ti/Ab)) AND reporting(Ti/Ab) AND medline(sb)

Reviewed of these articles was carried out by two researchers independently (AI and RM). Any disagreement was resolved through consensus.

Forty-one out of 42 articles mentioned how genetic variants were handled in the analysis, as dominant/recessive, per allele, per genotype or haplotype effect, but 26% of the articles did not mention which allele was the risk allele. Eighteen articles had longitudinal data, of which all but one explained the time horizon of the risk prediction.

Internal validation was performed only in 16 (38%) and external validation only in 9 (21%) studies. Among these, the great majority described in the methods sections how internal (94%; 15/16) or external (89%; 8/9) validation was performed. All articles that calculated the predictive ability of the model (n=37), also described the measures used. Less than half of the articles addressed model fit (19/42, 45%) and only 11 (26%) assessed reclassification. Seventy-four per cent (14/19) of articles that assessed model fit and 91% (10/11) of articles that assessed reclassification described the metrics used in methods section.

Results of unadjusted and adjusted association analysis, such as beta coefficients or odds ratio, were reported in 67% and 74% articles, respectively. In total only 48% of the 42 articles reported the results from both the unadjusted and adjusted association analyses. The intercept of the model was reported only in 17% of the articles that reported regression modelling. The distributions of the predicted risk and/or risk scores were reported in 62%.

To perform external validation of the prediction model, specific information is needed, including the definition of outcome and genetic variables (6b-c), the handling of genetic variants in the analyses and the definition of the risk variant (8a-b), the horizon of the risk prediction (9e), the intercept and adjusted estimates from the full risk model (16b-c). Only 10% (4/42) of the articles fully reported these seven items that would enable replication or application of the model by the reader.

Regarding the interpretation of results, 77% of articles discussed the clinical relevance of their findings; 86% reported the limitations of the study and their impact on the results, while only 60% discussed the representativeness of the study population within the target population.

DISCUSSION

In this study we evaluated the completeness of reporting in genetic risk prediction studies based on the 25 items proposed by the GRIPS statement in 2011. In total we identified 42 articles published in 2010 that empirically investigated genetic risk prediction. We found that the identification of genetic risk prediction studies in public bibliographic databases is hindered by inconsistencies in terminology. In addition, there is lack of key information necessary to understand fully the studies, enable comparison between them, and to replicate and apply the prediction models in other populations.

Three methodological issues may impact the interpretation of our results. First, we have included in this review only those genetic risk prediction studies that mentioned the development or validation of a prediction model in the abstract. GWAS that did not mention genetic risk prediction analysis in the abstract were not included but might have included as a secondary analysis the evaluation of predictive performance of a genetic risk score. GWAS and in particular meta-analyses thereof may become a prime setting where genetic risk prediction models are developed and validated in the future and their reporting would need to be meticulous and detailed enough to make this information transparent and usable rather than a mere short adjunct to long series of analyses typically presented in these papers. Furthermore, we did not include articles that constructed genetic risk scores to assess the variance explained by the genes or to elucidate the genetic architecture of the disease. It is likely that the reporting of such studies is less complete than in the analysed papers, since the predictive modelling is typically only a small part of the paper.

Second, our search strategy used multiple terms to capture the three defining aspects of genetic risk prediction studies, which include the keywords recommended by the GRIPS statement. Percentages showing the use of the recommended keywords in all 562 articles can be potentially biased by the choice of this search strategy (see **Table 2**). For example, 100% of our retrieved articles mentioned "risk" in their title or abstract, but this is not surprising as "risk" was included in our search strategy. Nevertheless, what can be concluded from the table is that the keyword "genetic" is more commonly used than "genomic", and that less than a third (31%) of the studies used the keyword "prediction"; instead, other words such as "predict" (12%), "predictive" (10%), or "predicted" (7%) were used (data not shown).

Third, some elements required a somewhat subjective evaluation or led us to make a judgement on the basis of indirect information. For example, whilst element 4a was reported in 100% of the articles, a strict evaluation of this element, which assessed if the data were novel or previously collected, indicated that only 52% of the articles explicitly specified when the data was collected. However, we scored a paper as "yes" when it mentioned, for example, that individuals included were part of a previous described cohort, case-control or population based study. We are aware that assessment based on indirect information could lead to an overestimation of the completeness of reporting of some elements.

It is important to stress that we only assessed the reporting and did not evaluate the methodological quality of genetic risk prediction studies (**Supplementary table 2**). We did not assess, for example, whether the authors correctly applied the measures of model performance because we only evaluated whether the articles reported or not the measures of model fit or predictive ability.

One of the specific aspects of reporting concerns the use of keywords that facilitate the retrieval of genetic risk prediction studies. Only 31% of the assessed articles used the combination "(genetic OR genomic) AND (risk) AND (prediction)" in the title or abstract. Although the keyword (prediction) was more frequently found in our selection of 42 articles compared to the excluded articles (**Table 2**), it is still uncommonly used by authors of genetic risk prediction studies. Interestingly, while 69% of the included articles did not use the word "prediction", words such as "predictive" or "predicted" were used instead. Increased use of the recommended key words is essential to facilitate identification of genetic risk prediction studies across thousands of publications, and thus facilitate the dissemination of this new genomic knowledge.

We found that the quality of reporting is good for general elements of study design and analysis, but that details are lacking about elements that characterize genetic risk prediction studies. General elements refer to relevant information that should be reported for all kind of study designs, such as the background, objectives, key aspects of study design, and discussion of results. However, reporting of elements that evaluate information specifically used in genetic risk prediction studies, such as the definition of the risk allele(s) and construction of the model, was less complete. This appears to be in line with experience for other reporting guidelines including CONSORT for clinical trials (56-64), STARD for diagnostic studies (65-67), and STROBE for observational studies (64, 68-70).We reviewed these 17 papers, which fulfilled the following inclusion criteria: 1) they evaluated the quality of reporting based on CONSORT, STROBE, STARD or REMARK guidelines, 2) mentioned the results from general and specific elements, and 3) addressed broad topics, such as cancer, surgery or psychiatric. We found that reporting for general elements was comprehensive, but for design-specific elements, such as method of allocation concealment or blinding for randomized control trials, adverse reaction reporting or variability of diagnostic accuracy for studies of diagnostic accuracy, or assessing risk of bias or sensitivity analyses in observational studies, the completeness of reporting was poor (**Table 3**).

Key information needed to perform external validation of genetic risk prediction studies and apply them in real life constitutes part of the design-relevant and specific elements. In order to perform external validation of the model or apply it, it is essential to have clear and complete reporting of the seven elements mentioned before on the results section (see bold type elements on **Table1**). Among the evaluated articles, only four studies reported all seven elements. Most articles (83%) did not report the intercept of the model, and 26% of articles did not provide information about the genetic variant considered as the risk variant. Based on these results, we conclude that the information reported prior to the GRIPS statement (5) by genetic risk prediction studies is insufficient to replicate their models and, therefore, to perform external validation.

The GRIPS statement was conceived in response to anticipated variable completeness of reporting of genetic risk prediction studies. We observed that the completeness of reporting of studies published in 2010, prior to the GRIPS statement, was suboptimal, with 90% of our sample of 42 articles presenting insufficient information about the risk models. This incomplete reporting renders the accurate assessment and the replication of their analyses difficult or even impossible. A clear and complete report of key information is central for the evaluation and comparison of genetic risk prediction studies and the possible future application in a clinical or public health scenario. We suggest that the GRIPS statement is a tool through which adequate reporting can be facilitated, and therefore recommend the adoption of the GRIPS statement by authors of genetic risk prediction studies.

REFERENCES

2. Guttmacher AE, and FS Collins, **Realizing the promise of genomics in biomedical research.** *JAMA*, 2005;**294**(11):1399-402.
3. Ng P.C, et al., **An agenda for personalized medicine.** *Nature*, 2009;**461**(7265):724-6.
4. Hamburg MA, and FS Collins, **The path to personalized medicine.** *N Engl J Med,* 2010;**363**(4): 301-4.
5. Janssens AC, and CM van Duijn, **Genome-based prediction of common diseases: methodological considerations for future research.** *Genome Med*, 2009;**1**(2):20.
6. Janssens AC, et al., **Strengthening the reporting of genetic risk prediction studies (GRIPS): explanation and elaboration.** *Eur J Hum Genet,* 2011.
7. Simera I, et al., **A catalogue of reporting guidelines for health research.** *Eur J Clin Invest*, 2010;**40**(1): 35-53.
8. Bossuyt PM, et al., **Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative.** *BMJ,* 2003;**326**(7379):41-4.
9. McShane LM, et al., **Reporting recommendations for tumor marker prognostic studies (REMARK).** *J Natl Cancer Inst*, 2005;**97**(16):1180-4.
10. von Elm E, et al., **The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies.** *Plos Medicine*, 2007;**4**(10):1623-1627.
11. Little J, et al., **STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement.** *PLoS Med*, 2009;**6**(2):e22.
12. Janssens AC, et al., **Strengthening the reporting of genetic risk prediction studies: the GRIPS statement.** *Eur J Hum Genet*, 2011.
13. Smidt N, et al., **The quality of diagnostic accuracy studies since the STARD statement: has it improved?** *Neurology*, 2006;**67**(5):792-7.
14. Mallett S, et al., **Reporting of prognostic studies of tumour markers: a review of published articles in relation to REMARK guidelines.** *Br J Cancer*, 2010;**102**(1):173-80.
15. Moher D, et al., **Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation.** *JAMA*, 2001;**285**(15):1992-5.
16. Wilczynski NL, **Quality of reporting of diagnostic accuracy studies: no change since STARD statement publication—before-and-after study.** *Radiology*, 2008;**248**(3):817-23.
17. Anderson JL, et al., **Joint effects of common genetic variants from multiple genes and pathways on the risk of premature coronary artery disease.** *Am Heart J*, 2010;**160**(2):250-256 e3.
18. Horstmann S, et al., **Polymorphisms in GRIK4, HTR2A, and FKBP5 show interactive effects in predicting remission to antidepressant treatment.** *Neuropsychopharmacology*, 2010;**35**(3):727-40.
19. Baca-Garcia E, et al., **Nucleotide variation in central nervous system genes among male suicide attempters.** *Am J Med Genet B Neuropsychiatr Genet*, 2010;**153B**(1):208-13.
20. Brugts JJ, et al., **Genetic determinants of treatment benefit of the angiotensin-converting enzyme-inhibitor perindopril in patients with stable coronary artery disease.** *Eur Heart J*, 2010;**31**(15): 1854-64.
21. Carayol J, et al., **Assessing the impact of a combined analysis of four common low-risk genetic variants on autism risk.** *Mol Autism*, 2010;**1**(1):4.
22. Comen E, et al., **Discriminatory accuracy and potential clinical utility of genomic profiling for breast cancer risk in BRCA-negative women.** *Breast Cancer Res Treat*, 2011;**127**(2):479-87.

23. Davies RW, et al., **Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies.** *Circ Cardiovasc Genet*, 2010;**3**(5):468-74.

24. de Miguel-Yanes JM, et al., **Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms.** *Diabetes Car*e, 2011;**34**(1): 121-5.

25. Dickinson AM, et al., **Impact of genomic risk factors on outcome after hematopoietic stem cell transplantation for patients with chronic myeloid leukemia.** *Haematologica*, 2010;**95**(6):922-7.

26. Fontaine-Bisson B, et al., **Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population.** *Diabetologia*, 2010;**53**(10):2155-62.

27. Haritunians T, et al., **Genetic predictors of medically refractory ulcerative colitis.** *Inflamm Bowel Dis*, 2010;**16**(11):1830-40.

28. Hecker LA, et al., **Genetic control of the alternative pathway of complement in humans and age-related macular degeneration.** *Hum Mol Genet*, 2010;**19**(1):209-15.

29. Helfand BT, et al., **Genetic prostate cancer risk assessment: common variants in 9 genomic regions are associated with cumulative risk.** *J Urol*, 2010;**184**(2):501-5.

30. Karlson EW, et al., **Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk.** *Ann Rheum Dis*, 2010;**69**(6):1077-85.

31. Lu Y, et al., **Exploring genetic determinants of plasma total cholesterol levels and their predictive value in a longitudinal study.** *Atherosclerosis*, 2010;**213**(1):200-5.

32. Mealiffe ME, et al., **Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information.** *J Natl Cancer Inst*, 2010;**102**(21):1618-27.

33. Okser S, et al., **Genetic variants and their interactions in the prediction of increased pre-clinical carotid atherosclerosis: the cardiovascular risk in young Finns study.** *PLoS Genet*, 2010;**6**(9).

34. Palmer CN, et al., **Combined effect of inflammatory gene polymorphisms and the risk of ischemic stroke in a prospective cohort of subjects with type 2 diabetes: a Go-DARTS study.** *Diabetes,* 2010;**59**(11):2945-8.

35. Wang P, et al., **Using genetic and clinical factors to predict tacrolimus dose in renal transplant recipients.** *Pharmacogenomics*, 2010;**11**(10): 1389-402.

36. Paynter NP, et al., **Association between a literature-based genetic risk score and cardiovascular events in women.** *JAMA*, 2010;**303**(7):631-7.

37. Peterson RE, et al., **Genetic risk sum score comprised of common polygenic variation is associated with body mass index.** *Hum Genet*, 2011;**129**(2):221-30.

38. Qi Q, et al., **Combined effects of 17 common genetic variants on type 2 diabetes risk in a Han Chinese population.** *Diabetologia*, 2010;**53**(10):2163-6.

39. Renstrom F, et al., **Genetic predisposition to long-term nondiabetic deteriorations in glucose homeostasis: Ten-year follow-up of the GLACIER study.** *Diabetes*, 2011;**60**(1):345-54.

40. Ripatti S, et al., **A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses.** *Lancet*, 2010;**376**(9750):1393-400.

41. Rotger M, et al., **Impact of single nucleotide polymorphisms and of clinical risk factors on new-onset diabetes mellitus in HIV-infected individuals.** *Clin Infect Dis*, 2010;**51**(9):1090-8.

42. Ruchat SM, et al., **Combining genetic markers and clinical risk factors improves the risk assessment of impaired glucose metabolism.** *Ann Med*, 2010;**42**(3):196-206.

43. Ruiz JR, et al., **Can we identify a power-oriented polygenic profile?** *J Appl Physiol*, 2010;**108**(3):561-6.

44. Seshadri S, et al., **Genome-wide analysis of genetic loci associated with Alzheimer disease.** *JAMA*, 2010;**303**(18):1832-40.

45. Soranzo N, et al., **Common variants at 10 genomic loci influence hemoglobin A(C) levels via glycemic and nonglycemic pathways.** *Diabetes*, 2010;**59**(12):3229-39.

46. Sun J, et al., **Inherited genetic markers discovered to date are able to identify a significant number of men at considerably elevated risk for prostate cancer.** *Prostate*, 2011;**71**(4):421-30.

47. Takahashi H, et al., **Prediction model for knee osteoarthritis based on genetic and clinical information.** *Arthritis Res Ther*, 2010;**12**(5):R187.

48. Talledo M, et al., **Evaluation of host genetic and viral factors as surrogate markers for HTLV-1-associated myelopathy/tropical spastic paraparesis in Peruvian HTLV-1-infected patients.** *J Med Virol*, 2010;**82**(3):460-6.

49. Talmud PJ, et al., **Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study.** *BMJ*, 2010;**340**:b4838.

50. Usvasalo A, et al., **Prognostic classification of patients with acute lymphoblastic leukemia by using gene copy number profiles identified from array-based comparative genomic hybridization data.** *Leuk Res*, 2010;**34**(11):1476-82.

51. Uusitupa MI, et al., **Impact of positive family history and genetic risk variants on the incidence of diabetes: the Finnish Diabetes Prevention Study.** *Diabetes Care*, 2011;**34**(2): 418-23.

52. van der Helm-van Mil AH, RE Toes, and TW Huizinga, **Genetic variants in the prediction of rheumatoid arthritis.** *Ann Rheum Dis*, 2010;**69**(9):1694-6.

53. Wacholder S, et al., **Performance of common genetic variants in breast-cancer risk models.** *N Engl J Med*, 2010;**362**(11):986-93.

54. Wang J, et al., **Identification of undiagnosed type 2 diabetic individuals by the Finnish diabetes risk score and biochemical and genetic markers: a population-based study of 7232 Finnish men.** *J Clin Endocrinol Metab*, 2010;**95**(8):3858-62.

55. Ward K, et al., **Validation of DNA-based prognostic testing to predict spinal curve progression in adolescent idiopathic scoliosis.** *Spine* (Phila Pa 1976), 2010;**35**(25):E1455-64.

56. Zheng W, et al., **Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women.** *J Natl Cancer Inst*, 2010;**102**(13):972-81.

57. Balasubramanian SP, et al., **Standards of reporting of randomized controlled trials in general surgery: can we do better?** *Ann Surg*, 2006;**244**(5):663-7.

58. Agha R, D Cooper, and G Muir, **The reporting quality of randomised controlled trials in surgery: a systematic review.** *Int J Surg*, 2007;**5**(6):413-22.

59. Rios LP, et al., **Quality of reporting of randomized controlled trials in general endocrinology literature.** *J Clin Endocrinol Metab*, 2008;**93**(10):3810-6.

60. Smith BA, et al., **Quality of reporting randomized controlled trials (RCTs) in the nursing literature: application of the consolidated standards of reporting trials (CONSORT).** *Nurs Outlook*, 2008;**56**(1): 31-37.

61. Han C, et al., **The impact of the CONSORT statement on reporting of randomized clinical trials in psychiatry.** *Contemp Clin Trials*, 2009;**30**(2):116-22.

62. Pandis N, A Polychronopoulou, and T Eliades, **An assessment of quality characteristics of randomised control trials published in dental journals.** *J Dent*, 2010;**38**(9):713-21.

63. Hopewell S, et al., **The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed.** *BMJ*, 2010;**340**:c723.

64. Kiehna EN, et al., **Standards for reporting randomized controlled trials in neurosurgery.** *J Neurosurg*, 2011;**114**(2):280-5.

65. Parsons NR, et al., **A systematic survey of the quality of research reporting in general orthopaedic journals.** *J Bone Joint Surg Br*, 2011;**93**(9):1154-9.

66. Siddiqui MA, A Azuara-Blanco, and J Burr, **The quality of reporting of diagnostic accuracy studies published in ophthalmic journals.** *Br J Ophthalmol*, 2005;**89**(3):261-5.

67. Coppus SF, et al., **Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative.** *Fertil Steril*, 2006;**86**(5):1321-9.

68. Selman TJ, et al., **The quality of reporting of primary test accuracy studies in obstetrics and gynaecology: application of the STARD criteria.** *BMC Womens Health*, 2011;**11**:8.

69. Langan S, et al., **The reporting of observational research studies in dermatology journals: a literature-based study.** *Arch Dermatol*, 2010;**146**(5):534-41.

70. Papathanasiou AA, and E Zintzaras, **Assessing the quality of reporting of observational studies in cancer.** *Ann Epidemiol*, 2010;**20**(1):67-73.

71. Poorolajal J, et al., **Quality of Cohort Studies Reporting Post the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement.** *Epidemiol Health*, 2011;**33**: e2011005.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Data extraction form based on the 25 items of the GRIPS statement.

Section	Sub-section	# Element	Element description	Score
Title and Abstract	Title and Abstract	1a	Identified as a study of risk prediction	Y/N
		1b	Use recommended key words in the abstract: (genetic OR genomic) AND (risk) AND (prediction)	Y/N
Introduction	Bakground and rationale	2a.1	States the role of genetic variants	Y/N
		2a.2	States the number of variants discovered	Y/N
		2b	Informs about models available or investigated	Y/N
	Objectives	3a	Specifies the study objectives	Y/N
		3b	State the specific model(s) that is/are investigated in the study	Y/N
		3c	Specifies if the study concerns to development or/and validation of the model (s)	Y/N
Methods	Study design and settings	4a.1	Specifies the type of study design (case-control, cohort, cross sectional)	Y/N
		4a.2	Data collection de novo or previously collected	Y/N
		4b.1	Specifies periods of recruitment/years	Y/N
		4b.2	Specifies periods of follow-up	Y/N/NA
		4c.1	Describes setting	Y/N
		4c.2	Describes locations	Y/N
	Participants	5a	Describes eligibility criteria for participants	Y/N
		5b.1	Describes sources of participants	Y/N
		5b.2	Describes methods of participants selection	Y/N
		5c.1	Defines diagnostic criteria for case-control studies	Y/N/NA
		5c.2	Defines diagnostic criteria for patient-cohort studies	Y/N/NA
	Variables definition	6a	Defines non-genetic risk factors	Y/N/NA
		6b	Defines outcome(s)	Y/N
		6c	Defines genetic variants	Y/N

Section	Sub-section	# Element	Element description	Score
	Variables assessment	7a	Describes methods of assessment for each variable	Y/N/NA
		7b	Describes genotyping and laboratory methdos	Y/N
	Variables coding	8a	Describes how genetic variants were handled in the analyses	Y/N
		8b	Explains which genetic variant is considered as the risk variant	Y/N
		8c	Describes how quantitative variables were handled in the analyses	Y/N/NA
		8d	Describes for groupings whether were chosen based on commonly used cut-offs or not	Y/N/NA
	Analysis risk model construction	9a	Specifies the variables that were initially considered	Y/N
		9b	Specifies the procedures followed for the final selection and specifies inclusion and deletion criteria for the variables, if applies	Y/N/NA
		9c	Model derivation for non-genetic variables	Y/N/NA
		9d	Model derivation for genetic variants	Y/N
		9e	Explains the horizon of the risk prediction	Y/N/NA
	Validation	10a	Describe internal validation	Y/N/NA
		10b	Describe external validation	Y/N/NA
	Missing data	11a	Specifies the percentage of missing values in their data	Y/N
		11b	Specifies how missing data were handled in the analyses	Y/N
	Statistical methods	12a	Describes measures of model fit	Y/N/NA
		12b	Describes measures of predictive ability	Y/N/NA
		12c	Describes measures of reclassification	Y/N/NA
	Other	13a	Describes all subgroups that were examined	Y/N/NA
		13b	Describes all subgroups that were examined and whether these subgroups were based on a priori or in an exploratory way	Y/N/NA
		13c	Describes if interaction effects were considered	Y/N/NA
Results	Participants	14a	Reports the number of individuals at each stage of the study	Y/N
		14b	Reports the mains reasons for non-participation	Y/N
		14c	Reports numbers of participants not genotyped	Y/N

Section	Sub-section	# Element	Element description	Score
		14d	Describes reasons for not genotyping	Y/N
	Population	15a	Reports clinical and demographic characteristics of study population	Y/N/NA
		15b	Reports risk factors of study population included in the model	Y/N
	Model estimates	16a	Reports unadjusted associations between the variables and the outcome	Y/N
		16b	Reports adjusted associations between the variables and the outcome	Y/N
		16c	Reports the intercept (α) of the model	Y/N
	Risk distributions	17	Reports distributions of predicted risks and/or risk scores	Y/N
	Assessment	18a	Reports measures of model fit	Y/N/NA
		18b	Reports measures of predictive ability	Y/N/NA
		18c	Reports measures of reclassification	Y/N/NA
	Validation	19a	Reports internal validation	Y/N/NA
		19b	Reports external validation	Y/N/NA
	Other analyses	20a	Reports results of subgroups, interaction or exploratory analyses	Y/N/NA
		20b	States the difference between exploratory and non-exploratory data is clear, if applies	Y/N/NA
Discussion	Limitations	21a	Discusses limitations of the study	Y/N
		21b	Discusses their impact on the results of the study	Y/N
	Interpretation	22	Gives an overall interpretation of results	Y/N
	Generalizability	23a	Discusses the representativeness of study population with the target population	Y/N
		23b	Discusses clinical relevance	Y/N
		23c	Discusses what information is needed to adopt the proposed model	Y/N
Other information	Supplementary information	24	Describes how other authors can access the model or protocols	Y/N
		25a	Gives the source of funding and the role of the funders for the present study	Y/N
		25b	States whether there are any conflicts of interest	Y/N

Supplementary Table 2. Description of the elements evaluated based on the 25 items of the GRIPS statement.

Section	Sub-section		Elements and sub-elements		Description of elements and sub-elements
Title and abstract	Title and abstract	1a	Identifies article as a risk prediction study	1a	"**YES**" is filled in when the following terms are used: risk assessment, genetic risk score, combined effects of SNP on disease risk. "**NO**" is filled in when the following terms are used: association, associated, and genetic predisposition.
		1b	Use recommended key words in the abstract: (genetic or genomic) and (risk) and (prediction)	1b	**"YES**" is filled in when [genetic or genomic] AND [risk] AND [prediction] are used in title or abstract.
Introduction	Background and rationale	2a	States the role and number of variants discovered	2a.1	States the role of genetic variants: when it can be concluded that the outcome has a genetic contribution or predisposition, the score **"YES**" is given.
				2a.2	States the number of variants discovered: the score **"YES**" is given when a specific number of variants, or gene names, or an approximation (e.g. >10 variants, many variants or multiple loci) are mentioned.
		2b	Informs about models available or investigated	2b	**"YES**" is filled in when it is clear that in previous studies genetic, clinical or both models have been proposed for the outcome of interest, or conversely, thatno models have been yet proposed. **"NO**" is filed in when this is not reported.
	Objectives	3a	Specifies the study objectives	3a	**"YES**" is filled in when the aim of the article is mentioned.
		3b	States the specific model(s) that is/ are investigated in the study	3b	**"YES**" is filled in when it can be concluded that the study investigates a clinical, genetic or both models.
		3c	Specifies if the study concerns the development of the model(s), a validation effort, or both	3c	**"YES**" is filled in when it can be concluded that the study concerns the development and/or validation of the model(s).
Methods	Study design and settings	4a	Describes the study design and whether the data was *de novo* or previously collected	4a.1	Describes the study design: when the type of study is mentioned, the score**"YES**" is given (e.g. cohort, case-control or cross-sectional study).

Section	Sub-section		Elements and sub-elements		Description of elements and sub-elements
				4a.2	Specifies whether the clinical data were de novo or previously collected: **"YES"** is filled in when is explicitly mentioned that clinical data were de novo or previously collected and when is mentioned, for example, that individuals included in the analyses were part of a previous described cohort, case-control or population based study.
		4b	States the dates of data collection and/or follow up, when applicable	4b.1	Dates of data-collection: the score **"YES"** is given when the study mentions the time period between the start and end of data-collection.
				4b.2	Duration of follow-up: the score **"YES"** is given when the study mentions the follow-up period. **"NA"** is filled in when the design of the study is case-control or cross-sectional.
		4c	Describes setting and location	4c.1	Describes the setting: **"YES"** is filled in when participant recruitment sites are mentioned (e.g. hospitals, outpatient clinics, screening centers or registries).
				4c.2	Describes the location: **"YES"** is filled in when the countries, regions or cities where the investigation took place are mentioned.
	Participants	5a	Describes eligibility criteria for participants	5a	**"YES"** is filled in when inclusion and exclusion criteria of participants are mentioned (e.g. age, sex, ancestry, ethnicity and/or geographical region of the participants, specific risk factors, and for cohorts of patients, diagnosis, disease duration or stage, and co-morbidity).
		5b	Describes sources of participants and/or methods of participants selection	5b.1	Sources of participants: when the populations from which the participants were selected are mentioned, the score **"YES"** is given (e.g. general population, population at risk, disease population).
				5.b.2	Methods of participants selection: the score **"YES"** is given when the methods of participants selection are mentioned (e.g. participants randomly invited, referred or self-selected).
		5c	For case-control studies, describes diagnostic criteria for both	5c.1	Describes diagnostic criteria for case-control studies: the score **"YES"** is given when the diagnostic criteria used for selecting cases and controls are mentioned. **"NA"** is filled in for cross-sectional or cohort studies.
			For patient-cohort studies, describes diagnostic criteria for cases	5c.2	Describes diagnostic criteria for patient-cohort studies: the score **"YES"** is given when the diagnostic criteria used to select cases from a patient-cohort study are mentioned. **"NA"** is filled in for cross-sectional or case-control studies.

Section	Sub-section		Elements and sub-elements		Description of elements and sub-elements
	Variables definition	6a	Defines non-genetic risk factors	6a	**"YES"** is filled in when non-genetic risk factors are mentioned (e.g. clinical characteristics as BMI, BP, family history etc. or laboratory measurements as cholesterol, glycaemia levels). **"NA"** is filled in when the prediction model does not include non-genetic risk factors.
		6b	Defines outcome(s)	6b	**"YES"** is filled in when the outcome(s) are mentioned. For diseases,a definition of the outcome according to diagnostic criteria should be mentioned .
		6c	Defines genetic variants	6c	**"YES"** is filled in when genetic variants are mentioned (e.g. SNPs, genotype, haplotype, insertions/deletions or CNV). All SNPs should be reported with their respective rs number, otherwise the score **"NO"** is given.
	Variables assessment	7a	Describes methods of assessment for each variable	7a	**"YES"** is filled in when procedures or questionnaires that were used are mentioned (e.g. how blood pressure was taken or which questionnaires were used). In case the methods have been previously published, a reference should be provided. **"NA"** is filled in when the prediction model does not include non-genetic risk factors.
		7b	Describes genotyping and laboratory methods	7b	**"YES"** is filled in when enough information to replicate the procedure is provided.
	Variables coding	8a	Describes how genetic variants were handled in the analyses	8a	**"YES"** is filled in when the handling of genetic variants is mentioned (e.g. using dominant/ recessive genetic model, per allele effect, per genotype categories or per haplotypes).
		8b	Explains which genetic variant is considered as the risk variant	8b	**"YES"** is filled in when the genetic variant considered as the risk variant is mentioned. The main article, tables and supplementary data were checked to assess this item.
		8c	Describes how quantitative variables were handled in the analyses	8c	**"YES"** is filled in when the handling of quantitative variables in the analysis is mentioned (e.g. as continuous or categorized variables). The main article, tables and supplementary data were checked to assess this item. **"NA"** is filled in when the prediction model does not include non-genetic risk factors.
		8d	If groupings were chosen describes whether these were bases on commonly used cut-offs	8d	**"YES"** is filled in when common or new cut-offs of categorized variables are mentioned. **"NA"** is filled in when the study does not use categorical variables.

Section	Sub-section		Elements and sub-elements		Description of elements and sub-elements
	Risk model construction	9a	Specifies the variables that were initially considered	9a	**"YES"** is filled in when the variables initially considered in the model are mentioned.
		9b	Specifies the procedures followed for variables selection and specifies inclusion and deletion criteria, if applicable	9b	**"YES"** is filled in when the procedures and criteria used for variables selection are mentioned (e.g. backward deletion or forward inclusion). Criteria should be mentioned in terms of p-values. **"NA"** is filled in when all variables initially considered in the model are part of the final model.
		9c	Specifies model derivation for non-genetic variables, if applicable	9c	**"YES"** is filled in when the statistical model chosen for non-genetic variables is mentioned (e.g., logistic regression analysis, Cox proportional hazards regression, risk score, machine learning techniques or Bayesian model averaging). When the study applies an already existing and recognized model, it should simply be stated which model. **"NA"** is filled in when the prediction model does not include non-genetic risk factors.
		9d	Specifies model derivation for genetic variables	9d	**"YES"** is filled in when the statistical model chosen for genetic variables is mentioned (e.g., logistic regression analysis, Cox proportional hazards regression, genetic risk score, machine learning techniques or Bayesian model averaging).
		9e	Explains the horizon of the risk prediction	9e	**"YES"** is filled in when the horizon of the risk prediction is mentioned (e.g. 5-year or lifetime risk). **"NA"** is filled in when the study has a cross-sectional design.
	Validation	10a	Describes internal validation	10a	**"YES"** is filled in when the procedures for internal validation are mentioned in the methods section. Internal validation was defined as the reassessing of the performance of the model in the same population. Procedures such cross-validation, bootstrapping techniques or split sample are used for that purpose. **"NO"** is filled in when the study performs internal validation analysis but does not mention which procedure was used for that purpose. **"NA"** is filled in when no internal validation analysis was performed.

Section	Sub-section	Elements and sub-elements		Description of elements and sub-elements
		10b Describes external validation	10b	**"YES"** is filled in when the procedures for external validation are mentioned in the methods section. External validation was defined as the reassessing of the performance of the model in a independent sample. Procedures such cross-validation, bootstrapping techniques or split sample are used for that purpose. **"NO"** is filled in when external validation analysis was performed but the procedure used for that purpose is not mentioned. **"NA"** is filled in when no external validation analysis was performed.
	Missing data	11a Specifies the percentage of missing values in their data	11a	**"YES"** is filled in when the percentage of missing data in the data is reported.
		11b Specifies how missing data were handled in the analyses	11b	**"YES"** is filled in the method used to deal with the missing data is mentioned (e.g. complete case analysis, imputation, reweighting).
	Statistical methods	12a describes measures of model fit	12a	**"YES"** is filled in when the measure(s) of model fit are mentioned in the methods section. Model fit measurements calculate how close predictions to the outcome are. Procedures such Hosmer Lemeshow statistic, R2, Log likelihood or Akaike information criteria (AIC) are used for that purpose. **"NO"** is filled in when model fit analysis was performed but the procedure used for that purpose is not mentioned. **"NA"** is filled in when no model fit analysis was performed.
		12b Describes measures of predictive ability	12b	**"YES"** is filled in when the measure(s) of predictive ability are mentioned in the methods section. Predictive ability measurements calculate how well the model discriminates between those with and those without the outcome. Procedures such area under the ROC curve (AUC), discrimination slope or Brier score are used for that purpose. **"NO"** is filled in when predictive ability analysis was performed but the procedure used for that purpose is not mentioned. **"NA"** is filled in when no predictive ability analysis is performed.

Section	Sub-section	Elements and sub-elements		Description of elements and sub-elements
		12c Describes measures of reclassification	12c	**"YES"** is filled in when the measure(s) of reclassification are mentioned in the methods section. Studies that perform reclassification analysis are those which assessed whether the improvement of risk models also reclassifies people into different risk categories. These measures of reclassification are calculated from a reclassification table. Procedures as Net reclassification improvement (NRI), Integrated discrimination improvement (IDI) or categories are used for that purpose. **"NO"** is filled in when the study performs reclassification analysis but does not mention which procedure was used for that purpose. **"NA"** is filled in when no reclassification analysis was performed.
	Other	13a Describes all subgroups that were examined	13a	**"YES"** is filled in when all subgroups examined are mentioned. Subgroups refers to the analysis of a subset of the participants classified according to a characteristic, for example, according to sex, BMI (≥25 or ≥30) or status of a disease (mild or moderate or severe). **"NO"** is filled in when subgroup analysis was performed but it is not mentioned which subgroups were examined. **"NA"** is filled in when the study does not assess any specific subgroup.
		13b Describes why these subgroups were examined and whether they were based on a priori information or if analyses were exploratory, if done	13b	**"YES"** is filled in when the rationale of subgroup analysis is mentioned. Additionally, it should be mentioned whether the analysis was planned or exploratory. **"NO"** is filled in when the study does not mention either the rationale of the analysis or whether the analysis was planned or exploratory. **"NA"** is filled in when the study does not address any specific subgroup.
		13c If interaction effects were considered mentions which and why	13c	**"YES"** is filled in when the interaction effects examined are mentioned and are justified. **"NO"** is filled in when the study performs interaction analysis but does not mention which and why. **"NA"** is filled in when the study does not address any interaction effects.

Section	Sub-section	Elements and sub-elements	Description of elements and sub-elements
Results	Participants	14a Reports the number of individuals at each stage of the study	14a **"YES"** is filled in when the number of participants at the start of the study and the number of participants analyzed is reported. In addition, when it can be concluded that all participants were part of the analysis, the score **"YES"** is given.
		14b Reports the mains reasons for non-participation	14b **"YES"** is filled in when the main reasons for non participation are reported (e.g. refused to participate, budget issues, missing data).
		14c Reports numbers of participants not genotyped	14c **"YES"** is filled in when the number of participants not genotyped is reported.
		14d Reports the reasons for not genotyping	14d **"YES"** is filled in when the reasons for not genotyping are reported (e.g. unavailability of DNA material, budget issues, genotyping quality issues).
	Population	15a Reports clinical and demographic characteristics of study population	15a **"YES"** is filled in when clinical and demographic characteristics of the study population are reported (e.g. sex, age, ethnicity, information on other risk factors or co-morbidity).
		15b Reports risk factors of study population included in the model	15b **"YES"** is filled in when the risk factors for the population included in the model are reported. **"NA"** is filled in when the prediction model does not include non-genetic risk factors.
	Model estimates	16a Reports unadjusted associations between the variables and the outcome	16a **"YES"** is filled in when unadjusted effect estimates are reported (e.g. odds ratios, hazard ratios, beta coefficients).
		16b Reports adjusted estimates from the full risk model(s)	16b **"YES"** is filled in when adjusted estimates from the full risk model(s) are reported (e.g. odds ratios, hazard ratios, beta coefficients).
		16c Reports the intercept (α) of the model, if applicable	16c **"YES"** is filled in when the intercept (α) of the model is reported.
	Risk distributions	17 Reports distributions of predicted risks and/or risk scores	17 **"YES"** is filled in when the distributions of predicted risks and/or risk scores are reported.
	Assessment	18a Reports measures of model fit	18a **"YES"** is filled in when the values of measures of model fit are reported. For more information see item 12a. **"NO"** is filled in when the values of measures of model fit are not reported. **"NA"** is filled in when the study does not perform any model fit analysis.

Section	Sub-section	Elements and sub-elements		Description of elements and sub-elements
		18b Reports measures of predictive ability	18b	**"YES"** is filled in when the values of predictive ability measures are reported. For more information see item 12b.
				"NO" is filled in when the values of predictive ability measures are not reported.
				"NA" is filled in when the study does not perform any predictive ability analysis.
		18c Reports measures of reclassification	18c	**"YES"** is filled in when the values of reclassification measures are reported. For more information see item 12c.
				"NO" is filled in whenthe values of reclassification measures are not reported.
				"NA" is filled in when the study does not perform any reclassification analysis.
	Validation	19a Reports internal validation	19a	**"YES"** is filled in when the values of internal validation measures are reported. For more information see item 10a.
				"NO" is filled in whenthe values of internal validation measures are not reported.
				"NA" is filled in when the study does not perform any internal validation analysis.
		19b Reports external validation	19b	**"YES"** is filled in when the values of external validation measures are reported. For more information see item 10b.
				"NO" is filled in when the values of external validation measures are not reported.
				"NA" is filled in when the study does not perform any external validation analysis.
	Other analyses	20a Reports results of subgroups, interaction or exploratory analyses	20a	**"YES"** is filled in when results of subgroups or interaction or exploratory analyses are reported. For more information see item 13.
				"NO" is filled in when results of subgroups or interaction or exploratory analyses are not reported.
				"NA" is filled in when the study does not perform any subgroups or interaction or exploratory analysis.
		20b States the difference between exploratory and non-exploratory data	20b	**"YES"** is filled in when the difference between exploratory and non-exploratory data is reported. For more information see item 13.
				"NO" is filled in when the difference between exploratory and non-exploratory is not reported.
				"NA" is filled in when the study does not perform any exploratory analysis.

Section	Sub-section		Elements and sub-elements		Description of elements and sub-elements
Discussion	Limitations	21	Discusses limitations of the study and their impact on the results of the study	21a	Discusses limitations of the study: when limitations are described the score **"YES"** is given. Limitations should include sources of potential bias and confounding or address issues in the study design and analyses (e.g., characteristics of the study population, selection of participants, procedures and measures used in data collection, length of follow-up, sample size and missing data).
				21b	Discusses the impact of limitations on the results: when the impact of the limitations on the results is mentioned the score **"YES"** is given.
	Interpretation	22	Gives an overall interpretation of results	22	**"YES"** is filled in when there is a comparison between the study with other studies, taking into account the main results, design and study conduction.
	Generalizability	23a	Discusses the representativeness of study population within the target population	23a	**"YES"** is filled in when possible differences in demographic variables, such as in sex, age, risk factors or ethnicity of the study population affects the applicability in a (future) target population for testing.
					"NO" is filled in when there is no mention of possible differences between study population and a (future) target population.
		23b	Discusses clinical relevance	23b	**"YES"** is filled in when the potential impact of genetic results on medical or public health decision making is mentioned.
					"NO" is filled in when the potential clinical impact is not mentioned.
		23c	Discusses what information is needed to adopt the proposed model	23c	**"YES"** is filed in when authors explain what information is needed to adopt the proposed model or suggest that further research should be done to adopt it.
					"NO" is filled when authors mention that the model is ready for implementation and not further research is needed.
Other	Supplementary information	24	Describes how other authors can access the model protocols	24	**"YES"** is filled in when relevant information to replicate the model is provided. For example, the β coefficients should be provided.
	Funding	25a	Gives the source of funding and the role of the funders for the present study	25a	**"YES"** is filled in when the source of funding to carry out the study and the role of the funders is mentioned.
		25b	States whether there are any conflicts of interest	25b	**"YES"** is filled in when authors state whether there are any conflicts of interest.

PART II
Applications

Chapter 7

Evaluation of risk prediction updates from commercial genome-wide scans

Raluca Mihaescu, Mandy van Hoek, Eric J.G. Sijbrands,
André G. Uitterlinden, Jacqueline C.M. Witteman, Albert Hofman,
Cornelia M. van Duijn, A. Cecile J.W. Janssens

Genetics in Medicine. 2009;11(8):588-94

ABSTRACT

Purpose: Commercial internet-based companies offer genome-wide scans to predict the risk of common diseases and personalize nutrition and lifestyle recommendations. These risk estimates are updated with every new gene discovery.

Methods: In order to assess the benefits of updating risk information in commercial genome-wide scans, we compared type 2 diabetes risk predictions based on *TCF7L2* alone, 18 polymorphisms alone, and 18 polymorphisms plus age, sex and body mass index. Analyses were performed using data from the Rotterdam study, a prospective, population-based study among individuals aged 55 years and older. Data were available from 5297 participants.

Results: The actual prevalence of type 2 diabetes in the study population was 20%. Predicted risks were below average for carriers of the *TCF7L2* CC genotype (predicted risk 17.6%) and above average for the CT and TT genotypes (20.8% and 28.0%). Including data on the 18 polymorphisms caused 34% of participants to be reclassified (i.e. switched between below and above average): 24% of the CC carriers changed to increased risk, 52% and 6% of the CT and TT carriers changed to decreased risk. Including information on age, sex and body mass index caused 29% to change categories (27%, 31% and 19% for CC, CT and TT carriers respectively). In total, 39% of participants changed categories once when risk factors were updated, and 11% changed twice, i.e. back to their initial risk category.

Conclusion: Updating risk factors may produce contradictory information about an individual's risk status over time, which is undesirable if lifestyle and nutritional recommendations vary accordingly.

INTRODUCTION

The accelerating rate of genomic discoveries is rapidly increasing our understanding of the genetic basis of common diseases. Recent genome-wide association studies have identified novel susceptibility variants for type 2 diabetes, age-related macular degeneration, cancer and many other common diseases.[1] These discoveries have fueled expectations about applications of predictive genetic tests in preventive and clinical health care (2, 3). It is envisioned that genetic tests will personalize medicine through targeted treatment for patients with common diseases and individualized lifestyle and dietary recommendations for high-risk individuals (4, 5).

Although genome-based clinical and public health applications still await empirical evidence, several companies already offer online genetic tests to predict an individual's risk of common diseases.[6] These tests are based on single susceptibility genes (e.g. DNA direct (7)), genetic profiles based on a limited number of variants (e.g. Sciona (8), Genovations (9)), genome-wide scans (e.g. 23andMe (10), Navigenics (11) deCODEme (12)), and whole genome sequencing (e.g. Knome (13)). It is widely acknowledged that testing single susceptibility genes is uninformative for predicting common diseases as, on their own, they only minimally affect disease risk (14-16) and most currently offered profiles based on a few selected variants are uninformative as they lack a firm scientific basis for the polymorphisms included (6).

Companies that offer genome-wide scans take a more rigorous approach in the selection of the variants, but the clinical validity and utility of their results may also be limited at present, as susceptibility genes for common diseases are still being discovered. Because of this, risk predictions from genome-wide scans frequently become outdated when scientific knowledge progresses. Therefore, commercial companies offer updates of the risk predictions when new susceptibility genes are discovered. Given that single new variants only have a minor contribution to disease risk, we might expect that risk predictions change minimally at each update. However, as many individuals will have disease risks that are only slightly higher or lower than average (17), even minor updates may reclassify people from below to above average disease risk or vice versa, and lifestyle and nutrition recommendations may vary accordingly.

We investigated the extent to which updating of risk predictions leads to reclassification of individuals from below to above average disease risk or vice versa. Taking type 2 diabetes as an example, we compared risk predictions based on a single gene, on multiple polymorphisms and on multiple polymorphisms combined with age, sex and body mass index. Analyses were performed using data from the Rotterdam Study, a population-based cohort of individuals aged 55 years and older.

MATERIALS AND METHODS

Subjects

The design and data collection of the Rotterdam Study have been described elsewhere (18). In short, the Rotterdam Study is a prospective, population-based, cohort study among 7,983 inhabitants of a Rotterdam suburb, designed to investigate determinants of chronic diseases. Participants were aged 55 years and older. Baseline examinations took place from 1990 until 1993, and follow-up examinations were performed in 1993-1994, 1997-1999 and 2002-2004. Between these exams, continuous surveillance on major disease outcomes was conducted. The medical ethics committee of the Erasmus Medical Center approved the study protocol and all participants gave their written informed consent.

Data Collection

The following polymorphisms (19) were genotyped: *TCF7L2* rs7903146 (MIM 602228), *CDKAL1* rs7754840 (MIM 611259), *CDKN2A/B* rs10811661 (MIM 600160, MIM 600431), *FTO* rs8050136 (MIM 610966), *HHEX* rs1111875 (MIM 604420), *IGF2BP2* rs4402960 (MIM 608289), *KCNJ11* rs5219 (MIM 600937), *PPARG* rs1801282 (MIM 601487), *SLC30A8* rs13266634 (MIM 611145), *ADAMTS9* rs4411878 (MIM 605421), *CDC123-CAMK1D* rs11257622 (MIM 607957), *CDKN2A/B* rs1412829, *JAZF1* rs1635852 (MIM 606246), *NOTCH2* rs1493694 (MIM 600275), *TCF2* rs4430796 (MIM 189907), *THADA* rs7578597 (MIM 611800), *TSPAN8-LGR5* rs1353362 (MIM 600769, MIM 606667) and *WFS1* rs10012946 (MIM 606201) (20-24). Details on genotyping techniques, genotype success and odds ratios for the genotyped variants have been published elsewhere (25).

At baseline, diagnostic criteria for prevalent cases of diabetes were a non-fasting or post-load glucose level (after oral glucose tolerance testing) ≥11.1mmol/l and/or treatment with anti-diabetic medication (oral medication or insulin) with a diagnosis of diabetes recorded by a general practitioner. During follow-up, incident cases of diabetes were diagnosed at fasting plasma glucose levels ≥7.0mmol/l, and/or non-fasting plasma glucose levels ≥11.1mmol/l, and/or treatment with anti-diabetic medication (oral medication or insulin) (26, 27) with a diagnosis of diabetes recorded by a general practitioner. Patients with a recorded diagnosis of type 1 diabetes were excluded from the present analyses (n=15). Body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared. Age and BMI were obtained from the baseline assessment.

Statistical Analyses

Predicted risks were obtained using logistic regression analyses with type 2 diabetes (prevalent and incident cases) as the dependent variable. All polymorphisms were entered as categorical variables in the analyses, allowing effect sizes to differ between heterozygous and homozygous carriers of the risk alleles. To evaluate how risk predictions change after adding more information, we compared first risk predictions based on the strongest genetic predictor of type 2 diabetes, *TCF7L2,* alone, 18 polymorphisms including *TCF7L2*, and 18 polymorphisms plus age, sex and BMI. Second, we compared risk predictions based on clinical factors, clinical factors and *TCF7L2*, and clinical factors plus all 18 polymorphisms. Predicted risks from the three models were evaluated by comparing risk distributions and discriminative accuracy, and by examining reclassification. To evaluate how risk predictions change when each polymorphism is added individually, we simulated 1000 random permutations of all possible orderings of the added polymorphisms. Discriminative accuracy, measured as the area under the receiver operating characteristic curve (AUC), indicates the degree to which the predicted risks can discriminate between individuals who will develop the disease and those who will not. AUC can range from 0.50 (equal to tossing a coin) to 1.00 (perfect discrimination). Reclassification was calculated as the percentage of individuals who switched from being at increased to being at decreased risk as compared to the average risk in the population or vice versa (28). Reclassification was assessed in individuals with complete genotype and clinical information. Analyses were performed using the SPSS software version 15.0.1 and R programming language version 2.8.0.

RESULTS

General characteristics

A total of 6544 participants were successfully genotyped for at least one polymorphism. Complete genotype information on all polymorphisms was available from 5297 participants, of whom 490 were incident and 545 were prevalent cases of type 2 diabetes (i.e. 20% had type 2 diabetes). Of those with complete genotype information, 41% were men, mean age was 69.5 years (standard deviation 9.1 years) and mean body mass index was 26.3 kg/m^2 (standard deviation 3.7 kg/m^2). Complete information on genotype, age, sex and BMI was available from 5111 participants. The average risk of type 2 diabetes in the population was defined as the actual prevalence (i.e. 20%).

Improving Risk Prediction at Population Level

Prediction based on the 18 polymorphisms and clinical characteristics yielded more differentiation in predicted risks than prediction based on one or multiple polymorphisms alone (**Figure 1**), which means that adding more risk factors yielded more extreme risk predictions. For example,

the 5% of the population indicated to be at highest risk had a predicted risk of 28.0% based on *TCF7L2*, but predicted risks of at least 29.7% based on the 18 polymorphisms and at least 36.8% based on the polymorphisms plus clinical factors. This increased differentiation is also reflected in higher AUCs. The AUC was 0.55 (95% CI 0.53-0.57) for prediction based on *TCF7L2*, 0.60 (95% CI 0.58-0.62) for prediction based on 18 polymorphisms, and 0.66 (95% CI 0.64-0.68) for prediction based on 18 polymorphisms plus age, sex and BMI. At the average risk, the main improvement in model performance was reflected in the increase in specificity. The specificity was 51.8% for prediction based on *TCF7L2*, 62.7% for prediction based on 18 polymorphisms, and 64.5% for prediction based on 18 polymorphisms plus age, sex, BMI (**Table 1**).

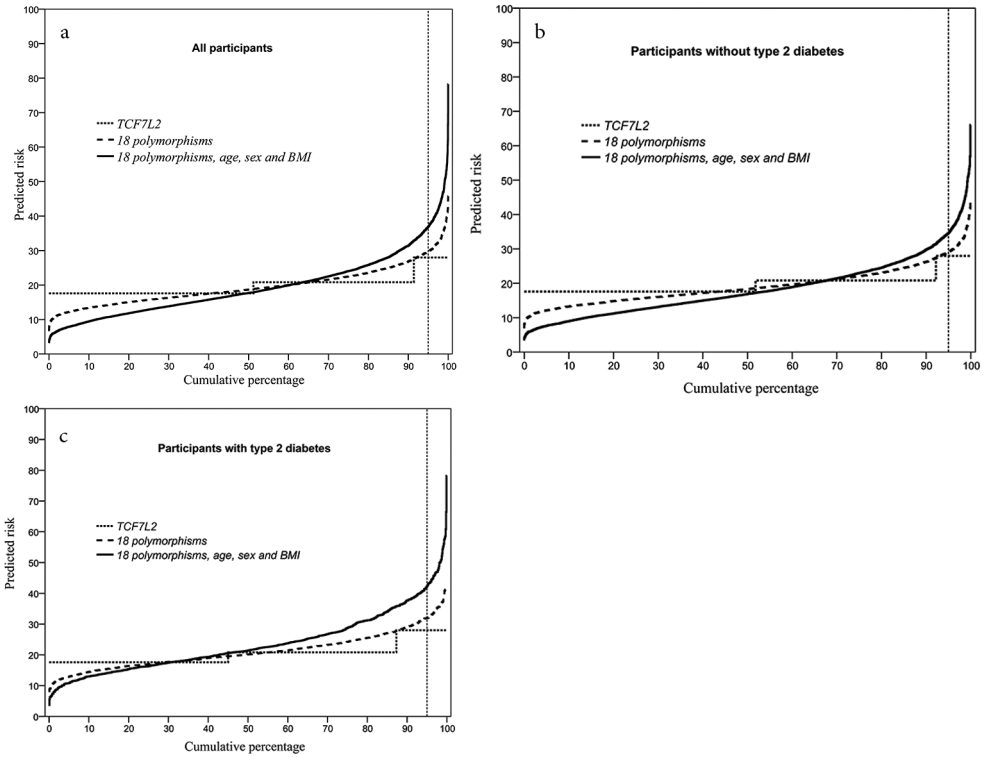
Reclassification

Predicted risks were lower than average for carriers of the *TCF7L2* CC genotype (predicted risk 17.6%) and higher than average for the CT and TT genotypes (20.8% and 28.0% respectively; **Figure 2**). As indicated by the larger standard deviations predicted risks diverged after adding novel risk factors, leading to reclassification. Based on testing the 18 polymorphisms, 33.5% of the participants were reclassified: 23.6% of non-carriers switched to the increased risk category and 43.6% of the heterozygous and homozygous carriers (51.6% and 5.6% respectively) switched to the decreased risk category (**Table 2**). Based on all polymorphisms, age, sex and BMI, 28.5% of participants switched their risk classification compared to prediction based on the 18 polymorphisms alone: the proportion of switchers was 26.5%, 31.4% and 19.1% for carriers of the CC, CT and TT genotypes respectively (data not shown). Overall, predicted risks changed from above to below average or vice versa in 50% of all individuals: 39% switched once and 11% switched twice i.e. back to their initial risk category (**Figure 3**).

Updating a Model Starting From Age, Sex and BMI

The AUC was 0.63 (95% CI 0.61-0.65) for prediction based on age, sex and BMI, 0.64 (95% CI 0.62-0.66) for prediction based on age, sex, BMI and *TCF7L2*, and 0.66 (95% CI 0.64-0.68) for prediction based on age, sex and BMI plus 18 polymorphisms. Starting from predictive testing based on age, sex and BMI, 13.2% of participants were reclassified after updating by *TCF7L2*. Based on age, sex, BMI and all polymorphisms, 16.3% of participants switched their risk classification compared to prediction based on age, sex, BMI and *TCF7L2*. Overall, predicted risks changed from above to below average or vice versa in 25.6% of all individuals: 21.7% switched once and 3.9% switched twice, i.e. back to their initial risk category.

Figure 1. Predictiveness curves (42) for *TCF7L2* alone, 18 polymorphisms alone, and 18 polymorphisms plus age, sex and body mass index.



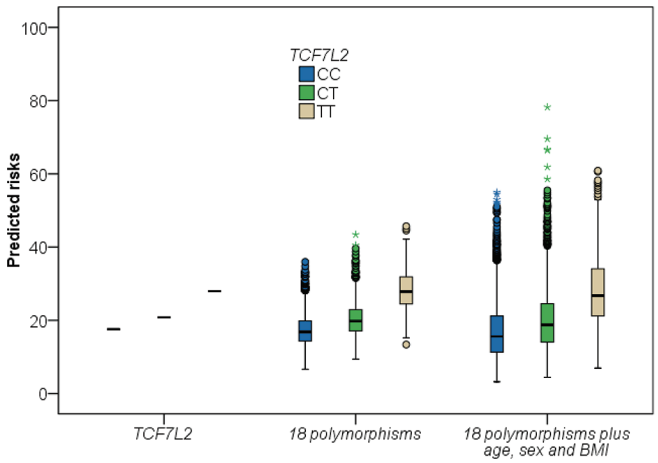
Predicted risks were obtained using logistic regression analysis. Cumulative percentage indicates the percentage of the population that has a predicted disease risk equal or lower than the risk value. E.g. based on genetic testing of 18 polymorphisms 90% (x-axis) of the individuals have a predicted risk lower than 26.8% (y-axis). Figure 1a shows the predictiveness curves for all participants, 1b for participants without type 2 diabetes, and 1c for participants with type 2 diabetes. BMI indicates body mass index (calculated as weight in kilograms divided by height in meters squared).

Table 1. Measures of model performance for dichotomizing predicted risk at average risk.

	Model based on *TCF7L2*	Model based on 18 polymorphisms	Model based on 18 polymorphisms, age, sex and body mass index
Sensitivity	55.7%	50.8%	57.2%
Specificity	51.8%	62.7%	64.5%
PPV	21.8%	24.7%	28.0%
NPV	82.9%	84.1%	86.2%

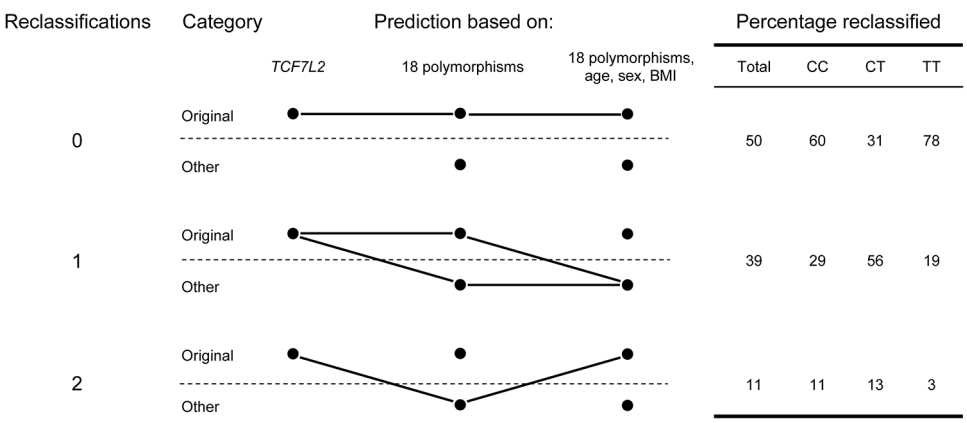
Measurements are based on individuals with complete genotype and clinical information. PPV indicates positive predictive value. NPV indicates negative predictive value.

Figure 2. Predicted risk of type 2 diabetes based on *TCF7L2* alone, 18 polymorphisms alone, and 18 polymorphisms plus age, sex and body mass index.



Predicted risks were obtained using logistic regression analysis. The bold line shows the median, the boxes indicate the interquartile ranges (25%-75% range), and the whiskers present 1.5 times the interquartile range. The points represent outliers and the asterisks represent extreme outliers that have values more than three times the interquartile range. BMI indicates body mass index (calculated as weight in kilograms divided by height in meters squared).

Figure 3. Patterns of reclassification that result from updating risk predictions.



Percentage reclassified	Total	CC	CT	TT
0	50	60	31	78
1	39	29	56	19
2	11	11	13	3

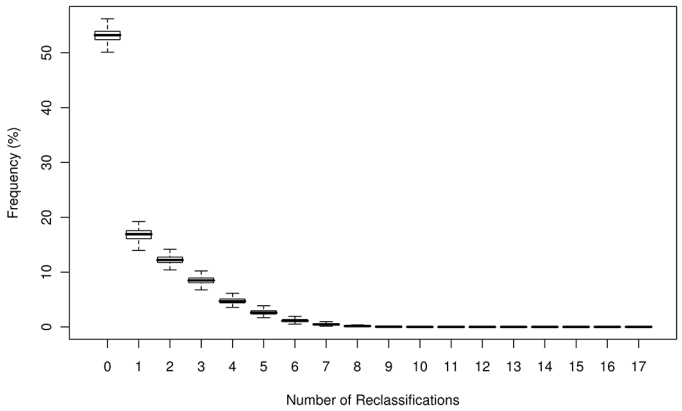
The number of reclassifications represents how many times a person switched between risk categories based on the three prediction models. E.g. a person did not reclassify (reclassification is 0) if they had above average or below average risks according to all three models. A person reclassified once (reclassification is 1) if they switched risk categories from the model based on *TCF7L2* to the model based on 18 polymorphisms, or from the model based on 18 polymorphisms to the model including clinical factors. The table explains what percentage of people reclassifies 0, 1 or 2 times, overall and by *TCF7L2* genotype. BMI indicates body mass index (calculated as weight in kilograms divided by height in meters squared).

Table 2. Risk stratification table for the pairwise comparison of two consecutively updated prediction models.

	Model based on 18 polymorphisms		Total		Model based on 18 polymorphisms, age, sex and body mass index		Total
Model based on *TCF7L2*	**Below average**	**Above average**		**Model based on 18 polymorphisms**	**Below average**	**Above average**	
Below average				Below average			
n (%)	**1966** (76.4%)	**607** (23.6%)	**2573** (100%)	n (%)	**2366** (77%)	**706** (23%)	**3072** (100%)
% of total	38.5%	11.9%	50.4%	% of total	46.3%	13.8%	60.1%
Cases, n	298	141	439	Cases, n	303	185	488
Observed risk	15.2%	23.2%	17.1%	Observed risk	12.8%	26.2%	15.9%
Above average				Above average			
n (%)	**1106** (43.6%)	**1432** (56.4%)	**2538** (100%)	n (%)	**717** (35.2%)	**1322** (64.8%)	**2039** (100%)
% of total	21.6%	28.0%	49.6%	% of total	14.7%	25.9%	39.9%
Cases, n	190	363	553	Cases, n	122	382	504
Observed risk	17.2%	25.3%	21.8%	Observed risk	17.0%	28.9%	24.7%
Total				Total			
n (%)	**3072** (60.1%)	**2039** (39.9%)	**5111** (100%)	n (%)	**3083** (60.3%)	**2028** (39.7%)	**5111** (100%)
% of total	60.1%	39.9%	100%	% of total	60.3%	39.7%	100%
Cases, n	488	504	992	Cases, n	425	567	992
Observed risk	15.9%	24.7%	19.4%	Observed risk	13.8%	28.0%	19.4%

Reclassification was calculated in participants for whom all data were available to avoid part of the reclassification being caused by differences in the average risk rather than differences in predicted risks.

Figure 4. Median percentage of reclassification in step by step update of prediction based on *TCF7L2* to prediction based on 18 polymorphisms.



The plot is obtained from a simulation of 1000 permutations of single polymorphism updates. Numbers on the x axis represent no reclassification (i.e. number of reclassification is 0) up to reclassification at each step (i.e. number of reclassification is 17). The bold line shows the median, the boxes indicate the interquartile ranges (25%-75% range), and the whiskers present 1.5 times the interquartile range.

Updating by Adding Each Polymorphism Individually

Finally, we considered risk updating by each additional polymorphism individually to the model that was based on testing *TCF7L2* alone, up to the model based on all 18 polymorphisms. Using 1000 random orderings in which the 17 polymorphisms can be added to the profile, we calculated that on average 47% (standard deviation 1.2%) of the participants ultimately switched at least once when risks were updated after every single polymorphism (**Figure 4**). Seventeen percent switched once, and 30% switched multiple times (range 2-15) from below to above average disease risk or vice versa. When *TCF7L2* was also added in a random order, on average 71% (standard deviation 8.3%) of the participants switched at least once (data not shown).

DISCUSSION

Using type 2 diabetes as an example, we showed that updating risk predictions by including more polymorphisms, age, sex, and BMI improved risk prediction at the population level as reflected in the higher AUC values. However, at the individual level, we found that 34% of the participants switched between risk categories when risks were updated from 1 to 18 polymorphisms and that 29% switched when age, sex and BMI were taken into consideration. In total, 39% of the participants switched risk categories once and 11% switched twice.

Before interpreting the public health relevance of these results, two methodological issues of our study, that may affect the degree of reclassification, should be pointed out. First, while we investigated 18 established type 2 diabetes polymorphisms, only about half were statistically significantly associated with type 2 diabetes risk in our population (25). This is in line with other studies that investigated the combined predictive value of the 18 polymorphisms, which also found that not all polymorphisms were statistically significant associated with the disease (29-32). If the effect sizes of all polymorphisms in our study had been the same as in the original studies that had identified their associations, we would have observed a larger variation in predicted risks (**Figure 2**) and likely also more reclassification.

Second, we focused on changes in risk prediction based on *TCF7L2* with that based on 18 polymorphisms, age, sex and BMI, which reflects the practice of commercial companies. In clinical settings, however, it is more logical to update a model based on recognized clinical risk factors. We showed that when risks were updated from age, sex and BMI to age, sex, BMI and *TCF7L2*, and further to age, sex, BMI and all 18 polymorphisms, 22% of individuals switched risk categories once, and 4% twice.

We compared risk prediction based on *TCF7L2* with that based on 18 polymorphisms, age, sex and BMI, but we considered only two risk updates: one based on adding 17 polymorphisms and one on adding age, sex and BMI. The percentage of reclassification was even higher when we considered risk updating by each additional polymorphism individually, as is done by the companies. The exact percentage of reclassification then varies with the order in which polymorphisms are added, and for 1000 random orderings of the 17 polymorphisms we calculated that 47% of the participants would have switched at least once when risks were updated after every single polymorphism, compared to 34% when the 17 polymorphisms were added in a single update.

The reason why people switch between risk categories is that the added polymorphisms may have different effects on disease risk compared to the polymorphisms already considered. **Figure 2** showed that individuals who are at increased risk according to their *TCF7L2* genotype may be at decreased risk of type 2 diabetes when all 18 polymorphisms are considered if they inherited protective genotypes on many other polymorphisms. In our analyses, the risk increase conferred by *TCF7L2* risk alleles was counterbalanced by protective effects of other alleles in 6% of the individuals, and was counterbalanced by young age, male sex and normal BMI in 19% of the individuals.

Our final prediction model included 18 polymorphisms, age, sex and BMI but it is important to realize that risk predictions can be further improved. Even with our current understanding of genomic factors, prediction of type 2 diabetes risk can be improved by also

considering family history and fasting plasma glucose levels (32), factors that are currently not considered by the companies that offer genome-wide scans. In the future, risk prediction may be improved by the addition of novel genetic factors, novel biomarkers, and with gene-gene and gene-environment interactions if these are demonstrated in future genetic epidemiological studies (33, 34). Thus, the risk predictions presented in this paper are not final and individuals may be subject to further reclassification as science advances.

Commercial companies assert that genome-wide scans will help consumers to learn their likelihood of developing a disease, but it is widely agreed that risk predictions and results from genetic tests are difficult for the lay public to understand (35, 36). To facilitate the interpretation of risk estimates, companies present the predicted risks together with the average risk of the disease for the total population or for a sex- and age-matched population. Individuals can thereby learn whether they are at higher or lower risk than others and, based on this information, may decide to make lifestyle and dietary changes. Individuals differ, however, in the way they value the information gained from genetic testing. Some may find a slight increase in predicted risk sufficiently motivating to adopt or maintain healthy behaviors, whereas others may not even change their behavior when they learn that their risk is markedly increased. A systematic review of the psychological and behavioral impact of genetic testing for hereditary nonpolyposis carcinoma, hereditary breast and ovarian cancer, and Alzheimer disease reported that, generally, 12 months after testing, perceived risk in carriers decreased to the pretest level or even below it (37). A study on the harms and benefits of *APOE* genotyping in first-degree relatives of patients with Alzheimer disease reported that disclosure of genotype status increased the motivation for risk reduction activities (38). Note that previous studies mainly addressed psychological and behavioral impact of genetic testing for monogenic and major gene disorders, and these findings cannot be directly translated to the impact of low-risk susceptibility genetic testing.

If individuals are informed that they have switched categories from above to below average risk of disease, or vice versa, their perceptions about the need for health behavior changes may vary accordingly. In current commercial genome scans, risks are updated on every new gene discovery and individuals may frequently reclassify over time. To date, it is unknown how individuals respond to variations in risk predictions over time and how it affects their perceptions about the threat of being at increased risk (39-41). Since health behavior changes are difficult to achieve, we might expect that individuals will become insensitive to risk information if they learn that their risk status may change over time, even without any lifestyle changes. Also, reclassification primarily focuses on changes in risk compared to the average risk, and less on the absolute risks of disease. The absolute risk should be important as well in decision making about healthy behavior and it is of interest to find out whether absolute or comparative

risk information influences health behavior change. Such potentially adverse consequences of updating risk predictions warrant further investigation.

The companies that offer genome-wide scans or whole genome sequencing for the prediction of multiple diseases take a higher scientific standard for the selection of susceptibility variants than those previously reviewed (6). They include only variants that have been consistently associated in multiple studies, and transparently present the polymorphisms that constitute genetic profiles for each disease, including references to scientific studies demonstrating their impact on disease risk. Nevertheless, with scientific advance their risk predictions may further improve, as causation of disease is better understood, and the benefit of these updates at the individual level are unclear. This does not imply that the introduction of genome-based applications in health care should wait until we completely understand the etiology of diseases, but we need to recognize that a premature introduction may have adverse effects.

REFERENCES

1. Manolio TA, Brooks LD, Collins FS. **A HapMap harvest of insights into the genetics of common disease.** *J Clin Invest.* 2008;**118**(5):1590-1605.
2. Brand A, Brand H, Schulte in den Baumen T. **The impact of genetics and genomics on public health.** *Eur J Hum Genet.* 2008;**16**(1):5-13.
3. Collins FS, McKusick VA. **Implications of the Human Genome Project for medical science.** *JAMA.* 2001;**285**(5):540-544.
4. Guttmacher AE, Collins FS. **Realizing the promise of genomics in biomedical research.** *JAMA.* 2005;**294**(11):1399-1402.
5. Collins FS, Green ED, Guttmacher AE, Guyer MS, Institute USNHGR. **A vision for the future of genomics research.** *Nature.* 2003;**422**(6934):835-847.
6. Janssens AC, Gwinn M, Bradley LA, Oostra BA, van Duijn CM, Khoury MJ. **A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions.** *Am J Hum Genet.* 2008;**82**(3):593-599.
7. DNA direct. Available at: http://www.dnadirect.com/ . Accessed November 20, 2008.
8. Sciona. Available at: http://www.sciona.com/ . Accessed November 20, 2008.
9. Genovations. Available at: http://www.genovations.com/ . Accessed November 20, 2008.
10. 23andMe. Available at: http://www.23andme.com/ . Accessed November 20, 2008.
11. Navigenics. Available at: http://www.navigenics.com . Accessed November 20, 2008.
12. deCODEme. Available at: http://www.decodeme.com/ . Accessed November 20, 2008.
13. Knome Web. Available at: http://www.knome.com/ . Accessed November 20, 2008.
14. Vineis P, Schulte P, McMichael AJ. **Misconceptions about the use of genetic tests in populations.** *Lancet.* 2001;**357**(9257):709-712.
15. Holtzman NA, Marteau TM. **Will genetics revolutionize medicine?** *N Engl J Med.* 2000;**343**(2):141-144.
16. Janssens AC, Gwinn M, Valdez R, Narayan KM, Khoury MJ. **Predictive genetic testing for type 2 diabetes.** *BMJ.* 2006;**333**(7567):509-510.
17. Janssens AC, van Duijn CM. **Genome-based prediction of common diseases: advances and prospects.** *Hum Mol Genet.* 2008;**17**(R2):R166-173.
18. Hofman A, Breteler MM, van Duijn CM, et al. **The Rotterdam Study: objectives and design update.** *Eur J Epidemiol.* 2007;**22**(11):819-829.
19. OMIM (Online Mendelian Inheritance in Man) Baltimore: Johns Hopkins University CfMG, 1996. http://www3.ncbi.nlm.nih.gov/omim/.
20. Diabetes Genetics Initiative of Broad Institute of Harvard and Mit LUaNIoBR, Saxena R, Voight BF, et al. **Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.** *Science.* 2007;**316**(5829):1331-1336.
21. Scott LJ, Mohlke KL, Bonnycastle LL, et al. **A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.** *Science.* 2007;316(5829):1341-1345.
22. Sladek R, Rocheleau G, Rung J, et al. **A genome-wide association study identifies novel risk loci for type 2 diabetes.** *Nature.* 2007;**445**(7130):881-885.
23. Zeggini E, Weedon MN, Lindgren CM, et al. **Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.** *Science.* 2007;**316**(5829):1336-1341.

24. Wellcome Trust Case Control C. **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature.* 2007;**447**(7145):661-678.

25. van Hoek M, Dehghan A, Witteman JC, et al. **Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study.** *Diabetes.* 2008;**57**(11):3122-3128.

26. **Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus.** *Diabetes Care.* 1997;**20**(7):1183-1197.

27. Alberti KG, Zimmet PZ. **Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation.** *Diabet Med.* 1998;**15**(7):539-553.

28. Cook NR. **Use and misuse of the receiver operating characteristic curve in risk prediction.** *Circulation.* 2007;**115**(7):928-935.

29. Vaxillaire M, Veslot J, Dina C, et al. **Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study.** *Diabetes.* 2008;57(1):244-254.

30. Lango H, Consortium UKTDG, Palmer CN, et al. **Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk.** *Diabetes.* 2008;**57**(11):3129-3135.

31. Lyssenko V, Jonsson A, Almgren P, et al. **Clinical risk factors, DNA variants, and the development of type 2 diabetes.** *N Engl J Med.* 2008;**359**(21):2220-2232.

32. Meigs JB, Shrader P, Sullivan LM, et al. **Genotype score in addition to common risk factors for prediction of type 2 diabetes.** *N Engl J Med.* 2008;**359**(21):2208-2219.

33. Janssens AC. **Is the time right for translation research in genomics?** *Eur J Epidemiol.* 2008;**23**(11):707-710.

34. Stolk RP, Rosmalen JG, Postma DS, et al. **Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study.** *Eur J Epidemiol.* 2008;**23**(1):67-74.

35. Erby LH, Roter D, Larson S, Cho J. **The rapid estimate of adult literacy in genetics (REAL-G): a means to assess literacy deficits in the context of genetics.** *Am J Med Genet A.* 2008;**146A**(2):174-181.

36. Scheuner MT, Sieverding P, Shekelle PG. **Delivery of genomic medicine for common chronic adult diseases: a systematic review.** *JAMA.* 2008;**299**(11):1320-1334.

37. Heshka JT, Palleschi C, Howley H, Wilson B, Wells PS. **A systematic review of perceived risks, psychological and behavioral impacts of genetic testing.** *Genet Med.* 2008;10(1):19-32.

38. Roberts JS, Cupples LA, Relkin NR, Whitehouse PJ, Green RC. **Genetic risk assessment for adult children of people with Alzheimer's disease: the Risk Evaluation and Education for Alzheimer's Disease (REVEAL) study.** *J Geriatr Psychiatry Neurol.* 2005;**18**(4):250-255.

39. Kuehn BM. **Risks and benefits of direct-to-consumer genetic testing remain unclear.** *JAMA.* 2008;**300**(13):1503-1505.

40. Kuehn BM. **Prenatal genome testing sparks debate.** *JAMA.* 2008;**300**(14):1637-1639.

41. Offit K. **Genomic profiles for disease risk: predictive or premature?** *JAMA.* 2008;**299**(11):1353-1355.

42. Pepe MS, Feng Z, Huang Y, et al. **Integrating the predictiveness of a marker with its performance as a classifier.** *Am J Epidemiol.* 2008;**167**(3):362-368.

Chapter 8

Comparison of performance of genetic risk prediction models for type 2 diabetes in nine European samples

Raluca Mihaescu, Pau Navarro, Inga Prokopenko, Jennifer Bragg-Gresham, Klaus Strassburger, Loïc Yengo, Mandy van Hoek, Najaf Amin, Beverley Balkau, Michael Boehnke, Harry Campbell, Guillaume Charpentier, Timothy M. Frayling, Martin Gögele, Harald Grallert, Caroline Hayward, Andrew A. Hicks, Albert Hofman, Thomas Illig, Cecilia M. Lindgren, Letizia Marullo, Ben A. Oostra, Irene Pichler, Ozren Polasek, Nigel W. Rayner, Igor Rudan, Eric J.G. Sijbrands, André G. Uitterlinden, Sarah H. Wild, Sara M. Willems, Anne U. Jackson, Lori L. Bonnycastle, Jaakko Tuomilehto, Stephane Cauchi, Wolfgang Rathmann, Laura Scott, Philippe Froguel, Cornelia M. van Duijn, Peter P. Pramstaller, Alan F. Wright, James F. Wilson, Mark I. McCarthy, A.Cecile J.W. Janssens

Submitted

ABSTRACT

Objective: The predictive ability of genetic variants in type 2 diabetes has been studied in several populations, but risk models have only rarely been validated in independent populations. We aimed to assess the generalizability of risk models that include genetic markers across nine European study populations.

Methods: Using logistic regression analyses, we constructed and validated three models for the prediction of type 2 diabetes risk, based on: 1) clinical variables age, sex and body mass index; 2) 37 well validated susceptibility variants derived from GWAS and available in all nine samples; 3) clinical and genetic factors jointly. Generalizability was evaluated by discrimination (area under receiver operating characteristic curve; AUC) across nine independent populations including one longitudinal cohort study, four case-control studies and four cross-sectional studies in small genetic isolates.

Results: The range of AUC values for the genetic models was 0.61-0.83 at model development and 0.49-0.70 at external validation. Genetic models developed in the longitudinal cohort and the case-control study with population-based cases and controls had moderate AUC at derivation that decreased slightly at validation (maximum decrease -0.13). Genetic models developed in case-control studies with non population-based cases and controls and genetic isolates generally had markedly higher AUC at derivation that decreased substantially (maximum decrease -0.33) at validation.

Conclusions: The observed predictive ability of genetic risk models showed considerable variability between studies due to differences in study design and population characteristics. Genetic risk prediction models should ideally be developed and validated in populations that resemble the target audience for testing.

Type 2 diabetes is a common disease caused by a complex interplay between multiple genetic, behavioral and environmental factors. Type 2 diabetes and its preclinical stage are associated with increased risk of cardiovascular disease. Given the availability of effective interventions to delay or prevent the onset of disease (1), a more active approach to the early detection of individuals at risk may be helpful (2). The identification of high risk individuals is currently based on established risk factors, including age, sex and body mass index (BMI) (3, 4).

Enabled by the recent progress in genome-wide association studies, genetic susceptibility variants are increasingly being investigated for their potential to improve the prediction of type 2 diabetes risk. These risk prediction models are based on genetic risk factors alone or on genetic variants added to known non-genetic risk factors (see appendix). In all studies, the predictive ability of the genetic models, as measured by the area under the receiver operating characteristic curve (AUC), was lower than that of the non-genetic models (5-22). Addition of genetic factors to the traditional risk factors resulted in a marginal increase of the predictive ability, which generally is considered of limited clinical value (23).

An essential step in the assessment of predictive ability is validation of the risk model. Validation reflects the ability to predict risks for individuals from the same sample in which the model was developed (internal validation) or from independent samples (external validation) (24). A risk model is more generalizable if calibration and discrimination are similar across samples (25). Calibration refers to how close the risks predicted by the model are to the observed risks, and discrimination refers to the distribution of predicted risks among affected and non-affected individuals. Our review of studies that used genetic variants to predict risk of type 2 diabetes revealed that none has externally validated their risk models (see appendix) (5-22, 26).

We assessed the predictive ability of three risk prediction models for type 2 diabetes in nine European samples. We compared risk predictions based on (1) age, sex and BMI; (2) on 37 common single nucleotide polymorphisms (SNPs); and (3) on age, sex, BMI and 37 SNPs. We used each sample to develop the risk models as well as to externally validate the models of all other samples. The generalizability of the risk models was assessed by comparing calibration and discrimination measures between the nine European samples.

RESEARCH DESIGN AND METHODS

Study samples and definitions of type 2 diabetes

Nine study datasets, which are part of the Diabetes Genetics Replication and Meta-analysis (DIAGRAM (27)) consortium, were used to develop and validate the risk prediction models for type 2 diabetes (see appendix). We had data available from one longitudinal cohort study (Rotterdam study), four case-control studies (DGDG, FUSION, KORA and WTCCC) and four

cross-sectional studies in genetic isolated populations (CROATIA-Vis, ERF study, MICROS and ORCADES).

The Rotterdam Study is a prospective, population-based, cohort study among inhabitants of a Rotterdam suburb, designed to investigate determinants of chronic diseases (28). Participants were aged 55 years and older. Continuous surveillance on major disease outcomes was conducted between the periodical follow-up examinations. Data from 5712 subjects were used for the current analyses.

The Diabetes Gene Discovery Group (DGDG) is a case-control study in subjects of French Caucasian origin (29). Diabetic subjects were recruited at the UMR8090 CNRS unit in Lille or at the Endocrinology-Diabetology Department of the Corbeil-Essonnes Hospital. Controls were drawn from participants in the Epidemiological Study on the Insulin Resistance syndrome (DESIR) program. Data from 679 cases and 697 controls were used for the current analyses.

The Finland–United States Investigation of NIDDM genetics (FUSION) study aims to discover genes associated with type 2 diabetes and intermediate quantitative traits in Finnish subjects (30). Unrelated cases were drawn from sibships with two or more siblings reporting type 2 diabetes (familial cases) and from the FINRISK 2002 study, a population-based risk factor survey (31, 32). Normoglucose tolerant (NGT) controls were drawn from spouses of familial cases, from individuals who were NGT at ages 65 and 70 years from Vantaa, Finland, and from Finrisk2002 (31, 32). Controls were approximately frequency matched to cases based on 5-year age category, sex and birth province. Data from 1,092 cases and 1,171 controls were used for the current analyses.

The Cooperative Health Research in the Region of Augsburg, Southern Germany, project (KORAgen) is a regional research platform for population-based studies, subsequent follow-up studies and family studies (33, 34). For the current analyses we used data from the KORA surveys S3, F3 (follow-up of S3) and S4. Data from 429 type 2 diabetes cases and 1432 non-diabetic age- and sex-matched controls, all of German ethnicity, were used for the current analyses.

The Wellcome Trust Case Control Consortium dataset (WTCCC) includes type 2 diabetes cases and population controls from UK (35, 36). The type 2 diabetes cases were selected from UK Caucasian subjects who are part of the Diabetes UK Warren 2 repository. Controls were derived from two sources in approximately equal numbers: the 1958 Birth Cohort and the UK National Blood Services (NBS) (35). Full information on age, sex and BMI was available only for controls from the NBS. Thus, data from 1903 cases and 1437 controls were used for the current analyses.

The Croatian study (CROATIA-Vis), the Erasmus Rucphen Family study (ERF), the Study of Micro-isolates in South Tyrol (MICROS) and the Orkney Complex Disease Study (ORCADES)

are part of the European Special Populations Research Network (EUROSPAN) project involving genetically isolated European populations. CROATIA-Vis includes individuals from the villages of Vis and Komiza on the Dalmatian island of Vis (37). 110 cases and 660 controls were used for the current analyses. The ERF study is a family-based study that includes inhabitants of a genetically isolated community in the southwestern area of the Netherlands (38). Data from 51 cases and 1178 controls were used for the current analyses. MICROS is a population-based survey that includes samples from three isolated villages located in Val Venosta area, Italy (39). Data from 43 cases and 1018 controls were used for the current analyses. ORCADES is an ongoing family-based, cross-sectional study in the isolated Scottish archipelago of Orkney (40). Data from 27 cases and 668 controls were used for the current analyses.

All participants gave written informed consent and the study protocols were approved by the relevant institutional or national ethics committees. All participants were of European ancestry.

Outcome measures (type 2 diabetes)

Diagnostic criteria varied in details but generally reflected the 1997 American Diabetes Association (ADA) criteria and/or the 1999 WHO criteria (41-43). Most studies used clinical criteria to diagnose type 2 diabetes and excluded individuals with type 1 diabetes and other forms of diabetes (see appendix).

Selection of SNPs and genotyping

The present analyses included 37 variants that were known genetic risk factors by May 2010 and were available in all nine samples. We selected type 2 diabetes loci as follows: 18 loci already established as genome wide significant based on previous reports: *NOTCH2*, *THADA*, *PPARG*, *ADAMTS9*, *IGF2BP2*, *WFS1*, *CDKAL1*, *JAZF1*, *SLC30A8*, *CDKN2A/B*, *CDC123/CAMK1D*, *HHEX/IDE*, *TCF7L2*, *KCNQ1* (rs163184), *KCNJ11*, *TSPAN8/LGR5*, *FTO*, *HNF1B (TCF2)* (29, 31, 35, 36, 44-46), two loci that were not initially replicated DIAGRAM(27) but have been established as genome-wide significant in recent reports: *MTNR1B*, *IRS1* (47-50) and 11 loci reported for the first time in the DIAGRAM meta-analysis (27): *BCL11A*, *ZBED3*, *KLF14*, *TP53INP1*, *CHCHD9*, *KCNQ1* (rs231362; linkage disequilibrium: r^2=0.01 between rs231362 and rs163184 in HapMap release 22 CEU), *CENTD2*, *HMGA2*, *HNF1A*, *ZFAND6*, *PRC1* (27). We added an independent SNP in the *KCNQ1* locus (rs2237892; linkage disequilibrium: r^2=0.064 between rs2237892 and rs163184; and r^2=0.008 between rs2237892 and rs231362 in HapMap release 22 CEU) that reached genome wide significance in East Asians (51). Also, we included five loci associated with type 2 diabetes reported in a recent meta-analysis done within the Meta-

Analyses of Glucose and Insulin-related traits Consortium (MAGIC): *GCK*, *DGKB-TMEM195*, *GCKR*, *ADCY5*, *PROX1* (52). Rs numbers are reported in the appendix.

Different platforms were used to genotype the study samples. Several studies used the Illumina platform Human Hap 300/300-duo+ Bead Array (DGDG, FUSION, and EUROSPAN) or the Human Hap 500 Bead Array (Rotterdam study). Other studies used Affymetrix GeneChip Human Mapping 500k Array Set (KORA and WTCCC). Genotyped SNPs were in Hardy-Weinberg equilibrium ($p>10^{-6}$ to 10^{-3}), the minimum call rate varied per study between 90 and 99%, and the minor allele frequency was greater than 1%. SNPs not found on the genotyping platform were imputed from HapMap CEU sample data. Imputation was performed using IMPUTE (DGDG, KORA and WTCCC) or MACH (Rotterdam study, FUSION and EUROSPAN). Details on genotyping and imputation techniques of individual studies can be found in the appendix.

Statistical analyses

Prediction models were constructed in each sample and applied to predict type 2 diabetes risks in all samples. In each sample, three risk models were constructed: (1) a model based on age, sex and BMI; (2) a model based on the 37 type 2 diabetes-associated SNPs; and (3) a model based on age, sex, BMI plus the 37 SNPs. Several studies included additional variables as covariates in the models to account for special features of their study designs: FUSION included birth province to account for the matching of cases and controls, and WTCCC included geographic origin as an extra variable. These variables were not considered in the prediction of type 2 diabetes risk in the other samples.

Predicted risks were obtained using logistic regression analyses with type 2 diabetes (prevalent and incident cases) as dependent variable. Participants included in the current analyses had complete phenotypic data and complete information on the 37 SNPs. All polymorphisms were entered as continuous variables, consistent with an allele dosage effect. In MICROS one SNP (i.e. rs13081389) was removed from the prediction models because the risk allele was rare in individuals without type 2 diabetes but present in all individuals with type 2 diabetes, resulting in a very high beta coefficient for this SNP. Whether this reflects a true difference in allele frequencies due to genetic drift or is a chance finding due to the small sample size is difficult to determine.

To apply the risk models in all other samples, the regression coefficients of all predictors were obtained from each dataset and for each model, resulting in 9 samples x 3 models. We constructed weighted risk scores by multiplying each predictor with its corresponding estimated regression (beta) coefficient. To obtain predicted risks for all individuals in the study we fitted

logistic regression models with weighted risk scores as predictor variable and type 2 diabetes status as outcome. In other words, we recalibrated the models at validation by obtaining new intercepts and slopes for the weighted risk scores. Birth province and geographic region were included as covariates in the regression models constructed and validated in FUSION and WTCCC.

The performance of the risk prediction models was assessed by comparing calibration and discriminative accuracy. Calibration was assessed using the Hosmer-Lemeshow (H-L) chi-square test, which compares the difference between the observed and expected number of people with events in categories (here deciles) of predicted risk. We also plotted observed events by deciles of predicted risk as a visual representation of the H-L test. If calibration is perfect, all points lay on the diagonal line. Discriminative accuracy was assessed by the AUC, which indicates the degree to which predicted risks can discriminate between individuals who will develop the disease and those who will not. AUC generally ranges from 0.50 (equal to tossing a coin) to 1.00 (perfect discrimination). A test with an AUC below 0.50 predicts worse than tossing a coin.

RESULTS

Study characteristics

The percentage of prevalent and incident cases of type 2 diabetes varied by study population and design. In the Rotterdam prospective cohort study prevalent and incident cases represented 19.6% of the study sample; in the population-based case-control study KORA the proportion was 23.1%, in other case-control studies the proportion was 49.3% in DGDG, 48.3% in FUSION and 57% in WTCCC; and in the cross-sectional studies in genetic isolates it was 14.3% in CROATIA-Vis, 4.1% in ERF and MICROS, and 3.9% in ORCADES. Descriptive characteristics and risk allele frequencies of each study sample are provided in the appendix.

Effects of risk factors on type 2 diabetes risk in the clinical, genetic and combined risk models

As described previously, type 2 diabetes risk increased with age, male sex and higher BMI, but the observed ORs varied substantially between samples (i.e. clinical models). The ORs for age (per year), sex (men vs. women) and BMI (per kg/m^2) ranged from 1.00 to 1.14, from 1.20 to 2.54, and from 1.04 to 1.52, respectively (see appendix for regression coefficients). Higher ORs, by definition, were seen in case-control studies and cross-sectional studies in genetic isolates with a larger difference in the distribution of age, sex and BMI between cases and controls. The

observed ORs of the genetic variants also varied substantially between studies (**Figure 1** and appendix).

Discrimination

Table 1 shows the AUC values of the risk prediction models both at model development (on the diagonal) and at external validation in all other datasets. As expected, within each sample, the model developed in its own sample had the highest AUC value (comparing values per row in **Table 1**). At model development, the cohort (Rotterdam study) had the lowest AUC values for the clinical, genetic and combined models, namely 0.64, 0.61 and 0.67. The maximum AUC values for these models were 0.90 (in WTCCC), 0.83 (in ORCADES) and 0.91 (in WTCCC and ORCADES), respectively. The genetic models generally had lower AUC values than the clinical models. Adding SNPs to the clinical models always increased the AUC, with increments varying between 0.01 and 0.09.

To test if the large AUCs of the genetic models obtained in the genetic isolates were due to the small sample size of the study populations we performed additional simulation analyses. As such, we down sampled the Rotterdam study to the number of cases and controls from the small genetic isolates. The median AUC and 95% confidence intervals from 200 simulations were: 0.69 (0.66-0.74) for replicating the CROATIA-Vis sample size; 0.75 (0.69-0.81) for ERF; 0.77 (0.71-0.82) for MICROS; and 0.83 (0.75-0.89) for ORCADES.

At validation, the AUC of the clinical, genetic and combined models generally decreased compared to derivation. Models developed in the cross-sectional studies in small genetic isolates showed the largest decrease in AUC. For example, the maximum decrease in AUC was 0.33, when the genetic model developed in ORCADES was validated in KORA, CROATIA-Vis or MICROS (see **Table 1b**). On the other hand, models developed in the cohort study (Rotterdam) and in the case-control study with population-based cases and controls (KORA) and in a case-control study with non population-based cases and controls (FUSION) showed a smaller decrease in AUC. The largest drop in AUC at validation of genetic models developed in these samples was 0.13 for the model developed in KORA and tested in ORCADES. Surprisingly, the AUC sometimes increased at validation. This happened when the external sample had a model with a higher AUC at derivation than the model that was validated. This was mainly observed in clinical and combined models and to a smaller extent in genetic models (see **Table 1a-c**).

Whereas at derivation the combined model always had higher AUC than the clinical model, at validation we found that combined models also had lower AUCs than models with clinical factors alone (see **Table 1a and 1c**). This was mainly due to the small sample size of the genetically isolated populations.

Table 1. Area under the ROC curve for the clinical, genetic and combined type 2 diabetes risk prediction models.

a. Clinical model

		Developed in								
Validated in		**Rotterdam**	**DGDG**	**FUSION**	**KORA**	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
	Rotterdam	**0.64**	0.63	0.60	0.63	0.63	0.61	0.64	0.63	0.62
	DGDG	0.82	**0.83**	0.79	0.83	0.80	0.77	0.81	0.80	0.80
	FUSION	0.68	0.69	**0.71**	0.69	0.63	0.57	0.66	0.65	0.69
	KORA	0.68	0.69	0.66	**0.69**	0.68	0.64	0.68	0.66	0.65
	WTCCC	0.88	0.86	0.74	0.87	**0.90**	0.88	0.89	0.88	0.83
	CROATIA-Vis	0.68	0.66	0.58	0.67	0.70	**0.70**	0.69	0.69	0.65
	ERF	0.80	0.79	0.72	0.79	0.80	0.79	**0.80**	0.80	0.78
	MICROS	0.84	0.82	0.73	0.82	0.84	0.84	0.84	**0.85**	0.82
	ORCADES	0.81	0.81	0.81	0.81	0.78	0.75	0.80	0.80	**0.82**

b. Genetic model

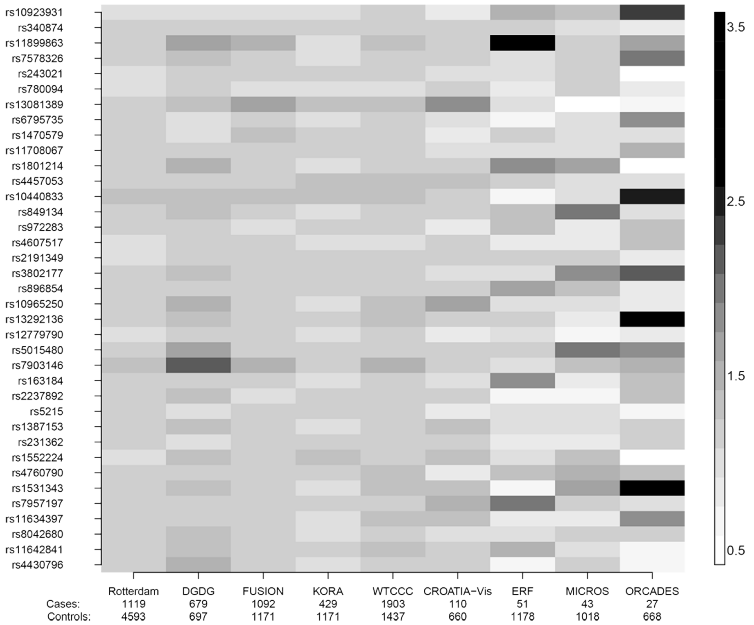
		Developed in								
Validated in		**Rotterdam**	**DGDG**	**FUSION**	**KORA**	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
	Rotterdam	**0.61**	0.58	0.59	0.57	0.59	0.54	0.51	0.53	0.53
	DGDG	0.70	**0.74**	0.70	0.62	0.69	0.62	0.54	0.61	0.54
	FUSION	0.62	0.61	**0.64**	0.60	0.62	0.55	0.52	0.55	0.52
	KORA	0.58	0.56	0.58	**0.63**	0.58	0.55	0.51	0.52	0.50
	WTCCC	0.68	0.66	0.67	0.65	**0.69**	0.61	0.61	0.61	0.61
	CROATIA-Vis	0.56	0.59	0.56	0.56	0.57	**0.68**	0.49	0.55	0.50
	ERF	0.55	0.57	0.54	0.56	0.56	0.51	**0.75**	0.60	0.56
	MICROS	0.60	0.62	0.60	0.58	0.59	0.58	0.55	**0.76**	0.50
	ORCADES	0.58	0.56	0.55	0.50	0.55	0.54	0.58	0.50	**0.83**

c. Combined clinical and genetic model

		Developed in								
Validated in		**Rotterdam**	**DGDG**	**FUSION**	**KORA**	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
	Rotterdam	**0.67**	0.66	0.64	0.65	0.66	0.59	0.58	0.61	0.57
	DGDG	0.85	**0.88**	0.83	0.83	0.83	0.74	0.70	0 .76	0.66
	FUSION	0.72	0.72	**0.74**	0.72	0.66	0.60	0.63	0.66	0.65
	KORA	0.70	0.71	0.68	**0.73**	0.70	0.62	0.63	0.63	0.59
	WTCCC	0.89	0.87	0.79	0.87	**0.91**	0.81	0.82	0.82	0.73
	CROATIA-Vis	0.69	0.68	0.61	0.68	0.71	**0.77**	0.64	0.68	0.59
	ERF	0.79	0.80	0.73	0.78	0.81	0.72	**0.89**	0.78	0.65
	MICROS	0.85	0.84	0.77	0.83	0.85	0.82	0.80	**0.90**	0.72
	ORCADES	0.80	0.81	0.79	0.80	0.79	0.68	0.69	0.75	**0.91**

Clinical model: age, sex and body mass index; Genetic model: 37 genetic variants; Combined clinical and genetic model: age, sex, body mass index and 37 genetic variants. Confidence intervals are included in Supplementary Table 6. Values in bold on the diagonal indicate the AUC at model development.

Figure 1. Odds ratios for the 37 variants included in the type 2 diabetes genetic risk prediction models.



The color legend depicts the odds ratio size from multivariable regression analyses with the 37 SNPs as predictor variables and type 2 diabetes as outcome variable. SNPs are ordered by chromosome number.

Calibration

All models generally showed good calibration at model development (see appendix). As expected, a few large H-L test values were observed at derivation and validation in the larger studies. Moreover, despite the small H-L test values inspection of the calibration plots indicates that models developed in the small studies were miscalibrated at validation in most external samples (see appendix).

DISCUSSION

In this study, we constructed genetic risk models for type 2 diabetes in nine samples, and evaluated the performance of these models in each sample, to investigate the impact of study design and study characteristics on the predictive ability of the risk models. The main finding of this study was that the magnitude of AUC varied enormously and was strongly dependent on study design, study characteristics and sample size. The highest AUC values were achieved in case-control studies and the cross-sectional studies in small genetic isolates. The AUC of genetic

models generally decreased at validation, with the largest decrease observed for models derived in the small genetic isolates. However this finding should be interpreted with caution since the large AUC at derivation was mainly due to their small sample size. The AUC of genetic models developed in the case-control study with population-based selection of cases and controls and in the longitudinal cohort study was lower at derivation but decreased to a smaller extent at validation.

It is important to acknowledge here that none of these case-control studies or the small genetic isolates were previously used to test the predictive ability of clinical risk factors. We used these case-control studies to provide insight into the biases introduced by different study designs. This is important because the current literature abounds in case-control studies that are used to assess the predictive ability of clinical and genetic predictive models (5, 10, 11, 20, 21, 53, 54). For purpose of risk variant detection, case-control designs, and particularly case-control designs that sample the extremes, i.e. the very healthy and the most severely ill, are more powerful than cohort studies of similar sample size. Studies with hyperselected cases and controls are not suitable for risk prediction because they overestimate the predictive power.

Before discussing the implications of our findings, we mention several methodological issues that could have affected our results. First, by downsampling the Rotterdam study we showed that the sample size was a strong determinant of the magnitude of AUC values. Thus, the very large AUC values in some of the genetic isolates may be explained by their small sample size. Second, the differences in AUC at model development may be explained by spectrum bias which refers to the change in performance of a test with the change in case-mix of cases or controls and with the prevalence of disease (55). Thus higher AUC values are expected in cases-control studies compared to cohort studies.

We observed a large variability in the discriminatory performance with the study design. The lowest AUC at model development was seen in the prospective cohort study. Models developed in case-control studies with non-population-based cases or controls showed good discriminatory ability, but their AUC dropped considerably when validated in the longitudinal cohort study or in the case-control study with random selection of cases and controls. The largest decrease in AUC at validation was observed for genetic models developed in the cross-sectional studies in small genetic isolated populations. Several possible explanations exist for the observed differences in predictive performance between derivation and validation across studies with different design.

First, poor discriminatory performance at external validation is partly explained by the fact that beta coefficients of risk factors differ between the samples. **Figure 1** shows that some SNPs had higher beta coefficients estimates in the case-control studies compared to the cohort study,

and most SNPs had more extreme estimates in the genetic isolates compared to the other studies. Higher beta coefficients estimates may reflect stronger genetic effects, e.g., explained by higher number of relatives with the disease in individuals with type 2 diabetes or, in genetic isolates, by true differences in allele frequencies due to founder effects and increased genetic homogeneity. Yet differences between beta coefficients estimates probably simply reflect sampling variation or low power to detect associations at common variant loci with modest effects in type 2 diabetes risk in cohorts with small sample size. Nevertheless, since prediction models are based on point estimates of the beta coefficients, different beta coefficients lead to different models with different predictive ability.

Second, AUC changes between derivation and validation when the selection criteria for the derivation sample are significantly different from the validation sample. For example, the discrepancy in AUC between the models derived in the Rotterdam study and the same models validated in case-control studies may be explained by the more homogeneous distribution of the clinical risk factors in the cohort study. Also, some case-control studies had stricter selection criteria for participant inclusion, e.g., while in most studies BMI varied across a wide range DGDG included cases and controls that all had a BMI below 30kg/m^2 and 27kg/m^2, respectively. For FUSION, the sampling design has effectively eliminated differences for age. The large variability in selection criteria for individuals included in each study has contributed to the change in performance at validation.

We have also investigated the calibration of the clinical, genetic and combined models both at derivation and validation. Several models showed miscalibration at external validation as indicated by the large H-L test values or by the visual inspection of the calibration plots. An explanation for the miscalibration observed at external validation may be that there were differences between the development and validation setting in the distribution of other measured clinical risk factors which were not included in the model. For example, the percentage of individuals with hypertension and prevalence of smoking differed between the samples considered in this study (see appendix). Thus in some instances calibration largely decreased at validation due to the differences in sample characteristics.

In conclusion, our analyses show that study design, sample size and characteristics of the samples included in the study lead to substantial variation in the discriminatory performance of risk prediction models. These results raise the question of what study design should be used to provide an unbiased assessment of the performance of genetic risk prediction models. Our analyses indicated that case-control studies with highly selected cases and controls and cross-sectional studies in small genetic isolates had the largest AUC values at derivation of genetic models, but they also displayed the greatest drop in AUC at validation. On the other hand, the

models developed in the longitudinal cohort study and the case-control study with random selection of cases and controls had a lower AUC at derivation but the AUC decreased by a smaller amount or even increased at validation. These results strongly suggest that genetic risk prediction models should be developed in samples with similar characteristics to the real clinical and public health settings for which genetic testing is envisioned. Furthermore, our results suggest that when the true effects of the genetic variants are similar across European populations it may be more reasonable to use the beta coefficients estimated from meta-analyses than from the individual datasets as weights in the genetic risk scores tested in European samples. However, while meta-analysis would help bypass the random variation it may mask important differences between populations. Further research should focus on testing the validity of weighted risk scores in different populations.

REFERENCES

1. Gillies CL, Abrams KR, Lambert PC, Cooper NJ, Sutton AJ, Hsu RT, Khunti K. **Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis.** *BMJ* 2007;**334**(7588)**:**299.
2. Tuomilehto J. **Nonpharmacologic therapy and exercise in the prevention of type 2 diabetes.** *Diabetes Care* 2009;**32 Suppl 2:**S189-93.
3. Rahman M, Simmons RK, Harding AH, Wareham NJ, Griffin SJ. **A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study.** *Fam Pract* 2008;**25**(3)**:**191-6.
4. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB, Sr. **Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study.** *Arch Intern Med* 2007;**167**(10)**:**1068-74.
5. Lango H, Consortium UKTDG, Palmer CN, Morris AD, Zeggini E, Hattersley AT, McCarthy MI, Frayling TM, Weedon MN. **Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk.** *Diabetes* 2008;**57**(11)**:**3129-35.
6. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB, Sr., Cupples LA. **Genotype score in addition to common risk factors for prediction of type 2 diabetes.** *N Engl J Med* 2008;**359**(21)**:**2208-19.
7. van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, Hofman A, Sijbrands EJ, Janssens AC. **Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study.** *Diabetes* 2008;**57**(11)**:**3122-8.
8. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, Berglund G, Altshuler D, Nilsson P, Groop L. **Clinical risk factors, DNA variants, and the development of type 2 diabetes.** *N Engl J Med* 2008;**359**(21)**:** 2220-32.
9. Vaxillaire M, Veslot J, Dina C, Proenca C, Cauchi S, Charpentier G, Tichet J, Fumeron F, Marre M, Meyre D, Balkau B, Froguel P. **Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study.** *Diabetes* 2008;**57**(1)**:**244-54.
10. 1Sparso T, Grarup N, Andreasen C, Albrechtsen A, Holmkvist J, Andersen G, Jorgensen T, Borch-Johnsen K, Sandbaek A, Lauritzen T, Madsbad S, Hansen T, Pedersen O. **Combined analysis of 19 common validated type 2 diabetes susceptibility gene variants shows moderate discriminative value and no evidence of gene-gene interaction.** *Diabetologia* 2009;**52**(7)**:**1308-14.
11. 1Miyake K, Yang W, Hara K, Yasuda K, Horikawa Y, Osawa H, Furuta H, Ng MC, Hirota Y, Mori H, Ido K, Yamagata K, Hinokio Y, Oka Y, Iwasaki N, Iwamoto Y, Yamada Y, Seino Y, Maegawa H, Kashiwagi A, Wang HY, Tanahashi T, Nakamura N, Takeda J, Maeda E, Yamamoto K, Tokunaga K, Ma RC, So WY, Chan JC, Kamatani N, Makino H, Nanjo K, Kadowaki T, Kasuga M. **Construction of a prediction model for type 2 diabetes mellitus in the Japanese population based on 11 genes with strong evidence of the association.** *J Hum Genet* 2009;**54**(4)**:**236-41.
12. 1Schulze MB, Weikert C, Pischon T, Bergmann MM, Al-Hasani H, Schleicher E, Fritsche A, Haring HU, Boeing H, Joost HG. **Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study.** *Diabetes Care* 2009;**32**(11)**:**2116-9.
13. 1Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, Hunter DJ, Hu FB. **Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry.** *Ann Intern Med* 2009; **150**(8)**:**541-50.

14. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, Kivimaki M, Humphries SE. **Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study.** *BMJ* 2010; **340:** b4838.

15. Wang J, Stancakova A, Kuusisto J, Laakso M. **Identification of undiagnosed type 2 diabetic individuals by the Finnish diabetes risk score and biochemical and genetic markers: a population-based study of 7232 Finnish men.** *J Clin Endocrinol Metab* 2010;**95**(8)**:**3858-62.

16. Fontaine-Bisson B, Renstrom F, Rolandsson O, Payne F, Hallmans G, Barroso I, Franks PW. **Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population.** *Diabetologia* 2010;**53**(10)**:**2155-62.

17. Ruchat SM, Vohl MC, Weisnagel SJ, Rankinen T, Bouchard C, Perusse L. **Combining genetic markers and clinical risk factors improves the risk assessment of impaired glucose metabolism.** *Ann Med* 2010;**42**(3)**:** 196-206.

18. de Miguel-Yanes JM, Shrader P, Pencina MJ, Fox CS, Manning AK, Grant RW, Dupuis J, Florez JC, D'Agostino RB, Sr., Cupples LA, Meigs JB, Investigators M, Investigators D. **Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms.** *Diabetes Care* 2011;**34**(1)**:**121-5.

19. Rotger M, Gsponer T, Martinez R, Taffe P, Elzi L, Vernazza P, Cavassini M, Bernasconi E, Hirschel B, Furrer H, Weber R, Ledergerber B, Egger M, Telenti A, Tarr PE, Swiss HIVCS. **Impact of single nucleotide polymorphisms and of clinical risk factors on new-onset diabetes mellitus in HIV-infected individuals.** *Clin Infect Dis* 2010;**51**(9)**:**1090-8.

20. Xu M, Bi Y, Xu Y, Yu B, Huang Y, Gu L, Wu Y, Zhu X, Li M, Wang T, Song A, Hou J, Li X, Ning G. **Combined effects of 19 common variations on type 2 diabetes in Chinese: results from two community-based studies.** *PloS one* 2010;**5**(11)**:**e14022.

21. Rees SD, Hydrie MZ, Shera AS, Kumar S, O'Hare JP, Barnett AH, Basit A, Kelly MA. **Replication of 13 genome-wide association (GWA)-validated risk variants for type 2 diabetes in Pakistani populations.** *Diabetologia* 2011;**54**(6)**:**1368-74.

22. Janipalli CS, Kumar MV, Vinay DG, Sandeep MN, Bhaskar S, Kulkarni SR, Aruna M, Joglekar CV, Priyadharshini S, Maheshwari N, Yajnik CS, Chandak GR. **Analysis of 32 common susceptibility genetic variants and their combined effect in predicting risk of Type 2 diabetes and related traits in Indians.** *Diabet Med* 2012;**29**(1)**:**121-7.

23. Meigs JB. **Prediction of type 2 diabetes: the dawn of polygenetic testing for complex disease.** *Diabetologia* 2009;**52**(4)**:**568-70.

24. Altman DG, Vergouwe Y, Royston P, Moons KG. **Prognosis and prognostic research: validating a prognostic model.** *BMJ* 2009;**338:**b605.

25. Harrell FE, Jr., Lee KL, Mark DB. **Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.** *Stat Med* 1996;**15**(4)**:**361-87.

26. Tabara Y, Osawa H, Kawamoto R, Onuma H, Shimizu I, Makino H, Kohara K, Miki T. **Genotype risk score of common susceptible variants for prediction of type 2 diabetes mellitus in Japanese: the Shimanami Health Promoting Program (J-SHIPP study). Development of type 2 diabetes mellitus and genotype risk score.** *Metabolism* 2011;**60**(11)**:**1634-40.

27. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Bostrom K, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haeften TW, van Herpt T, van Vliet-Ostaptchouk JV, Bragi Walters G, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllensten U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohlke KL, Morris AD, Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI. **Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis.** *Nat Genet* 2010;**42**(7)**:** 579-89.

28. Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC. **The Rotterdam Study: objectives and design update.** *Eur J Epidemiol* 2007; **22**(11)**:**819-29.

29. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. **A genome-wide association study identifies novel risk loci for type 2 diabetes.** *Nature* 2007;**445**(7130)**:**881-5.

30. Valle T, Tuomilehto J, Bergman RN, Ghosh S, Hauser ER, Eriksson J, Nylund SJ, Kohtamaki K, Toivanen L, Vidgren G, Tuomilehto-Wolf E, Ehnholm C, Blaschak J, Langefeld CD, Watanabe RM, Magnuson V, Ally DS, Hagopian WA, Ross E, Buchanan TA, Collins F, Boehnke M. Mapping genes for NIDDM. **Design of the Finland-United States Investigation of NIDDM Genetics (FUSION) Study.** *Diabetes Care* 1998;**21**(6)**:**949-58.

31. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. **A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.** *Science* 2007;**316**(5829)**:**1341-5.

32. Saaristo T, Peltonen M, Lindstrom J, Saarikoski L, Sundvall J, Eriksson JG, Tuomilehto J. **Cross-sectional evaluation of the Finnish Diabetes Risk Score: a tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome.** *Diab Vasc Dis Res* 2005;**2**(2)**:**67-72.

33. Wichmann HE, Gieger C, Illig T. K**ORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes.** *Gesundheitswesen* 2005;**67 Suppl 1:**S26-30.

34. Huth C, Heid IM, Vollmert C, Gieger C, Grallert H, Wolford JK, Langer B, Thorand B, Klopp N, Hamid YH, Pedersen O, Hansen T, Lyssenko V, Groop L, Meisinger C, Doring A, Lowel H, Lieb W, Hengstenberg C, Rathmann W, Martin S, Stephens JW, Ireland H, Mather H, Miller GJ, Stringham HM, Boehnke M, Tuomilehto J, Boeing H, Mohlig M, Spranger J, Pfeiffer A, Wernstedt I, Niklason A, Lopez-Bermejo A, Fernandez-Real JM, Hanson RL, Gallart L, Vendrell J, Tsiavou A, Hatziagelaki E, Humphries SE, Wichmann HE, Herder C, Illig T. **IL6 gene promoter polymorphisms and type 2 diabetes: joint analysis of individual participants' data from 21 studies.** *Diabetes* 2006;**55**(10)**:**2915-21.

35. Wellcome Trust Case Control Consortium. **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007;**447**(7145)**:**661-78.

36. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT. **Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes**. *Science* 2007;**316**(5829)**:** 1336-41.

37. Rudan I, Campbell H, Rudan P. **Genetic epidemiological studies of eastern Adriatic Island isolates, Croatia: objective and strategies.** *Coll Antropol* 1999;**23**(2)**:**531-46.

38. Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, van Duijn CM. **Linkage disequilibrium in young genetically isolated Dutch population.** *Eur J Hum Genet* 2004;**12**(7)**:**527-34.

39. Pattaro C, Marroni F, Riegler A, Mascalzoni D, Pichler I, Volpato CB, Dal Cero U, De Grandi A, Egger C, Eisendle A, Fuchsberger C, Gogele M, Pedrotti S, Pinggera GK, Stefanov SA, Vogl FD, Wiedermann CJ, Meitinger T, Pramstaller PP. **The genetic study of three population microisolates in South Tyrol (MICROS): study design and epidemiological perspectives.** *BMC medical genetics* 2007;**8:**29.

40. Helgason A, Nicholson G, Stefansson K, Donnelly P. **A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift.** *Ann Hum Gene*t 2003;**67**(Pt 4)**:** 281-97.

41. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Car*e 1997; **20**(7)**:**1183-97.

42. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 2003; **26 Suppl 1:**S5-20.

43. World Health Organization. **Definitions, diagnosis and classification of diabetes mellitus, report of a WHO consultation, part 1: diagnosis and classification of diabetes mellitus.** Geneva:WHO;1999.

44. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. **Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.** *Science* 2007;**316**(5829)**:**1331-6.

45. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Wellcome Trust Case Control C, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. **Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes.** *Nat Genet* 2008;**40**(5)**:**638-45.

46. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, Ng DP, Holmkvist J, Borch-Johnsen K, Jorgensen T, Sandbaek A, Lauritzen T, Hansen T, Nurbaya S, Tsunoda T, Kubo M, Babazono T, Hirose H, Hayashi M, Iwamoto Y, Kashiwagi A, Kaku K, Kawamori R, Tai ES, Pedersen O, Kamatani N, Kadowaki T, Kikkawa R, Nakamura Y, Maeda S. **SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations.** *Nat Genet* 2008;**40**(9)**:**1098-102.

47. Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, Thorleifsson G, Loos RJ, Manning AK, Jackson AU, Aulchenko Y, Potter SC, Erdos MR, Sanna S, Hottenga JJ, Wheeler E, Kaakinen M, Lyssenko V, Chen WM, Ahmadi K, Beckmann JS, Bergman RN, Bochud M, Bonnycastle LL, Buchanan TA, Cao A, Cervino A, Coin L, Collins FS, Crisponi L, de Geus EJ, Dehghan A, Deloukas P, Doney AS, Elliott P, Freimer N, Gateva V, Herder C, Hofman A, Hughes TE, Hunt S, Illig T, Inouye M, Isomaa B, Johnson T, Kong A, Krestyaninova M, Kuusisto J, Laakso M, Lim N, Lindblad U, Lindgren CM, McCann OT, Mohlke KL, Morris AD, Naitza S, Orru M, Palmer CN, Pouta A, Randall J, Rathmann W, Saramies J, Scheet P, Scott LJ, Scuteri A, Sharp S, Sijbrands E, Smit JH, Song K, Steinthorsdottir V, Stringham HM, Tuomi T, Tuomilehto J, Uitterlinden AG, Voight BF, Waterworth D, Wichmann HE, Willemsen G, Witteman JC, Yuan X, Zhao JH, Zeggini E, Schlessinger D, Sandhu M, Boomsma DI, Uda M, Spector TD, Penninx BW, Altshuler D, Vollenweider P, Jarvelin MR, Lakatta E, Waeber G, Fox CS, Peltonen L, Groop LC, Mooser V, Cupples LA, Thorsteinsdottir U, Boehnke M, Barroso I, Van Duijn C, Dupuis J, Watanabe RM, Stefansson K, McCarthy MI, Wareham NJ, Meigs JB, Abecasis GR. **Variants in MTNR1B influence fasting glucose levels.** *Nat Genet* 2009; **41**(1)**:** 77-81.

48. Lyssenko V, Nagorny CL, Erdos MR, Wierup N, Jonsson A, Spegel P, Bugliani M, Saxena R, Fex M, Pulizzi N, Isomaa B, Tuomi T, Nilsson P, Kuusisto J, Tuomilehto J, Boehnke M, Altshuler D, Sundler F, Eriksson JG, Jackson AU, Laakso M, Marchetti P, Watanabe RM, Mulder H, Groop L. **Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion.** *Nat Genet 2009*;**41**(1)**:** 82-8.

49. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C, Sparso T, Holmkvist J, Marchand M, Delplanque J, Lobbens S, Rocheleau G, Durand E, De Graeve F, Chevre JC, Borch-Johnsen K, Hartikainen AL, Ruokonen A, Tichet J, Marre M, Weill J, Heude B, Tauber M, Lemaire K, Schuit F, Elliott P, Jorgensen T, Charpentier G, Hadjadj S, Cauchi S, Vaxillaire M, Sladek R, Visvikis-Siest S, Balkau B, Levy-Marchal C, Pattou F, Meyre D, Blakemore AI, Jarvelin MR, Walley AJ, Hansen T, Dina C, Pedersen O, Froguel P. **A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk.** *Nat Genet* 2009;**41**(1)**:**89-94.

50. Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proenca C, Bacot F, Balkau B, Belisle A, Borch-Johnsen K, Charpentier G, Dina C, Durand E, Elliott P, Hadjadj S, Jarvelin MR, Laitinen J, Lauritzen T, Marre M, Mazur A, Meyre D, Montpetit A, Pisinger C, Posner B, Poulsen P, Pouta A, Prentki M, Ribel-Madsen R, Ruokonen A, Sandbaek A, Serre D, Tichet J, Vaxillaire M, Wojtaszewski JF, Vaag A, Hansen T, Polychronakos C, Pedersen O, Froguel P, Sladek R. **Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia.** *Nat Genet* 2009;**41**(10)**:**1110-5.

51. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, Hirota Y, Mori H, Jonsson A, Sato Y, Yamagata K, Hinokio Y, Wang HY, Tanahashi T, Nakamura N, Oka Y, Iwasaki N, Iwamoto Y, Yamada Y, Seino Y, Maegawa H, Kashiwagi A, Takeda J, Maeda E, Shin HD, Cho YM, Park KS, Lee HK, Ng MC, Ma RC, So WY, Chan JC, Lyssenko V, Tuomi T, Nilsson P, Groop L, Kamatani N, Sekine A, Nakamura Y, Yamamoto K, Yoshida T, Tokunaga K, Itakura M, Makino H, Nanjo K, Kadowaki T, Kasuga M. **Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus.** Nat Genet 2008;**40**(9)**:**1092-7.

52. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Magi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JR, Egan JM, Lajunen T, Grarup N, Sparso T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proenca C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Payne F, Roccasecca RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Bottcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day IN, de Geus EJ, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, Gyllensten U, Hadjadj S, Hallmans G, Hammond N, Han X, Hartikainen AL, Hassanali N, Hayward C, Heath SC, Hercberg S, Herder C, Hicks AA, Hillman DR, Hingorani AD, Hofman A, Hui J, Hung J, Isomaa B, Johnson PR, Jorgensen T, Jula A, Kaakinen M, Kaprio J, Kesaniemi YA, Kivimaki M, Knight B, Koskinen S, Kovacs P, Kyvik KO, Lathrop GM, Lawlor DA, Le Bacquer O, Lecoeur C, Li Y, Lyssenko V, Mahley R, Mangino M, Manning AK, Martinez-Larrad MT, McAteer JB, McCulloch LJ, McPherson R, Meisinger C, Melzer D, Meyre D, Mitchell BD, Morken MA, Mukherjee S, Naitza S, Narisu N, Neville MJ, Oostra BA, Orru M, Pakyz R, Palmer CN, Paolisso G, Pattaro C, Pearson D, Peden JF, Pedersen NL, Perola M, Pfeiffer AF, Pichler I, Polasek O, Posthuma D, Potter SC, Pouta A, Province MA, Psaty BM, Rathmann W, Rayner NW, Rice K, Ripatti S, Rivadeneira F, Roden M, Rolandsson O, Sandbaek A, Sandhu M, Sanna S, Sayer AA, Scheet P, Scott LJ, Seedorf U, Sharp SJ, Shields B, Sigurethsson G, Sijbrands EJ, Silveira A, Simpson L, Singleton A, Smith NL, Sovio U, Swift A, Syddall H, Syvanen AC, Tanaka T, Thorand B, Tichet J, Tonjes A, Tuomi T, Uitterlinden AG, van Dijk KW, van Hoek M, Varma D, Visvikis-Siest S, Vitart V, Vogelzangs N, Waeber G, Wagner PJ, Walley A, Walters GB, Ward KL, Watkins H, Weedon MN, Wild SH, Willemsen G, Witteman JC, Yarnell JW, Zeggini E, Zelenika D, Zethelius B, Zhai G, Zhao JH, Zillikens MC, Borecki IB, Loos RJ, Meneton P, Magnusson PK, Nathan DM, Williams GH, Hattersley AT, Silander K, Salomaa V, Smith GD, Bornstein SR, Schwarz P, Spranger J, Karpe F, Shuldiner AR, Cooper C, Dedoussis GV, Serrano-Rios M, Morris AD, Lind L, Palmer LJ, Hu FB, Franks PW, Ebrahim S, Marmot M, Kao WH, Pankow JS, Sampson MJ, Kuusisto J, Laakso M, Hansen T, Pedersen O, Pramstaller PP, Wichmann HE, Illig T, Rudan I, Wright AF, Stumvoll M, Campbell H, Wilson JF, Bergman RN, Buchanan TA, Collins FS, Mohlke KL, Tuomilehto J, Valle TT, Altshuler D, Rotter JI, Siscovick DS, Penninx BW, Boomsma DI, Deloukas P, Spector TD, Frayling TM, Ferrucci L, Kong A, Thorsteinsdottir U, Stefansson K, van Duijn CM, Aulchenko YS, Cao A, Scuteri A, Schlessinger D, Uda M, Ruokonen A, Jarvelin MR, Waterworth DM, Vollenweider P, Peltonen L, Mooser V, Abecasis GR, Wareham NJ, Sladek R, Froguel P, Watanabe RM, Meigs JB, Groop L, Boehnke M, McCarthy MI, Florez JC, Barroso I. **New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.** *Nat Genet* 2010;**42**(2)**:**105-16.

53. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, Rayner NW, Shields B, Owen KR, Hattersley AT, Frayling TM. **Combining information from common type 2 diabetes risk polymorphisms improves disease prediction.** *PLoS medicine* 2006;**3**(10)**:**e374.

54. Hu C, Zhang R, Wang C, Wang J, Ma X, Lu J, Qin W, Hou X, Wang C, Bao Y, Xiang K, Jia W. **PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population.** *PloS one* 2009;**4**(10)**:**e7643.

55. Ransohoff DF, Feinstein AR. **Problems of spectrum and bias in evaluating the efficacy of diagnostic tests.** *N Engl J Med* 1978;**299**(17)**:**926-30.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Assessment of calibration and discrimination in genetic risk prediction studies of type 2 diabetes.

Study reference	Non-genetic factors	# genetic variants	Calibration	Discrimination (AUC)			Validation
				Non-genetic only	Genetic only	Both	
Weedon et al. (1)	NA	3	No	NA	0.58	NA	No
Lyssenko et al. (2, 3)	BMI, FPG	2	No	0.68	NA	0.68	No
Lango et al. (4)	Age, sex, BMI	18	No	0.78	0.60	0.80	No
Meigs et al. (5)	Age, sex, BMI, family history, FPG, SBP, HDL, TG	18	H-L test Calibration plots	0.900	0.581 (adjusted for sex)	0.901	No
Cauchi et al. (6)	Age, sex, BMI	15	No	NA	NA	0.86	No
Balkau et al. (7)	Current smoker (only in men), waist circumference, hypertension, fasting glucose, γ-glutamyltransferase (GGT), and family history and BMI (only in women)	2	H-L test	Men: 0.850 Women: 0.917	NA	Men: 0.851 Women: 0.912	Internal validation: bootstrap sampling External validation of clinical score developed in this study
van Hoek et al. (8)	Age, sex, BMI	18	No	0.66	0.60	0.68	No
Lyssenko et al. (9)	Age, sex, BMI, family history, BP, TG, FPG	11	H-L test	0.743	0.626	0.753	Study is external validation of clinical risk scores. No validation of risk models developed in this study
	Age, sex, BMI, family history, BP, TG, FPG, HDL and waist circumference	11	H-L test	0.786	0.682	0.801	

Study reference	Non-genetic factors	# genetic variants	Calibration	Discrimination (AUC)			Validation
				Non-genetic only	Genetic only	Both	
Vaxillaire et al. (10)	Age, sex, BMI	3	No	0.82	0.56	0.83	No
Hu et al. (11)	Age, sex, BMI	11	No	0.614	0.621	0.668	No
Lin et al. (12)	Age, sex, family history, physical activity, triacyglycerol/HDL ratio, waist-hip ratio	15	H-L test	0.86	Weighted risk score: 0.59 Unweighted: 0.57	0.87	Internal validation: tenfold cross-validation
Sparso et al. (13)	Age, sex, BMI	19	No	0.92	0.60	0.93	Internal validation: bootstrapping
Miyake et al. (14)	Age, sex, BMI	11	No	0.68	0.63	0.72	No
Schulze et al. (15)	German Diabetes Risk Score (DRS): age, waist circumference, height, history of hypertension, physical activity, smoking, consumption of red meat, whole-grain bread, coffee and alcohol; glucose, HbA1c, TG, HDL cholesterol, γ-glutamyltransferase, alanine aminotransferase and hs-CRP	20	H-L test	0.90	NA	0.90	Study is external validation of clinical risk score. No validation of risk models developed in this study
Cornelis et al. (16)	Age, sex, BMI, family history, smoking, alcohol intake and physical activity	10	No	0.78	NA	0.79	No
Talmud et al. (17)	Cambridge T2D risk score: Age, sex, BMI, drug treatment, family history, smoking status	20	H-L test	0.72	0.55	0.73	Study is external validation of clinical risk scores. No validation of risk models developed in this study
	Framingham offspring T2D risk score: Age, sex, BMI, parental history of T2D, HDL, TG, fasting glucose)			0.78		0.78	

Study reference	Non-genetic factors	# genetic variants	Calibration	Discrimination (AUC)			Validation
				Non-genetic only	Genetic only	Both	
Wang et al. (18)	Finnish Diabetes Risk Score (FINDRISC): age, BMI, waist circumference, antihypertensive medication, physical activity, previously known high glucose, and daily consumption of vegetables, and fruits or berries	19	H-L test	0.727	0.552	0.730	Study is external validation of clinical risk score. No validation of risk models developed in this study
	FINDRISC plus TG, HDL, adiponectin, ALT			0.772		0.772	
Fontaine-Bisson et al. (19)	Age, sex	17 (T2D), 13 (glucose/ insulin), 26 (lipid), 17 (obesity), 73 (all)	No	NA	NA	0.591, 0.543, 0.565, 0.557, 0.626	No
Ruchat et al. (20)	Age, sex, BMI, SBP, DBP, smoking status	6	No	0.81	NA	0.85	No
de Miguel-Yanes et al. (21)	1) Stratified by age: sex, FH, BMI, FPG, SBP, HDL, TG	40	H-L test	1) <50 y: 0.908 ≥50 y: 0.883	1) <50 y: 0.657 ≥50 y: 0.590 (Adjusted for sex)	1) <50 y: 0.911 ≥50 y: 0.884	Internal validation: bootstrapping
	2) Age, sex, FH, BMI, FPG, SBP, HDL, TG			2) 0.903	2) 0.606	2) 0.906	
Qi et al. (22)	Age, sex, region, BMI, FH, smoking, alcohol use, physical activity, HDL, triacylglycerol	17	No	0.77	0.62	0.79	No

Study reference	Non-genetic factors	# genetic variants	Calibration	Discrimination (AUC)			Validation
				Non-genetic only	Genetic only	Both	
Rotger et al. (23)	Age, sex, BMI, antiretroviral therapy group, CD4+ T cell count, HDL, TG	4 22	No	0.75	NA	0.78 0.77	No
Xu et al. (24)	1) Case-control sample: age, sex, BMI, FH 2) Cohort sample: age, sex, BMI, FH	4	No	0.714 0.634	NA	0.730 0.663	No
Rees et al. (25)	Age, sex, BMI, country of residence	28	No	0.71	NA	0.74	No
Hivert et al. (26)	Age, sex, ethnic background, treatment arm, waist circumference	34	Calibration tests for the Cox model	0.628	NA	0.631	No
Tabara et al. (27)	Age, sex, BMI, glucose	10	No	NA	NA	NA	No
Janipalli et al. (28)	Age, sex, BMI, waist-hip ratio	32	No	0.959	0.634	0.963	No

Abbreviations of risk factors: ALT, alanine aminotransferase; BMI, body mass index; BP, blood pressure; FH, family history of T2D; FPG, fasting plasma glucose; GGT, γ-glutamyltransferase; HDL, high-density lipoprotein cholesterol; hs-CRP, high sensitivity c-reactive protein; HT, hypertension; SBP, systolic blood pressure; TG, triglycerides.
Other abbreviations: H-L test, Hosmer-Lemeshow chi square calibration test; NA, not available.

SUPPLEMENTAL REFERENCES

1. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, Rayner NW, Shields B, Owen KR, Hattersley AT, Frayling TM. **Combining information from common type 2 diabetes risk polymorphisms improves disease prediction.** *PLoS medicine* 2006;**3**(10):e374.
2. Lyssenko V, Almgren P, Anevski D, Orho-Melander M, Sjogren M, Saloranta C, Tuomi T, Groop L. **Genetic prediction of future type 2 diabetes.** *PLoS medicine* 2005;**2**(12): e345.
3. Janssens AC, Gwinn M, Khoury MJ, Subramonia-Iyer S. **Does genetic testing really improve the prediction of future type 2 diabetes?** *PLoS medicine* 2006;**3**(2):e114; author reply e27.
4. Lango H, Consortium UKTDG, Palmer CN, Morris AD, Zeggini E, Hattersley AT, McCarthy MI, Frayling TM, Weedon MN. **Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk.** Diabetes 2008;**57**(11):3129-35.
5. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB, Sr., Cupples LA. **Genotype score in addition to common risk factors for prediction of type 2 diabetes.** *N Engl J Med* 2008;**359**(21):2208-19.
6. Cauchi S, Meyre D, Durand E, Proenca C, Marre M, Hadjadj S, Choquet H, De Graeve F, Gaget S, Allegaert F, Delplanque J, Permutt MA, Wasson J, Blech I, Charpentier G, Balkau B, Vergnaud AC, Czernichow S, Patsch W, Chikri M, Glaser B, Sladek R, Froguel P. **Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value.** *PloS one* 2008;**3**(5):e2031.
7. Balkau B, Lange C, Fezeu L, Tichet J, de Lauzon-Guillain B, Czernichow S, Fumeron F, Froguel P, Vaxillaire M, Cauchi S, Ducimetiere P, Eschwege E. **Predicting diabetes: clinical, biological, and genetic approaches: data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR).** *Diabetes Care* 2008;**31**(10):2056-61.
8. van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, Hofman A, Sijbrands EJ, Janssens AC. **Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study.** *Diabetes* 2008;**57**(11):3122-8.
9. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, Berglund G, Altshuler D, Nilsson P, Groop L. **Clinical risk factors, DNA variants, and the development of type 2 diabetes.** *N Engl J Med* 2008;**359**(21):2220-32.
10. Vaxillaire M, Veslot J, Dina C, Proenca C, Cauchi S, Charpentier G, Tichet J, Fumeron F, Marre M, Meyre D, Balkau B, Froguel P. **Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study.** *Diabetes* 2008;**57**(1):244-54.
11. Hu C, Zhang R, Wang C, Wang J, Ma X, Lu J, Qin W, Hou X, Wang C, Bao Y, Xiang K, Jia W. **PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population.** *PloS one* 2009;**4**(10): e7643.
12. Lin X, Song K, Lim N, Yuan X, Johnson T, Abderrahmani A, Vollenweider P, Stirnadel H, Sundseth SS, Lai E, Burns DK, Middleton LT, Roses AD, Matthews PM, Waeber G, Cardon L, Waterworth DM, Mooser V. **Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score–the CoLaus Study.** *Diabetologia* 2009;**52**(4):600-8.
13. Sparso T, Grarup N, Andreasen C, Albrechtsen A, Holmkvist J, Andersen G, Jorgensen T, Borch-Johnsen K, Sandbaek A, Lauritzen T, Madsbad S, Hansen T, Pedersen O. **Combined analysis of 19 common validated type 2 diabetes susceptibility gene variants shows moderate discriminative value and no evidence of gene-gene interaction.** *Diabetologia* 2009;**52**(7):1308-14.

14. Miyake K, Yang W, Hara K, Yasuda K, Horikawa Y, Osawa H, Furuta H, Ng MC, Hirota Y, Mori H, Ido K, Yamagata K, Hinokio Y, Oka Y, Iwasaki N, Iwamoto Y, Yamada Y, Seino Y, Maegawa H, Kashiwagi A, Wang HY, Tanahashi T, Nakamura N, Takeda J, Maeda E, Yamamoto K, Tokunaga K, Ma RC, So WY, Chan JC, Kamatani N, Makino H, Nanjo K, Kadowaki T, Kasuga M. **Construction of a prediction model for type 2 diabetes mellitus in the Japanese population based on 11 genes with strong evidence of the association.** *J Hum Genet* 2009;**54**(4):236-41.

15. Schulze MB, Weikert C, Pischon T, Bergmann MM, Al-Hasani H, Schleicher E, Fritsche A, Haring HU, Boeing H, Joost HG. **Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study.** *Diabetes Care* 2009;**32**(11):2116-9.

16. Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, Hunter DJ, Hu FB. **Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry.** *Ann Intern Med* 2009; **150**(8):541-50.

17. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, Kivimaki M, Humphries SE. **Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study.** *BMJ* 2010; 340:b4838.

18. Wang J, Stancakova A, Kuusisto J, Laakso M. **Identification of undiagnosed type 2 diabetic individuals by the Finnish diabetes risk score and biochemical and genetic markers: a population-based study of 7232 Finnish men.** *J Clin Endocrinol Metab* 2010;**95**(8):3858-62.

19. Fontaine-Bisson B, Renstrom F, Rolandsson O, Payne F, Hallmans G, Barroso I, Franks PW. **Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population.** *Diabetologia* 2010;**53**(10):2155-62.

20. Ruchat SM, Vohl MC, Weisnagel SJ, Rankinen T, Bouchard C, Perusse L. **Combining genetic markers and clinical risk factors improves the risk assessment of impaired glucose metabolism.** *Ann Med* 2010;**42**(3): 196-206.

21. de Miguel-Yanes JM, Shrader P, Pencina MJ, Fox CS, Manning AK, Grant RW, Dupuis J, Florez JC, D'Agostino RB, Sr., Cupples LA, Meigs JB, Investigators M, Investigators D. **Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms.** *Diabetes Care* 2011;**34**(1):121-5.

22. Qi Q, Li H, Wu Y, Liu C, Wu H, Yu Z, Qi L, Hu FB, Loos RJ, Lin X. **Combined effects of 17 common genetic variants on type 2 diabetes risk in a Han Chinese population.** *Diabetologia* 2010;**53**(10):2163-6.

23. Rotger M, Gsponer T, Martinez R, Taffe P, Elzi L, Vernazza P, Cavassini M, Bernasconi E, Hirschel B, Furrer H, Weber R, Ledergerber B, Egger M, Telenti A, Tarr PE, Swiss HIVCS. **Impact of single nucleotide polymorphisms and of clinical risk factors on new-onset diabetes mellitus in HIV-infected individuals.** *Clin Infect Dis* 2010;**51**(9):1090-8.

24. Xu M, Bi Y, Xu Y, Yu B, Huang Y, Gu L, Wu Y, Zhu X, Li M, Wang T, Song A, Hou J, Li X, Ning G. **Combined effects of 19 common variations on type 2 diabetes in Chinese: results from two community-based studies.** *PloS one* 2010;**5**(11):e14022.

25. Rees SD, Hydrie MZ, Shera AS, Kumar S, O'Hare JP, Barnett AH, Basit A, Kelly MA. **Replication of 13 genome-wide association (GWA)-validated risk variants for type 2 diabetes in Pakistani populations.** *Diabetologia* 2011;**54**(6):1368-74.

26. Hivert MF, Jablonski KA, Perreault L, Saxena R, McAteer JB, Franks PW, Hamman RF, Kahn SE, Haffner S, Consortium D, Meigs JB, Altshuler D, Knowler WC, Florez JC, Diabetes Prevention Program Research G. **Updated genetic score based on 34 confirmed type 2 diabetes Loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program.** *Diabetes* 2011;**60**(4):1340-8.

27. Tabara Y, Osawa H, Kawamoto R, Onuma H, Shimizu I, Makino H, Kohara K, Miki T. **Genotype risk score of common susceptible variants for prediction of type 2 diabetes mellitus in Japanese: the Shimanami Health Promoting Program (J-SHIPP study). Development of type 2 diabetes mellitus and genotype risk score.** *Metabolism* 2011;**60**(11):1634-40.

28. Janipalli CS, Kumar MV, Vinay DG, Sandeep MN, Bhaskar S, Kulkarni SR, Aruna M, Joglekar CV, Priyadharshini S, Maheshwari N, Yajnik CS, Chandak GR. **Analysis of 32 common susceptibility genetic variants and their combined effect in predicting risk of Type 2 diabetes and related traits in Indians.** *Diabet Med* 2012;**29**(1):121-7.

Supplementary Table 2. Description of studies that contributed to the current analyses (study design, outcome definition, exclusion criteria for type 1 diabetes and details on genotyping and imputation techniques).

Study	**ROTTERDAM**	**DGDG**	**FUSION**	**KORA**
Collection type	Population-based	Case-control	Case-control	Population-based
Sample size	7983	679 cases (266 from theUMR8090 CNRS unit in Lille; 413 from the Endocrinology-Diabetology Department of the Corbeil-Essonnes Hospital) 697 controls	1161 cases (372 from Finrisk2002) 1174 controls (651 from Finrisk2002)	S3: 4856 F3: about 4000 participants from S3 S4: 4261
Follow-up times	Baseline: 1990 - 1993 Follow-up: 1993 - 1994 1997 - 1999 2002 - 2004	NA	NA	S3: 1994/1995 F3: 2004/2005 S4: 1999-2001
Individuals without T2D (Random selection or in subgroup)	Random selection	Population based controls mentained normal glucose tolerance over time	Controls were matched to cases on 5-year age category, sex and birth province. some of the controls are spouses of the cases and all controls are NGT.	Random selection
T2D definition	At baseline, prevalent cases of diabetes were diagnosed by a non-fasting or postload glucose level (after oral glucose tolerance testing) ≥11.1 mmol/l and/or treatment with oral antidiabetic medication or insulin. During follow-up, incident diabetes was diagnosed as fasting plasma glucose levels ≥7.0 mmol/l, non-fasting plasma glucose levels ≥11.0 mmol/l and/or treatment with oral antidiabetic medication or insulin. All diagnoses of T2D were confirmed by a general practitioner.	(i) T2D diagnosed according to 1997 American Diabetes Association (ADA) criteria; (ii) a family history of diabetes in first degree relatives; and (iii) BMI <30 kg/m^2.	WHO 1999 criteria of fasting plasma glucose ≥7.0 mmol/l or 2-h plasma glucose ≥11.1 mmol/l, by report of diabetes medication use, or based on medical record review. Part of the case sample (which is all unrelateds) was selected from sibships with type 2 diabetes, and there was an upper age limit on the case recruitment.	Cases were identified by self report of T2D in a personal interview which was validated by a questionnaire mailed to the treating physician and/or by medical chart review.

Study	ROTTERDAM	DGDG	FUSION	KORA
Collection type	Population-based	Case-control	Case-control	Population-based
Control definition	At baseline, controls from the 1990-1993 cohort were defined on the basis of NGT or nonfasting glucose level <11.0 mmol/l, and from the 1999 cohort on the basis of FPG <7.0 mmol/l. During the follow-up, controls from either cohort had to maintain FPG <7.0 mmol/l and have no physician diagnosis of diabetes.	(i) age at examination >45yr; (ii) normal fasting glucose according to 1997 ADA criteria (FPG<5.7 mmol/l); and (iii) BMI <27 kg/m².	Normal glucose tolerance as defined by WHO 1999 criteria.	Controls were non-diabetic as defined by self-report.
Exclusion criteria	Patients registered in general practitioner records as having type 1 diabetes (n = 15). Islet cell specific antibodies were not measured.	Cases with a diagnosis of diabetes below 45 years of age were screened for known MODY mutations. Cases from the Corbeil-Essonnes Hospital were also systematically tested for fasting C-peptide levels: if fasting C-peptide levels were below 0.4 mg/l, subjects were also tested for anti-GAD antibodies: those with anti- GAD antibodies >10U/ml were excluded. A subset (76/266) of the cases from UMR8090 CNRS were tested for anti-islet-antibodies (positive if >5.5% islet beta-cells stained positive) and/or anti-insulin antibodies and positive cases were excluded.	Cases were excluded if they had: (i)known or probable type 1 diabetes amongst first degree relatives; (ii) insulin treatment initiated within 10 years of disease diagnosis, detectable levels of anti-GAD antibodies and fasting C-peptide ≤0.30 nmol/l; or (iii) insulin treatment initiated within 4 years of diagnosis and fasting C-peptide ≤0.30 nmol/l.	No islet autoantibodies were measured.
Genotyping platform	Illumina Human Hap 550 Bead Array	Illumina Human Hap 300 Bead Array	Illumina Human Hap 300 Bead Array	Affymetrix GeneChip Human Mapping 500k Array Set

Study	ROTTERDAM	DGDG	FUSION	KORA
Collection type	Population-based	Case-control	Case-control	Population-based
MAF	>0.01	>0.01	>0.01	>0.01
HWE	$>10^{-6}$	$>10^{-4}$	$>10^{-6}$	$>10^{-6}$
Call rate	>0.98	>0.95	>0.90	>0.95
Imputation program	MACH	IMPUTE	MACH1	IMPUTE
Imputation quality	r^2>0.5	Proper-info > 0.5	r^2>0.3	Proper-info >0.5

Study	WTCCC	CROATIA-Vis	ERF	MICROS	ORCADES
Collection type	Case-control	Population-based	Population-based	Population-based	Population-based
Sample size	1924 T2D cases 1437 NBS controls (1501 UK 1958Birth cohort controls weren't included in the present study due to limited availability of the additional phenotypic data required)	1008	3000	1340	1019
Follow-up times	NA	NA	NA	NA	NA
Individuals without T2D (Random selection or in subgroup)	Controls were selected without reference to T2D status, and were representative of gender and each geographical region.	Genetically isolated population	Genetically isolated population	Genetically isolated population	Genetically isolated population

Study	WTCCC	CROATIA-Vis	ERF	MICROS	ORCADES
Collection type	Case-control	Population-based	Population-based	Population-based	Population-based
	Current prescribed treatment with sulphonylureas, biguanides, other oral agents and/or insulin or, in the case of individuals treated with diet alone, historical or contemporary laboratory evidence of hyperglycemia (as defined by WHO).				
T2D definition	Approximately 30% of cases were explicitly recruited as part of multiplex sibships and ~25% were offspring in parent offspring "trios" or "duos" (that is families comprising only one parent complemented by additional sibs). Isolated cases were (compared to population-based cases) of relatively early onset and had a high proportion of T2D parents and/or siblings.	FPG >=7.0mmol/l and/ or use of anti-diabetic medication.	FPG >=7.0mmol/l and/ or use of anti-diabetic medication.	FPG >=7.0mmol/l and/ or use of anti-diabetic medication.	FPG >=7.0mmol/l and/ or use of anti-diabetic medication.

Study	WTCCC	CROATIA-Vis	ERF	MICROS	ORCADES
Collection type	Case-control	Population-based	Population-based	Population-based	Population-based
Control definition	Ascertainment was distributed throughout the UK by the three national UK Blood Services. All subjects gave written informed consent and the project protocols were approved by the relevant research ethics committees in the UK.	Normoglycemic and no history of medication for T2D.	Normoglycemic and no history of medication for T2D.	Normoglycemic and did no history of medication for T2D.	Normoglycemic and no history of medication for T2D.
Exclusion criteria	Known forms of diabetes, e.g. maturity-onset diabetes of the young and mitochondrial diabetes. First-degree relatives with type 1 diabetes. An interval of <1 year between diagnosis and institution of regular insulin therapy. Positive testing for antibodies to glutamic acid decarboxylase (anti-GAD). An anti-GAD titer >10 U (corresponding to ~8 SD above the mean of 88 normal control subjects) in duplicate samples was considered positive.	No islet autoantibodies were measured.	No islet autoantibodies were measured.	No islet autoantibodies were measured.	No islet autoantibodies were measured.
Genotyping platform	Affymetrix GeneChip Human Mapping 500k Array Set	Illumina Human Hap 300 Bead Array	Illumina Human Hap 300 Bead Array	Illumina Human Hap 300 Bead Array	Illumina Human Hap 300 Bead Array

Study	WTCCC	CROATIA-Vis	ERF	MICROS	ORCADES
Collection type	Case-control	Population-based	Population-based	Population-based	Population-based
MAF	>0.01	>0.01	>0.01	>0.01	>0.01
HWE	$>10^{-3}$	$>10^{-6}$	$>10^{-6}$	$>10^{-6}$	$>10^{-6}$
Call rate	>0.95 for MAF>0.05 >0.99 for MAF<0.05	>0.98	>0.98	>0.98	>0.98
Imputation program	IMPUTE	MACH1	MACH1	MACH1	MACH1
Imputation quality	Proper-info >0.4	$r^2>0.5$	$r^2>0.5$	$r^2>0.5$	$r^2>0.5$

Supplementary Table 3. Descriptive characteristics for each study.

Study (reference)	Rotterdam (Hofman. Eur J Epidemiol 2007)			DGDG (Sladek. Nature 2007)			FUSION (Valle. Diabetes Care 1998)		
	All participants	**Subjects without type 2 diabetes**	**Incident and prevalent cases with type 2 diabetes**	**All participants**	**Subjects without type 2 diabetes**	**Subjects with type 2 diabetes**	**All participants**	**Subjects without type 2 diabetes**	**Subjects with type 2 diabetes**
n	5712	4593	1119	1376	697	679	2263	1171	1092
Men (%)	41.3	41.1	42.2	50.4	40.3	60.8	52.8	48.8	57.1
Age (years, SD)	69.0 (8.8)	68.6 (8.7)	70.9 (8.8)	56.4 (8.7)	53.5 (5.7)	59.5 (10.1)	63.2 (7.5)	63.6 (7.4)	62.7 (7.6)
BMI (kg/m2, SD)	26.3 (3.7)	26.0 (3.6)	27.4 (4.0)	24.5 (2.7)	23.2 (1.8)	25.9 (2.8)	28.6 (4.5)	27.1 (3.9)	30.2 (4.7)
Waist circumference (cm, SD)	90.5 (11.1)	89.7 (11.0)	94.2 (10.9)	82.4 (11.1)	79.1 (8.7)	92.2 (11.6)	96.5 (12.9)	91.4 (11.5)	102.0 (12.0)
Systolic blood pressure (mmHg, SD)	139.2 (22.2)	137.8 (21.9)	145.2 (22.5)	134.1 (17.8)	129.7 (16.7)	139.6 (17.6)	147.1 (21.1)	142.8 (19.4)	151.7 (21.8)
Diastolic blood pressure (mmHg, SD)	73.7 (11.4)	73.6 (11.3)	74.2 (11.9)	79.3 (9.7)	79.5 (10.1)	79.2 (9.3)	83.1 (10.8)	82.1 (10.2)	84.2 (11.3)
Hypertension (%)[a]	55.2	51.7	69.8	49.3	33.8	68.2	74.5	64.6	85.1
Total cholesterol (mmol/l, SD)	6.6 (1.21)	6.6 (1.22)	6.6 (1.17)	5.57 (0.99)	5.79 (0.89)	5.35 (1.04)	5.8 (1.12)	5.8 (1.00)	5.7 (1.23)
HDL cholesterol (mmol/l, SD)	1.3 (0.37)	1.4 (0.37)	1.2 (0.35)	1.56 (0.47)	1.75 (0.44)	1.36 (0.4)	1.3 (0.40)	1.5 (0.41)	1.2 (0.33)
Current smoking (%)	22.9	22.7	23.9	25.4	13.1	38.1	5.9	4.6	7.3
Former smoking (%)	41.3	41.6	39.9	16.8	21.4	12.2	13.1	4.5	22.3

Study (reference)	KORA (Wichmann. *Gesundheitswesen* 2005)			WTCCC (WTCCC. Nature 2007)			CROATIA-Vis (Rudan. *Coll Antropol* 1999)		
	All participants	Subjects without type 2 diabetes	Incident and prevalent cases with type 2 diabetes	All participants	Subjects without type 2 diabetes	Subjects with type 2 diabetes	All participants	Subjects without type 2 diabetes	Subjects with type 2 diabetes
n	1861	1432	429	3340	1437	1903	770	660	110
Men (%)	50.7	48.3	59	53.9	48.3	58.1	42.20	41.80	44.60
Age (years, SD)	62.6 (9.9)	61.8 (10.2)	65.1 (8.2)	52.0 (13.4)	43.4 (12.4)	58.5 (10.1)	56.5 (15.4)	55.0 (15.4)	65.2 (12.6)
BMI (kg/m2, SD)	28.5 (4.6)	27.7 (4.3)	30.9 (5.0)	29.1 (6.0)	26.2 (4.3)	31.2 (6.1)	27.3 (4.3)	27.2 (4.2)	28.4 (4.3)
Waist circumference (cm, SD)	97.3 (12.7)	95.4 (12.2)	103.3 (12.7)	NA	NA	NA	96.0 (11.8)	95.2 (11.7)	100.5 (11.3)
Systolic blood pressure (mmHg, SD)	135.8 (20.5)	133.6 (20.0)	142.9 (20.4)	NA	NA	NA	138.1 (23.6)	135.8 (22.2)	151.8 (27.1)
Diastolic blood pressure (mmHg, SD)	82.4 (11.0)	82.6 (10.8)	81.4 (11.6)	NA	NA	NA	80.8 (10.7)	80.6 (10.4)	81.6 (12.7)
Hypertension (%)[a]	63	57.6	80.8	NA	NA	NA	51.9	48.6	71.8
Total cholesterol (mmol/l, SD)	5.8 (1.08)	5.8 (1.02)	5.8 (1.24)	NA	NA	NA	5.1 (1.00)	5.1 (0.99)	4.9 (1.01)
HDL cholesterol (mmol/l, SD)	1.5 (0.44)	1.5 (0.43)	1.3 (0.39)	NA	NA	NA	1.1 (0.16)	1.1 (0.16)	1.1 (0.15)
Current smoking (%)	14.4	13.4	17.8	NA	NA	NA	27.9	29.1	20.1
Former smoking (%)	36.3	35.2	40	NA	NA	NA	28.4	29.3	23.6

Study (reference)	ERF (Aulchenko. Eur J Hum Genet 2004)			MICROS (Pataro. BMC medical genetics 2007)			ORCADES (Helgason. Ann Hum Genet 2003)		
	All participants	Subjects without type 2 diabetes	Incident and prevalent cases with type 2 diabetes	All participants	Subjects without type 2 diabetes	Subjects with type 2 diabetes	All participants	Subjects without type 2 diabetes	Subjects with type 2 diabetes
n	1229	1178	51	1061	1018	43	695	668	27
Men (%)	39.80	39.30	51.00	43.30	42.90	51.20	46.60	45.80	66.70
Age (years, SD)	48.4 (14.7)	47.9 (14.6)	60.3 (11.5)	45.0 (16.0)	44.3 (15.7)	62.9 (12.6)	53.8 (15.3)	53.5 (15.4)	62.4 (10.6)
BMI (kg/m2, SD)	26.7 (4.7)	26.6 (4.6)	30.7 (5.2)	25.6 (4.8)	25.4 (4.5)	30.6 (8.6)	27.8 (4.9)	27.5 (4.7)	33.1 (5.8)
Waist circumference (cm, SD)	86.96(13.56)	86.4(13.2)	100.6(13.5)	88.0 (13.1)	87.6 (13.1)	99.9 (7.3)	93.7 (13.9), n=694	93.1 (13.4), n=667	109.7 (16.3)
Systolic blood pressure (mmHg, SD)	139.3(20.7)	138.6(20.5)	155.6(19.4)	132.9 (20.3)	132.3 (20.0)	151.0 (21.1)	131 (18)	131 (18)	140 (14)
Diastolic blood pressure (mmHg, SD)	79.9(10.0)	79.7(10.0)	83.6(9.6)	79.4 (11.2)	79.3 (11.2)	82.6 (11.2)	76.4 (10.1), n=693	76.2 (10.1), n=666	78.9 (8.9)
Hypertension (%)[a]	46.1	44.3	79	17.9	16.7	54.5	41.7	40.3	77.8
Total cholesterol (mmol/l, SD)	5.58(1.10)	5.6(1.1)	5.1(1.2)	5.38 (1.08)	5.37 (1.08)	5.52 (1.03)	5.45 (1.13)	5.49 (1.12)	4.65 (0.97)
HDL cholesterol (mmol/l, SD)	1.28(0.36)	1.28(0.36)	1.06(0.29)	1.53 (0.38)	1.53 (0.38)	1.40 (0.40)	1.56 (0.43)	1.57 (0.43)	1.26 (0.37)
Current smoking (%)	41.5	41.6	38.9	NA	NA	NA	8.65	8.55, n=57/667	11.1
Former smoking (%)	29.1	28.6	40.7	NA	NA	NA	33.1	32.4, n=216/667	57.9

[a]Hypertension was defined as systolic blood pressure ≥140 or diastolic blood pressure ≥90 or blood pressure lowering medication with indication hypertension

Supplementary Table 4. Risk allele frequencies in individuals without and with type 2 diabetes reported per study.

					ROTTERDAM		DGDG		FUSION		KORA		WTCCC		CROATIA-Vis		ERF		MICROS		ORCADES	
					Observed risk allele frequency																	
SNP	Nearest gene	Chr	Risk allele	Nonrisk allele	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca
rs10440833	*CDKAL1*	6	A	T	0.27	0.30	0.28	0.33	0.34	0.38	0.27	0.33	0.27	0.32	0.26	0.30	0.31	0.23	0.27	0.26	0.20	0.30
rs10923931	*NOTCH2*	1	T	G	0.11	0.12	0.11	0.11	0.13	0.13	0.09	0.08	0.11	0.12	0.10	0.08	0.11	0.15	0.10	0.11	0.10	0.15
rs10965250	*CDKN2A/B*	9	G	A	0.82	0.83	0.81	0.86	0.85	0.87	0.84	0.83	0.84	0.87	0.86	0.90	0.87	0.85	0.82	0.83	0.85	0.80
rs11634397	*ZFAND6*	15	G	A	0.67	0.67	0.65	0.68	0.61	0.63	0.66	0.66	0.65	0.69	0.64	0.70	0.62	0.60	0.69	0.62	0.66	0.70
rs11642841	*FTO*	16	A	C	0.39	0.40	0.42	0.47	0.44	0.44	0.41	0.43	0.42	0.47	0.45	0.45	0.44	0.51	0.42	0.38	0.38	0.31
rs11708067	*ADCY5*	3	A	G	0.74	0.76	0.81	0.83	0.82	0.83	0.80	0.81	0.78	0.79	0.85	0.83	0.69	0.68	0.82	0.81	0.76	0.81
rs11899863	*THADA*	2	C	T	0.89	0.89	0.89	0.92	0.95	0.96	0.90	0.89	0.89	0.90	0.90	0.91	0.94	0.98	0.86	0.86	0.92	0.93
rs12779790	*CDC123/ CAMK1D*	10	G	A	0.18	0.18	0.17	0.20	0.19	0.22	0.18	0.16	0.22	0.24	0.19	0.16	0.18	0.19	0.17	0.12	0.25	0.24
rs13081389	*PPARG*	3	A	G	0.93	0.94	0.95	0.96	0.90	0.93	0.93	0.95	0.95	0.96	0.96	0.97	0.94	0.93	0.95	1.00	0.92	0.91
rs13292136	*CHCHD9*	9	C	T	0.92	0.93	0.94	0.95	0.86	0.88	0.94	0.95	0.93	0.94	0.97	0.98	0.94	0.94	0.95	0.94	0.89	0.94
rs1387153	*MTNR1B*	11	T	C	0.27	0.29	0.27	0.32	0.32	0.35	0.27	0.26	0.27	0.28	0.26	0.31	0.28	0.29	0.28	0.29	0.24	0.26
rs1470579	*IGF2BP2*	3	C	A	0.30	0.32	0.33	0.33	0.29	0.34	0.27	0.30	0.32	0.35	0.30	0.26	0.29	0.30	0.26	0.27	0.29	0.32
rs1531343	*HMGA2*	12	C	G	0.10	0.11	0.10	0.13	0.07	0.08	0.10	0.10	0.09	0.11	0.12	0.14	0.14	0.11	0.11	0.17	0.09	0.15
rs1552224	*CENTD2*	11	A	C	0.84	0.84	0.85	0.88	0.76	0.79	0.86	0.89	0.84	0.85	0.87	0.91	0.76	0.76	0.84	0.86	0.89	0.87
rs163184	*KCNQ1*	11	G	T	0.48	0.49	0.47	0.51	0.63	0.65	0.48	0.49	0.47	0.49	0.48	0.45	0.47	0.56	0.47	0.42	0.52	0.57
rs1801214	*WFS1*	4	T	C	0.57	0.60	0.61	0.69	0.55	0.58	0.60	0.60	0.60	0.61	0.54	0.59	0.60	0.72	0.63	0.72	0.53	0.38
rs2191349	*DGKB-TMEM195*	7	T	G	0.54	0.54	0.54	0.56	0.45	0.46	0.54	0.57	0.54	0.57	0.56	0.59	0.54	0.57	0.58	0.62	0.50	0.43
rs2237892	*KCNQ1*	11	C	T	0.95	0.95	0.94	0.95	0.95	0.95	0.96	0.96	0.94	0.94	0.95	0.95	0.96	0.96	0.94	0.90	0.94	0.96
rs231362	*KCNQ1*	11	G	A	0.49	0.51	0.47	0.48	0.48	0.52	0.42	0.46	0.47	0.48	0.48	0.49	0.45	0.43	0.50	0.44	0.49	0.48

					ROTTERDAM		DGDG		FUSION		KORA		WTCCC		CROATIA-Vis		ERF		MICROS		ORCADES	
					Observed risk allele frequency																	
SNP	Nearest gene	Chr	Risk allele	Nonrisk allele	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca
rs243021	*BCL11A*	2	A	G	0.46	0.47	0.45	0.48	0.45	0.47	0.46	0.49	0.45	0.48	0.50	0.49	0.42	0.43	0.52	0.55	0.47	0.39
rs340874	*PROX1*	1	C	T	0.56	0.58	0.55	0.56	0.43	0.44	0.53	0.55	0.59	0.61	0.47	0.49	0.50	0.54	0.60	0.55	0.62	0.57
rs3802177	*SLC30A8*	8	G	A	0.70	0.72	0.69	0.76	0.61	0.65	0.79	0.81	0.72	0.73	0.71	0.69	0.68	0.69	0.71	0.81	0.73	0.82
rs4430796	*HNF1B (TCF2)*	17	G	A	0.48	0.50	0.50	0.59	0.35	0.37	0.56	0.56	0.50	0.51	0.49	0.52	0.43	0.36	0.51	0.56	0.51	0.47
rs4457053	*ZBED3*	5	G	A	0.28	0.30	0.18	0.21	0.22	0.25	0.26	0.31	0.28	0.31	0.29	0.34	0.25	0.27	0.28	0.31	0.35	0.34
rs4607517	*GCK*	7	A	G	0.18	0.17	0.18	0.20	0.10	0.11	0.19	0.19	0.19	0.18	0.15	0.15	0.20	0.17	0.24	0.23	0.23	0.29
rs4760790	*TSPAN8/LGR5*	12	A	G	0.28	0.30	0.28	0.32	0.21	0.22	0.28	0.29	0.28	0.33	0.27	0.24	0.31	0.36	0.30	0.36	0.27	0.32
rs5015480	*HHEX/IDE*	10	C	T	0.59	0.61	0.58	0.67	0.53	0.56	0.60	0.63	0.60	0.62	0.60	0.63	0.57	0.62	0.56	0.71	0.62	0.72
rs5215	*KCNJ11*	11	C	T	0.36	0.38	0.39	0.38	0.45	0.49	0.37	0.40	0.35	0.38	0.38	0.35	0.33	0.33	0.33	0.31	0.33	0.26
rs6795735	*ADAMTS9*	3	C	T	0.59	0.62	0.59	0.56	0.65	0.67	0.55	0.56	0.59	0.62	0.52	0.52	0.58	0.47	0.53	0.54	0.59	0.69
rs7578326	*IRS1*	2	A	G	0.64	0.65	0.65	0.71	0.63	0.65	0.65	0.66	0.65	0.67	0.59	0.64	0.67	0.66	0.61	0.64	0.65	0.72
rs780094	*GCKR*	2	C	T	0.63	0.62	0.55	0.57	0.66	0.65	0.57	0.57	0.60	0.60	0.55	0.58	0.69	0.65	0.57	0.58	0.65	0.69
rs7903146	*TCF7L2*	10	T	C	0.28	0.34	0.29	0.47	0.18	0.23	0.29	0.33	0.31	0.39	0.29	0.33	0.29	0.32	0.32	0.37	0.26	0.30
rs7957197	*HNF1A*	12	T	A	0.81	0.82	0.79	0.80	0.78	0.79	0.81	0.84	0.82	0.84	0.84	0.88	0.82	0.88	0.75	0.75	0.79	0.80
rs8042680	*PRC1*	15	A	C	0.31	0.33	0.32	0.38	0.32	0.35	0.32	0.31	0.30	0.32	0.33	0.35	0.37	0.37	0.32	0.35	0.36	0.41
rs849134	*JAZF1*	7	A	G	0.52	0.54	0.50	0.55	0.50	0.53	0.52	0.54	0.49	0.53	0.52	0.54	0.57	0.64	0.50	0.64	0.54	0.55
rs896854	*TP53INP1*	8	T	C	0.52	0.53	0.48	0.50	0.47	0.50	0.52	0.56	0.48	0.51	0.51	0.53	0.50	0.60	0.51	0.55	0.53	0.44
rs972283	*KLF14*	7	G	A	0.51	0.54	0.55	0.57	0.54	0.54	0.52	0.54	0.51	0.54	0.47	0.43	0.44	0.49	0.56	0.55	0.51	0.57

Ca, cases; Co, controls.

Supplementary Table 5. Beta coefficients for the three type 2 diabetes risk prediction models.

Clinical model									
	ROTTERDAM	**DGDG**	**FUSION**	**KORA**	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
Intercept	-6.353	-15.021	-54.418	-7.615	-12.250	-6.016	-10.759	-10.459	-11.191
AGE	0.030	0.079	-0.003	0.032	0.133	0.050	0.062	0.065	0.035
SEX_Male	0.240	0.467	0.501	0.265	0.220	0.178	0.364	0.586	0.931
BMI	0.102	0.419	0.180	0.148	0.229	0.041	0.143	0.125	0.180
Genetic model									
	ROTTERDAM	DGDG	FUSION	KORA	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
Intercept	-4.673	-8.671	5.230	-4.656	0.872	-6.704	-6.913	-5.393	-6.787
rs10923931_T	0.025	-0.045	0.024	-0.149	0.091	-0.321	0.341	0.240	0.880
rs11899863_C	0.066	0.477	0.399	-0.065	0.254	0.045	1.289	0.039	0.440
rs13081389_A	0.126	0.222	0.458	0.284	0.296	0.547	-0.106	0.000	-0.563
rs6795735_C	0.130	-0.065	0.059	0.030	0.144	0.006	-0.400	0.019	0.581
rs1470579_C	0.121	-0.003	0.228	0.131	0.173	-0.316	0.039	-0.022	-0.093
rs1801214_T	0.139	0.346	0.120	-0.009	0.062	0.151	0.599	0.501	-1.078
rs10440833_A	0.190	0.260	0.214	0.251	0.256	0.112	-0.383	-0.143	0.884
rs849134_A	0.080	0.215	0.143	0.029	0.144	0.101	0.225	0.684	-0.055
rs3802177_G	0.128	0.252	0.168	0.139	0.092	-0.060	-0.055	0.549	0.757
rs10965250_G	0.071	0.429	0.155	-0.042	0.250	0.462	-0.151	0.024	-0.321
rs12779790_G	0.031	0.154	0.161	-0.124	0.178	-0.316	0.032	-0.513	-0.275
rs5015480_C	0.087	0.444	0.142	0.126	0.118	0.181	0.135	0.665	0.610
rs7903146_T	0.300	0.761	0.329	0.181	0.360	0.164	0.004	0.224	0.433
rs163184_G	0.036	0.107	0.056	0.012	0.056	-0.126	0.580	-0.215	0.308
rs2237892_C	0.053	0.206	-0.082	0.150	0.042	0.147	-0.392	-0.559	0.222

Genetic model									
	ROTTERDAM	DGDG	FUSION	KORA	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
rs5215_C	0.057	-0.076	0.166	0.114	0.153	-0.209	0.002	0.020	-0.489
rs4760790_A	0.092	0.154	0.068	0.066	0.192	-0.187	0.240	0.333	0.321
rs11642841_A	0.080	0.202	0.037	0.093	0.219	0.062	0.388	-0.052	-0.411
rs4430796_G	0.067	0.322	0.116	-0.018	0.167	0.151	-0.423	0.138	-0.601
rs1387153_T	0.136	0.317	0.162	-0.039	0.069	0.299	-0.003	0.008	0.092
rs7578326_A	0.053	0.300	0.064	0.017	0.077	0.180	-0.139	0.096	0.671
rs243021_A	0.027	0.112	0.069	0.130	0.157	-0.108	0.027	0.130	-0.700
rs4457053_G	0.141	0.041	0.185	0.302	0.253	0.275	0.119	-0.013	-0.149
rs972283_G	0.107	0.049	0.017	0.060	0.111	-0.184	0.192	-0.224	0.282
rs896854_T	0.040	0.105	0.117	0.173	0.101	0.092	0.457	0.245	-0.256
rs13292136_C	0.174	0.289	0.095	0.091	0.229	0.131	0.063	-0.295	0.991
rs231362_G	0.035	0.021	0.185	0.101	0.046	0.104	-0.159	-0.260	0.056
rs1552224_A	0.030	0.307	0.186	0.289	0.070	0.309	0.008	0.218	-0.711
rs1531343_C	0.159	0.204	0.151	-0.095	0.263	0.282	-0.397	0.533	0.956
rs7957197_T	0.069	0.175	0.111	0.175	0.121	0.424	0.672	0.094	0.024
rs11634397_G	0.038	0.058	0.078	0.028	0.233	0.229	-0.239	-0.315	0.598
rs8042680_A	0.083	0.299	0.153	-0.039	0.100	0.017	-0.086	0.100	0.092
rs4607517_A	-0.007	0.139	0.187	0.031	-0.042	0.048	-0.286	-0.167	0.266
rs2191349_T	0.016	0.134	0.034	0.133	0.124	0.133	0.071	0.154	-0.314
rs780094_C	-0.038	0.170	-0.007	-0.014	0.023	0.098	-0.224	0.114	-0.196
rs11708067_A	0.151	0.160	0.127	0.066	0.135	-0.141	-0.048	-0.017	0.424
rs340874_C	0.092	0.076	0.085	0.069	0.124	0.146	0.178	-0.119	-0.287

Combined clinical and genetic model									
	ROTTERDAM	DGDG	FUSION	KORA	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
Intercept	-9.864	-24.410	-11.509	-11.757	-13.143	-12.119	-18.740	-15.966	-13.994
AGE	0.031	0.079	0.000	0.034	0.140	0.055	0.078	0.075	0.048
SEX_Male	0.235	0.642	0.525	0.276	0.241	0.135	0.168	0.772	0.991
BMI	0.105	0.440	0.187	0.157	0.243	0.049	0.167	0.162	0.219
rs10923931_T	0.020	-0.100	0.094	-0.164	0.296	-0.227	0.418	0.360	0.792
rs11899863_C	0.050	0.568	0.473	-0.057	0.432	0.076	1.710	0.288	0.124
rs13081389_A	0.139	0.071	0.454	0.243	0.296	0.575	-0.025	0.000	-0.840
rs6795735_C	0.114	-0.033	0.104	0.074	0.194	-0.087	-0.399	0.244	0.339
rs1470579_C	0.123	0.035	0.203	0.131	0.248	-0.280	0.057	0.206	-0.010
rs1801214_T	0.143	0.150	0.107	0.004	0.051	0.146	0.641	0.429	-1.355
rs10440833_A	0.204	0.275	0.190	0.334	0.320	0.123	-0.495	-0.201	0.918
rs849134_A	0.085	0.301	0.142	0.054	0.185	0.119	0.285	0.879	-0.202
rs3802177_G	0.146	0.440	0.243	0.229	0.198	-0.069	-0.076	0.522	0.757
rs10965250_G	0.078	0.220	0.177	-0.048	0.191	0.472	-0.204	0.041	-0.569
rs12779790_G	0.028	0.071	0.161	-0.109	0.357	-0.352	-0.007	-0.823	-0.047
rs5015480_C	0.074	0.487	0.155	0.177	0.151	0.275	0.381	0.753	0.920
rs7903146_T	0.327	0.758	0.401	0.149	0.538	0.233	-0.033	0.322	0.539
rs163184_G	0.042	0.148	0.042	-0.012	0.117	-0.180	0.577	-0.165	0.383
rs2237892_C	0.064	0.067	-0.047	0.054	0.113	0.229	-0.250	-0.786	-0.505
rs5215_C	0.057	-0.019	0.187	0.133	0.152	-0.265	-0.021	-0.055	-0.199
rs4760790_A	0.093	0.388	0.064	0.100	0.250	-0.221	0.232	0.422	0.383
rs11642841_A	0.063	0.188	-0.017	0.095	0.121	0.005	0.408	-0.290	-0.569
rs4430796_G	0.057	0.318	0.125	0.042	0.290	0.196	-0.551	0.231	-1.031
rs1387153_T	0.126	0.457	0.155	-0.024	0.287	0.365	0.112	0.013	0.042

Combined clinical and genetic model									
	ROTTERDAM	DGDG	FUSION	KORA	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
rs7578326_A	0.061	0.237	0.102	0.084	-0.064	0.154	-0.003	0.059	0.776
rs243021_A	0.033	0.133	0.093	0.122	0.081	-0.126	0.084	0.082	-0.571
rs4457053_G	0.136	0.231	0.172	0.320	0.336	0.324	-0.027	0.186	-0.145
rs972283_G	0.110	0.091	0.007	0.082	0.198	-0.237	0.376	-0.225	0.014
rs896854_T	0.037	0.127	0.194	0.140	0.097	0.106	0.441	0.257	-0.356
rs13292136_C	0.166	0.519	0.163	0.167	0.239	0.159	0.228	-0.282	1.084
rs231362_G	0.031	0.148	0.167	0.152	0.088	0.106	-0.048	-0.231	0.152
rs1552224_A	0.054	0.264	0.168	0.247	0.202	0.354	0.225	0.184	-0.840
rs1531343_C	0.178	0.334	0.180	-0.049	0.280	0.478	-0.471	0.626	1.288
rs7957197_T	0.066	0.139	0.116	0.167	0.100	0.380	0.861	0.386	0.245
rs11634397_G	0.047	-0.024	0.076	0.027	0.253	0.268	-0.106	-0.196	0.743
rs8042680_A	0.083	0.289	0.090	-0.033	0.158	0.140	-0.052	0.086	-0.263
rs4607517_A	-0.006	0.103	0.163	0.058	-0.082	-0.030	-0.443	-0.187	0.548
rs2191349_T	0.037	0.126	-0.002	0.117	0.159	0.132	0.077	0.333	0.009
rs780094_C	-0.042	0.168	0.009	-0.055	-0.027	0.087	-0.296	0.142	-0.246
rs11708067_A	0.145	0.161	0.155	0.104	0.271	-0.069	-0.158	0.029	0.319
rs340874_C	0.103	0.027	0.067	0.022	0.097	0.206	0.164	-0.045	-0.403

Supplementary Table 6. Area under ROC curves (95% confidence intervals) for the clinical, genetic and combined type 2 diabetes prediction models.

Clinical model										
		Developed in								
Validated in		**Rotterdam**	**DGDG**	**FUSION**	**KORA**	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
	Rotterdam	**0.64** (**0.62 – 0.65**)	0.63 (0.62 – 0.65)	0.6 (0.58 – 0.62)	0.63 (0.62 – 0.65)	0.63 (0.61 – 0.65)	0.61 (0.59 – 0.63)	0.64 (0.62 – 0.65)	0.63 (0.61 – 0.65)	0.62 (0.60 – 0.64)
	DGDG	0.82 (0.80 – 0.84)	**0.83** (**0.80 – 0.85**)	0.79 (0.76 – 0.81)	0.83 (0.80 – 0.85)	0.8 (0.77 – 0.82)	0.77 (0.74 – 0.79)	0.81 (0.79 – 0.83)	0.8 (0.78 – 0.83)	0.8 (0.78 – 0.82)
	FUSION	0.68 (0.66 – 0.70)	0.69 (0.67 – 0.71)	**0.71** (**0.69 – 0.73**)	0.69 (0.67 – 0.71)	0.63 (0.61 – 0.65)	0.57 (0.54 – 0.59)	0.66 (0.64 – 0.68)	0.65 (0.63 – 0.67)	0.69 (0.67 – 0.71)
	KORA	0.68 (0.66 – 0.71)	0.69 (0.67 – 0.72)	0.66 (0.63 – 0.69)	**0.69** (**0.66 – 0.72**)	0.68 (0.66 – 0.71)	0.64 (0.62 – 0.67)	0.68 (0.65 – 0.71)	0.66 (0.63 – 0.69)	0.65 (0.63 – 0.68)
	WTCCC	0.88 (0.87 – 0.89)	0.86 (0.85 – 0.88)	0.74 (0.73 – 0.76)	0.87 (0.85 – 0.88)	**0.9** (**0.89 – 0.91**)	0.88 (0.87 – 0.89)	0.89 (0.88 – 0.90)	0.88 (0.87 – 0.89)	0.83 (0.81 – 0.84)
	CROAS	0.68 (0.63 – 0.73)	0.66 (0.61 – 0.72)	0.58 (0.52 – 0.64)	0.67 (0.62 – 0.72)	0.7 (0.65 – 0.75)	**0.7** (**0.65 – 0.75**)	0.69 (0.64 – 0.74)	0.69 (0.64 – 0.74)	0.65 (0.60 – 0.70)
	ERF	0.8 (0.75 – 0.85)	0.79 (0.74 – 0.85)	0.72 (0.65 – 0.79)	0.79 (0.74 – 0.85)	0.8 (0.75 – 0.86)	0.79 (0.73 – 0.85)	**0.8** (**0.75 – 0.86**)	0.8 (0.74 – 0.86)	0.78 (0.72 – 0.84)
	MICROS	0.84 (0.78 – 0.89)	0.82 (0.75 – 0.88)	0.73 (0.65 – 0.81)	0.82 (0.76 – 0.89)	0.84 (0.79 – 0.90)	0.84 (0.79 – 0.89)	0.84 (0.79 – 0.90)	**0.85** (**0.79 – 0.90**)	0.82 (0.76 – 0.88)
	ORCADES	0.81 (0.72 – 0.90)	0.81 (0.72 – 0.90)	0.81 (0.71 – 0.90)	0.81 (0.72 – 0.91)	0.78 (0.70 – 0.87)	0.75 (0.67 – 0.84)	0.8 (0.72 – 0.89)	0.8 (0.71 – 0.88)	**0.82** (**0.73 – 0.90**)

Genetic model		Developed in								
Validated in		Rotterdam	DGDG	FUSION	KORA	WTCCC	CROATIA-Vis	ERF	MICROS	ORCADES
		0.61	0.58	0.59	0.57	0.59	0.54	0.51	0.53	0.53
	Rotterdam	**(0.59 – 0.63)**	(0.56 – 0.60)	(0.57 – 0.61)	(0.55 – 0.59)	(0.58 – 0.61)	(0.52 – 0.55)	(0.50 – 0.53)	(0.52 – 0.55)	(0.51 – 0.55)
		0.7	**0.74**	0.7	0.62	0.69	0.62	0.54	0.61	0.54
	DGDG	(0.67 – 0.72)	**(0.72 – 0.77)**	(0.67 – 0.73)	(0.59 – 0.65)	(0.66 – 0.72)	(0.59 – 0.65)	(0.51 – 0.57)	(0.58 – 0.64)	(0.51 – 0.57)
		0.62	0.61	**0.64**	0.6	0.62	0.55	0.52	0.55	0.52
	FUSION	(0.60 – 0.64)	(0.59 – 0.64)	**(0.62 – 0.67)**	(0.58 – 0.62)	(0.60 – 0.64)	(0.53 – 0.57)	(0.50 – 0.54)	(0.52 – 0.57)	(0.50 – 0.55)
		0.58	0.56	0.58	**0.63**	0.58	0.55	0.51	0.52	0.5
	KORA	(0.55 – 0.61)	(0.53 – 0.59)	(0.55 – 0.61)	**(0.60 – 0.66)**	(0.55 – 0.62)	(0.52 – 0.58)	(0.48 – 0.54)	(0.49 – 0.55)	(0.47 – 0.53)
		0.68	0.66	0.67	0.65	**0.69**	0.61	0.61	0.61	0.61
	WTCCC	(0.66 – 0.69)	(0.64 – 0.68)	(0.65 – 0.69)	(0.63 – 0.67)	**(0.67 – 0.71)**	(0.59 – 0.63)	(0.59 – 0.63)	(0.59 – 0.63)	(0.59 – 0.63)
		0.56	0.59	0.56	0.56	0.57	**0.68**	0.49	0.55	0.5
	CROAS	(0.50 – 0.62)	(0.53 – 0.65)	(0.50 – 0.62)	(0.51 – 0.62)	(0.51 – 0.63)	**(0.63 – 0.74)**	(0.43 – 0.55)	(0.49 – 0.60)	(0.44 – 0.55)
		0.55	0.57	0.54	0.56	0.56	0.51	**0.75**	0.6	0.56
	ERF	(0.46 – 0.63)	(0.49 – 0.64)	(0.46 – 0.62)	(0.48 – 0.64)	(0.48 – 0.64)	(0.43 – 0.59)	**(0.69 – 0.81)**	(0.52 – 0.68)	(0.49 – 0.63)
		0.6	0.62	0.6	0.58	0.59	0.58	0.55	**0.76**	0.5
	MICROS	(0.52 – 0.68)	(0.54 – 0.71)	(0.51 – 0.68)	(0.49 – 0.67)	(0.50 – 0.67)	(0.50 – 0.67)	(0.45 – 0.64)	**(0.69 – 0.83)**	(0.42 – 0.58)
		0.58	0.56	0.55	0.5	0.55	0.54	0.58	0.5	**0.83**
	ORCADES	(0.47 – 0.69)	(0.47 – 0.66)	(0.44 – 0.65)	(0.39 – 0.61)	(0.44 – 0.65)	(0.43 – 0.65)	(0.48 – 0.68)	(0.41 – 0.59)	**(0.77 – 0.90)**

Combined clinical and genetic model

		Developed in								
Validated in		**Rotterdam**	**DGDG**	**FUSION**	**KORA**	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
		0.67	0.66	0.64	0.65	0.66	0.59	0.58	0.61	0.57
	Rotterdam	**(0.66 – 0.69)**	(0.64 – 0.68)	(0.62 – 0.66)	(0.63 – 0.67)	(0.64 – 0.68)	(0.57 – 0.61)	(0.56– 0.60)	(0.59 – 0.62)	(0.55 – 0.59)
		0.85	**0.88**	0.83	0.83	0.83	0.74	0.7	0 .76	0.66
	DGDG	(0.83 – 0.87)	**(0.86 – 0.89)**	(0.81 – 0.85)	(0.81 – 0.85)	(0.81 – 0.85)	(0.71 – 0.77)	(0.67 – 0.72)	(0.73 – 0.78)	(0.63 – 0.69)
		0.72	0.72	**0.74**	0.72	0.66	0.6	0.63	0.66	0.65
	FUSION	(0.70 – 0.74)	(0.70 – 0.74)	**(0.72 – 0.76)**	(0.70 – 0.74)	(0.64 – 0.68)	(0.57 – 0.62)	(0.61 – 0.66)	(0.63 – 0.68)	(0.63 – 0.67)
		0.7	0.71	0.68	**0.73**	0.7	0.62	0.63	0.63	0.59
	KORA	(0.68 – 0.73)	(0.68 – 0.73)	(0.65 – 0.71)	**(0.70 – 0.75)**	(0.67 – 0.73)	(0.59 – 0.65)	(0.60 – 0.66)	(0.60 – 0.66)	(0.56 – 0.62)
		0.89	0.87	0.79	0.87	**0.91**	0.81	0.82	0.82	0.73
	WTCCC	(0.88 – 0.90)	(0.86 – 0.88)	(0.77 – 0.80)	(0.85 – 0.88)	**(0.91 – 0.92)**	(0.80 – 0.83)	(0.80 – 0.83)	(0.80 – 0.83)	(0.72 – 0.75)
		0.69	0.68	0.61	0.68	0.71	**0.77**	0.64	0.68	0.59
	CROAS	(0.64 – 0.74)	(0.63 – 0.73)	(0.55 – 0.66)	(0.63 – 0.73)	(0.65 – 0.76)	**(0.72 – 0.81)**	(0.58 – 0.69)	(0.63 – 0.73)	(0.53 – 0.64)
		0.79	0.8	0.73	0.78	0.81	0.72	**0.89**	0.78	0.65
	ERF	(0 .73 – 0.85)	(0.75 – 0.85)	(0.66 – 0.79)	(0.72 – 0.84)	(0.76 – 0.86)	(0.65 – 0.79)	**(0.84 – 0.93)**	(0.72 – 0.84)	(0.58 – 0.71)
		0.85	0.84	0.77	0.83	0.85	0.82	0.8	**0.9**	0.72
	MICROS	(0.78 – 0.91)	(0.77 – 0.90)	(0.69 – 0.84)	(0.77 – 0.89)	(0.79 – 0.91)	(0.75 – 0.89)	(0.73 – 0.87)	**(0.85 – 0.95)**	(0.64 – 0.79)
		0.8	0.81	0.79	0.8	0.79	0.68	0.69	0.75	**0.91**
	ORCADES	(0.71 – 0.88)	(0.72 – 0.90)	(0.70 – 0.88)	(0.72 – 0.89)	(0.70 – 0.87)	(0.59 – 0.77)	(0.58 – 0.80)	(0.66 – 0.85)	**(0.86 – 0.96)**

Supplementary Table 7. Hosmer-Lemeshow chi-square tests (and associated P values) for the clinical, genetic and combined type 2 diabetes risk prediction models.

Clinical model										
		Developed in								
Tested in		**Rotterdam**	**DGDG**	**FUSION**	**KORA**	**WTCCC**	**CROATIA-Vis**	**ERF**	**MICROS**	**ORCADES**
	Rotterdam	**17.66 (0.02)**	21.54 (0.01)	12.93 (0.11)	21.44 (0.01)	37.23 (<0.001)	24.30 (0.002)	26.80 (0.001)	23.84 (0.002)	10.64 (0.22)
	DGDG	38.57 (<0.001)	**50.00 (<0.001)**	51.99 (<0.001)	49.19 (<0.001)	68.67 (<0.001)	78.63 (<0.001)	42.31 (<0.001)	37.05 (<0.001)	32.45 (<0.001)
	FUSION	8.71 (0.37)	3.48 (0.90)	**6.60 (0.58)**	5.52 (0.70)	6.14 (0.63)	24.67 (0.002)	7.59 (0.48)	4.44 (0.82)	5.98 (0.65)
	KORA	17.03 (0.03)	11.23 (0.19)	9.75 (0.28)	**14.01 (0.08)**	14.31 (0.07)	13.63 (0.09)	10.43 (0.24)	5.78 (0.67)	7.70 (0.46)
	WTCCC	50.07 (<0.001)	46.22 (<0.001)	7.43 (0.49)	46.63 (<0.001)	**15.95 (0.04)**	18.58 (0.02)	17.14 (0.03)	9.73 (0.28)	37.43 (<0.001)
	CROAS	4.84 (0.77)	5.63 (0.69)	2.18 (0.98)	6.79 (0.56)	7.64 (0.47)	**2.54 (0.96)**	3.48 (0.90)	1.45 (0.99)	8.01 (0.43)
	ERF	8.83 (0.36)	9.13 (0.33)	8.37 (0.40)	9.11 (0.33)	7.91 (0.44)	14.41 (0.07)	**3.54 (0.90)**	11.61 (0.17)	11.92 (0.16)
	MICROS	2.95 (0.94)	5.31 (0.72)	4.37 (0.82)	5.85 (0.66)	2.20 (0.97)	6.77 (0.56)	6.60 (0.58)	**7.71 (0.46)**	3.00 (0.93)
	ORCADES	12.81 (0.12)	10.73 (0.22)	11.70 (0.17)	12.71 (0.12)	7.89 (0.44)	11.22 (0.19)	14.94 (0.06)	12.02 (0.15)	**4.84 (0.78)**
Genetic model										
		Developed in								
Tested in		Rotterdam	DGDG	FUSION	KORA	WTCCC	CROATIA-Vis	ERF	MICROS	ORCADES
	Rotterdam	**6.23 (0.62)**	11.06 (0.20)	9.04 (0.34)	7.46 (0.49)	4.04 (0.85)	8.83 (0.36)	14.19 (0.08)	12.37 (0.14)	11.85 (0.16)
	DGDG	8.64 (0.37)	**10.07 (0.26)**	4.00 (0.86)	2.78 (0.95)	9.15 (0.33)	10.13 (0.26)	3.79 (0.88)	5.03 (0.75)	3.99 (0.86)
	FUSION	6.07 (0.64)	9.10 (0.33)	**6.74 (0.57)**	14.50 (0.07)	12.30 (0.14)	3.74 (0.88)	6.07 (0.64)	9.51 (0.30)	3.85 (0.87)
	KORA	5.64 (0.69)	7.52 (0.48)	5.49 (0.70)	**7.27 (0.51)**	7.48 (0.49)	8.84 (0.36)	13.43 (0.10)	8.87 (0.35)	16.01 (0.04)
	WTCCC	18.06 (0.02)	18.85 (0.02)	19.29 (0.01)	34.78 (<0.001)	**26.96 (0.001)**	113.37 (<0.001)	169.45 (<0.001)	137.62 (<0.001)	182.66 (<0.001)
	CROAS	7.06 (0.53)	8.70 (0.37)	4.58 (0.80)	6.85 (0.55)	9.27 (0.32)	**10.33 (0.24)**	10.85 (0.21)	8.90 (0.35)	0.42 (0.99)
	ERF	7.20 (0.52)	7.89 (0.44)	2.15 (0.98)	2.80 (0.95)	3.95 (0.86)	14.68 (0.07)	**8.97 (0.35)**	11.64 (0.17)	9.62 (0.29)
	MICROS	7.29 (0.51)	7.36 (0.50)	3.42 (0.91)	5.80 (0.67)	11.07 (0.20)	7.12 (0.52)	5.57 (0.70)	**8.55 (0.38)**	5.37 (0.72)
	ORCADES	10.04 (0.26)	10.05 (0.26)	13.88 (0.09)	5.77 (0.67)	6.06 (0.64)	8.12 (0.42)	7.38 (0.50)	8.57 (0.38)	**5.33 (0.72)**

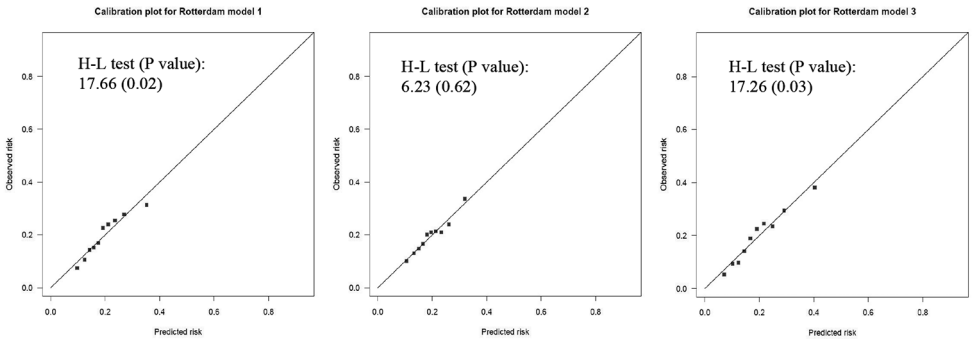
Combined clinical and genetic model

		Developed in								
Tested in		Rotterdam	DGDG	FUSION	KORA	WTCCC	CROATIA-Vis	ERF	MICROS	ORCADES
	Rotterdam	**17.26 (0.03)**	9.74 (0.28)	3.47 (0.90)	11.64 (0.17)	29.79 (0.002)	7.00 (0.54)	13.65 (0.09)	6.08 (0.64)	23.95 (0.002)
	DGDG	17.61 (0.02)	**17.99 (0.02)**	20.97 (0.01)	19.24 (0.01)	41.98 (<0.001)	14.62 (0.07)	9.78 (0.28)	10.39 (0.24)	5.53 (0.70)
	FUSION	7.28 (0.51)	11.11 (0.20)	**10.72 (0.22)**	8.11 (0.42)	9.65 (0.29)	11.95 (0.15)	3.86 (0.87)	3.21 (0.92)	13.88 (0.09)
	KORA	9.57 (0.30)	10.33 (0.24)	6.00 (0.65)	**9.65 (0.29)**	24.07 (0.002)	2.09 (0.98)	7.14 (0.52)	3.67 (0.89)	13.27 (0.10)
	WTCCC	12.09 (0.15)	35.52 (<0.001)	4.19 (0.84)	24.48 (0.002)	**11.33 (0.18)**	10.39 (0.24)	6.35 (0.61)	9.02 (0.34)	7.98 (0.44)
	CROAS	8.02 (0.43)	3.51 (0.90)	4.59 (0.80)	4.79 (0.78)	2.84 (0.94)	**10.08 (0.26)**	5.44 (0.71)	8.84 (0.36)	8.42 (0.40)
	ERF	5.91 (0.66)	7.84 (0.45)	8.05 (0.43)	5.96 (0.65)	7.18 (0.52)	3.82 (0.87)	**2.57 (0.96)**	5.70 (0.68)	9.62 (0.29)
	MICROS	6.78 (0.56)	10.62 (0.22)	7.17 (0.52)	4.94 (0.76)	5.95 (0.65)	6.10 (0.64)	6.86 (0.55)	**11.10 (0.20)**	6.75 (0.56)
	ORCADES	7.81 (0.45)	7.72 (0.46)	13.92 (0.08)	6.45 (0.60)	10.77 (0.22)	8.55 (0.38)	5.21 (0.74)	6.50 (0.59)	**8.39 (0.40)**

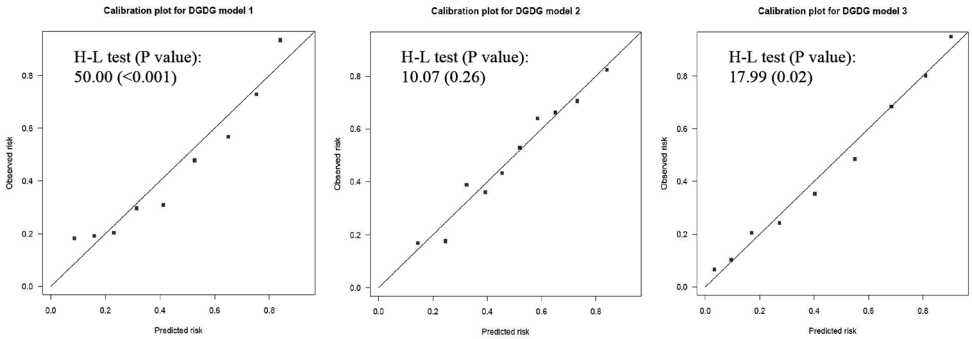
Clinical model: age, sex and body mass index; Genetic model: 37 genetic variants; Combined clinical and genetic model: age, sex, body mass index and 37 genetic variants.

Supplementary Figure 1. Calibration plots at derivation for the clinical (model 1), genetic (model 2) and combined (model 3) models.

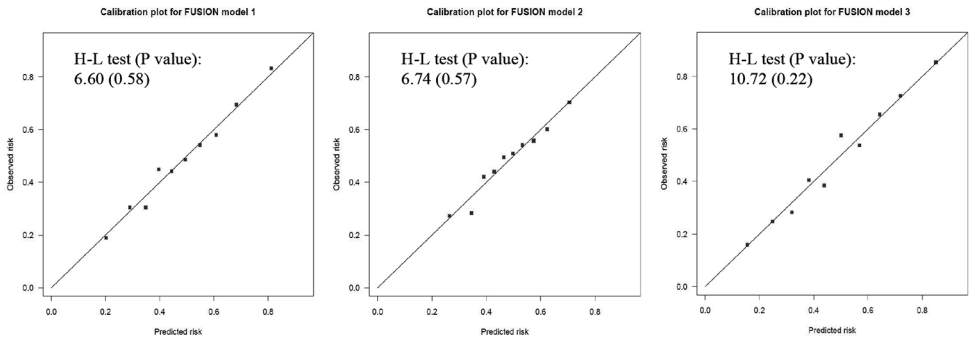
1a. Rotterdam



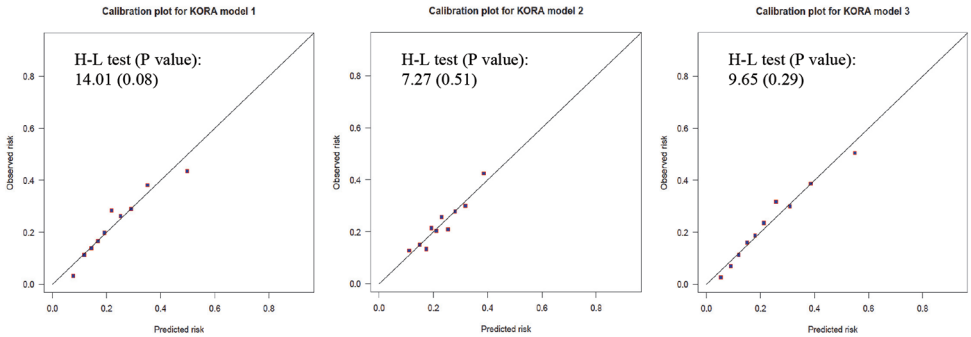
1b. DGDG



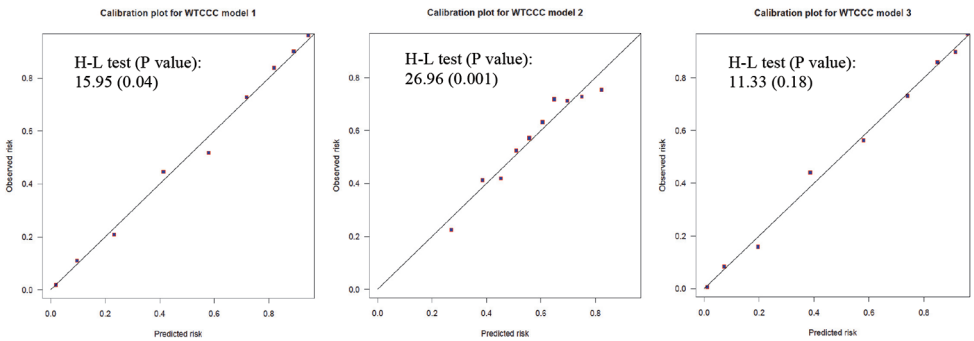
1c. FUSION



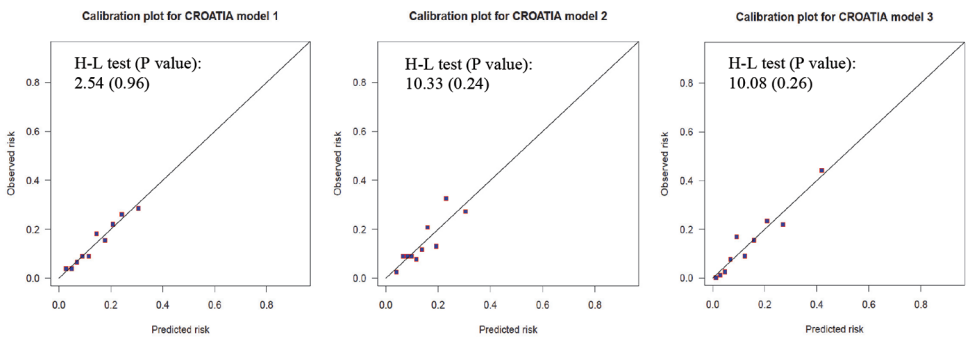
1d. KORA



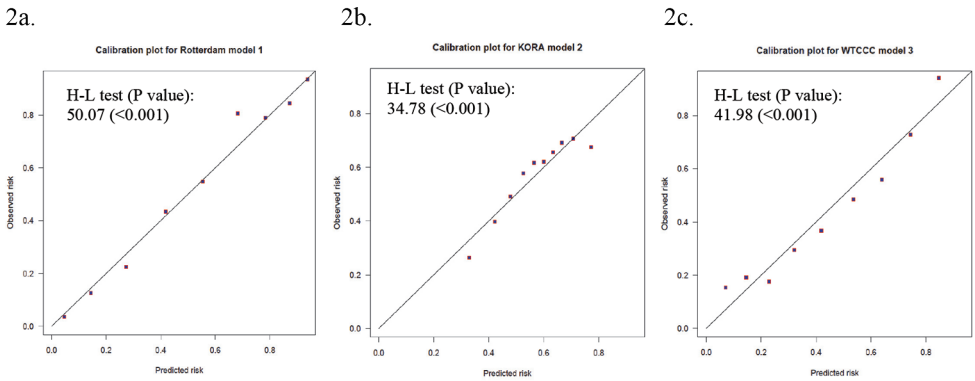
1e. WTCCC



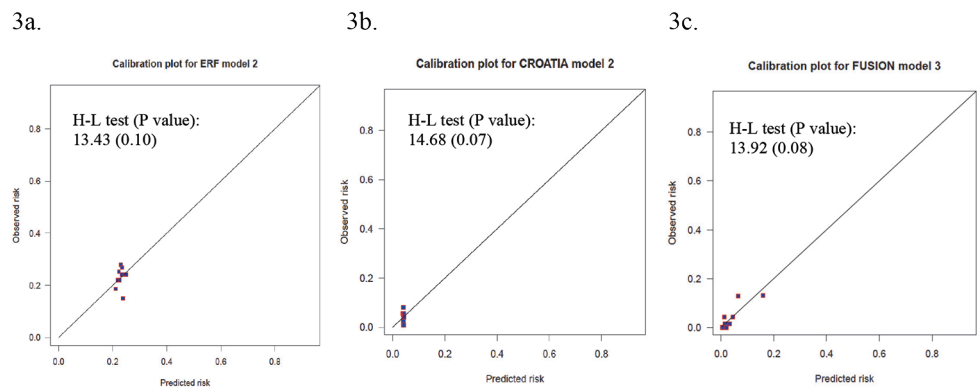
1f. CROATIA-Vis



Supplementary Figure 2. Examples of calibration plots at validation when the Hosmer-Lemeshow test indicated large miscalibration not visible on inspection of calibration plots. 2a. Clinical model developed in Rotterdam and tested in WTCCC; 2b. Genetic model developed in KORA and tested in WTCCC; 2c. Combined model developed in WTCCC and tested in DGDG.



Supplementary Figure 3. Examples of calibration plots at validation when the Hosmer-Lemeshow test indicated good calibration but inspection of calibration plots showed miscalibration. 3a. Genetic model developed in ERF and tested in KORA; 3b. Genetic model developed in CROATIA-Vis and tested in ERF; 3c. Combined model developed in FUSION and tested in ORCADES.



Chapter 9

Variations in predicted risks in personal genome testing for common complex diseases

Rachel R.J. Kalf, Raluca Mihaescu, Suman Kundu, Peter de Knijff,
Robert C. Green, A. Cecile J. W. Janssens

Genetics in Medicine. 2013;doi:10.1038/gim.2013.80

ABSTRACT

Purpose: The promise of personalized genomics for common complex diseases depends, in part, on the ability to predict genetic risks on the basis of single nucleotide polymorphisms. We examined and compared the methods of three companies (23andMe, deCODEme, and Navigenics) that have offered direct-to-consumer personal genome testing.

Methods: We simulated genotype data for 100,000 individuals on the basis of published genotype frequencies and predicted disease risks using the methods of the companies. Predictive ability for six diseases was assessed by the AUC.

Results: AUC values differed among the diseases and among the companies. The highest values of the AUC were observed for age-related macular degeneration, celiac disease, and Crohn disease. The largest difference among the companies was found for celiac disease: the AUC was 0.73 for 23andMe and 0.82 for deCODEme. Predicted risks differed substantially among the companies as a result of differences in the sets of single nucleotide polymorphisms selected and the average population risks selected by the companies, and in the formulas used for the calculation of risks.

Conclusion: Future efforts to design predictive models for the genomics of common complex diseases may benefit from understanding the strengths and limitations of the predictive algorithms designed by these early companies.

INTRODUCTION

It is envisioned that genome testing will personalize medicine, not only for the diagnosis and treatment of monogenic or Mendelian disorders, but also for the prevention of common complex diseases such as type 2 diabetes, age-related macular degeneration, and heart attack. Since 2007, personal genome tests have been offered directly to consumers via the Internet to educate and empower consumers about the risk of common diseases (1-4).

Common complex diseases are caused by an interplay between multiple genetic and non-genetic factors (5). Genomewide association studies are rapidly discovering variants implicated in common diseases but to date still leave a large part of the heritability unexplained because the identified single nucleotide polymorphisms (SNPs) generally have minor effects on disease risk (6). Consequently, genetic risk models based on known SNPs typically have a low to moderate predictive ability for most diseases. Exceptions do occur when one or more variants have a strong effect on disease risk, as in age-related macular degeneration and type 1 diabetes (5, 7, 8).

The predictive ability of direct-to-consumer personal genome tests has not been demonstrated in empirical studies. Insights concerning the concordance of personal genome tests conducted by different companies are available from a few reports of individuals who had sent their saliva to more than one company (9, 10). These reports showed that predicted risks differed among companies and were divergent for some traits in some individuals (9, 10). Differences in predicted risks were attributed to variations in the selection of the SNPs used, their effect sizes, and the average population risks of disease that were used to calculate disease risks (9, 11-13). As genotyping and sequencing become less expensive, they will be entering the medical mainstream. The methods used for estimating the predictive ability of common variants to generate risk information will be an important concern. In anticipation of this, we conducted an in-depth analysis and comparison of the approaches of the companies that pioneered the predictive use of genotyping in order to better understand the strengths and limitations of the methods they used to compute estimates.

We assessed and compared predicted risks and the predictive ability of personal genome testing offered by three companies: 23andMe, deCODEme, and Navigenics. The study was conducted in a hypothetical population of 100,000 individuals. Predicted risks were calculated using the methods of the companies, which were obtained from their websites. The predictive ability of the genetic risk models was quantified by the area under the receiver operating characteristic curve (AUC).

MATERIALS AND METHODS

Predicted risks and the predictive ability of personal genome tests from 23andMe, deCODEme, and Navigenics were assessed for six diseases: age-related macular degeneration, atrial fibrillation, celiac disease, Crohn disease, prostate cancer, and type 2 diabetes, which for all companies constitute a subset of all diseases tested. These diseases were chosen because of differences in the effect sizes of the SNPs discovered to date and differences in average population risks. Age-related macular degeneration and celiac disease are influenced by a few SNPs with strong effects on disease risk, whereas the other diseases are influenced by many SNPs with relatively weak effects. Celiac disease and Crohn disease are rare disorders, whereas the others are more common.

Because there are no prospective empirical data on the predictive ability of personal genome tests, we used hypothetical data to answer our research questions. A detailed description of the construction of the data sets, the calculation of predicted risks, and our efforts to verify correct interpretation of the risk calculation methods is provided in the **Supplementary Materials and Methods.**

METHODS

Simulated data

Construction of genotype data. Simulated data sets were constructed using a modeling procedure that has been validated and described in more detail elsewhere (14, 15) In short, this procedure creates genotypes for a hypothetical population of 100,000 individuals. For each SNP, genotypes are assigned randomly to individuals in such a way that genotype or allele frequencies in the 100,000 individuals match prespecified input values (see **Supplementary Materials and Methods**).

Calculation of predicted risks. Predicted risks were calculated using the methods of 23andMe, deCODEme, and Navigenics, which were described on their websites or in downloadable white papers (16-18) To calculate disease risks, all three methods require information on the average "population risk" and on the odds ratios and genotype or allele frequencies of the SNPs included in the test. The average population risks and the SNPs were obtained from the websites of the companies, and the odds ratios of the SNPs were extracted from the scientific studies referenced on the websites (accessed January 2012) (1-3). Genotype and allele frequencies were obtained from HapMap release 24 for 23andMe, cited scientific studies for deCODEme, and the company's website for Navigenics. The companies first compute the likelihood ratio or relative risk for each SNP using the odds ratio and genotype or allele frequencies. To generate predicted

risks, these likelihood ratios or relative risks are combined with the average population risk (see **Supplementary Materials and Methods**). All risks were calculated for Caucasian men.

Data analysis

To compare predicted risks among the three companies, we constructed one large data set with genotypes for the 113 SNPs tested by the three companies for all six diseases on the basis of genotype frequencies from HapMap release 28. For each individual, predicted risks were obtained using the formulas of the three companies, which yielded 18 predicted risks (6 diseases × 3 companies) per person.

To assess and compare the predictive ability, we used the genotype frequencies that the companies each used for the calculation of the likelihood ratios or relative risks (see above). Hence, we constructed hypothetical populations for each company and each disease separately. The predictive ability was quantified by the AUC (19). The AUC values range from 0.5 (random prediction) to 1.0 (perfect prediction). The AUC represents the probability that a random individual who will develop the disease has a higher predicted risk than a random individual who will not develop the disease. For the calculation of the AUC, disease status was randomly assigned to individuals on the basis of their predicted risks, in such a way that for individuals with the same disease risk, the percentage of individuals who will develop the disease equals that risk when the subgroup of individuals with that risk would have been sufficiently large.[14] In other words, the simulation method assumes perfect calibration of the prediction models. To illustrate the predictive ability, we obtained the distribution of predicted risks for people who will develop the disease and those who will not across the three risk categories that 23andMe distinguishes in the presentation of disease risks on the personal webpages of their consumers. The thresholds for these categories of decreased, typical, and elevated risk are 20% below and above the average population risks (relative risks 0.83 and 1.2) (1).

Finally, we assessed the agreement between the companies in classifying each individual to the same risk category. We used the original large data set, constructed for the comparison of predicted risks among the companies, to assess the agreement in classification across the three risk categories that 23andMe distinguishes. All analyses were performed using R version 2.12.1 (20).

RESULTS

Table 1 shows that 23andMe, deCODEme, and Navigenics used similar average population risks for the prediction of disease risks, except for age-related macular degeneration and celiac disease. For celiac disease, deCODEme used an average population risk that was eightfold higher than that used by 23andMe and 16-fold higher than that used by Navigenics. The number of SNPs

that were used for the calculation of the risk varied substantially among the companies. For the calculation of type 2 diabetes risks, 23andMe used 11 SNPs, deCODEme 21, and Navigenics 18; and for prostate cancer the companies used 12, 26, and 9 SNPs, respectively. For four diseases, deCODEme used the most SNPs, and for all six the company used twice as many SNPs as 23andMe used. The **Supplementary Table S1** shows that most SNPs tested by 23andMe or Navigenics were tested by two or more companies but that deCODEme tested many SNPs that were not covered by the other companies.

Table 2 shows that for each disease the AUC of the tests differed among the companies. The largest difference was observed for celiac disease (0.73 for 23andMe and 0.82 for deCODEme). The AUC values were also substantially different among the diseases. The AUC values were around 0.80 for age-related macular degeneration, celiac disease, and Crohn disease, but only around 0.60 for atrial fibrillation, prostate cancer, and type 2 diabetes. **Table 3** illustrates the predictive ability using the risk categories defined by 23andMe. When the AUC values are higher, individuals who will develop the disease more often have elevated risks and individuals who will not develop the disease more often have decreased risks of disease. When the AUC values are closer to 0.50, the distribution of predicted risks across the risk categories is more similar, which reflects that the risk model does not discriminate between the two groups.

Table 1. Average population risks and number of single nucleotide polymorphisms used by 23andMe, deCODEme and Navigenics in the prediction of risks for six multifactorial diseases.

	Average population risk (%)			Number of SNPs		
Diseases	**23andMe**	**deCODEme**	**Navigenics**	**23andMe**	**deCODEme**	**Navigenics**
Age-related macular degeneration	6.5	8	3.1	3	6	6
Atrial fibrillation	27.2	25	26	2	6	2
Celiac disease	0.12	1	0.06	4	8	10
Crohn disease	0.53	0.5	0.58	12	30	27
Prostate cancer	17.8	16	17	12	26	9
Type 2 diabetes	25.7	25	25	11	21	18

SNPs, single nucleotide polymorphisms.

Table 2. Area under the receiver operating characteristic curve for the prediction of six multifactorial diseases by 23andMe, deCODEme and Navigenics.

Diseases	23andMe	deCODEme	Navigenics
Age-related macular degeneration	0.76	0.81	0.82
Atrial fibrillation	0.58	0.62	0.58
Celiac disease	0.73	0.82	0.80
Crohn disease	0.76	0.80	0.77[a]
Prostate cancer	0.61	0.68	0.60
Type 2 diabetes	0.59	0.64	0.63[a]

[a]Calculated using an approximation of the formula described by Navigenics in their white paper (see **Supplementary Materials and Methods**).

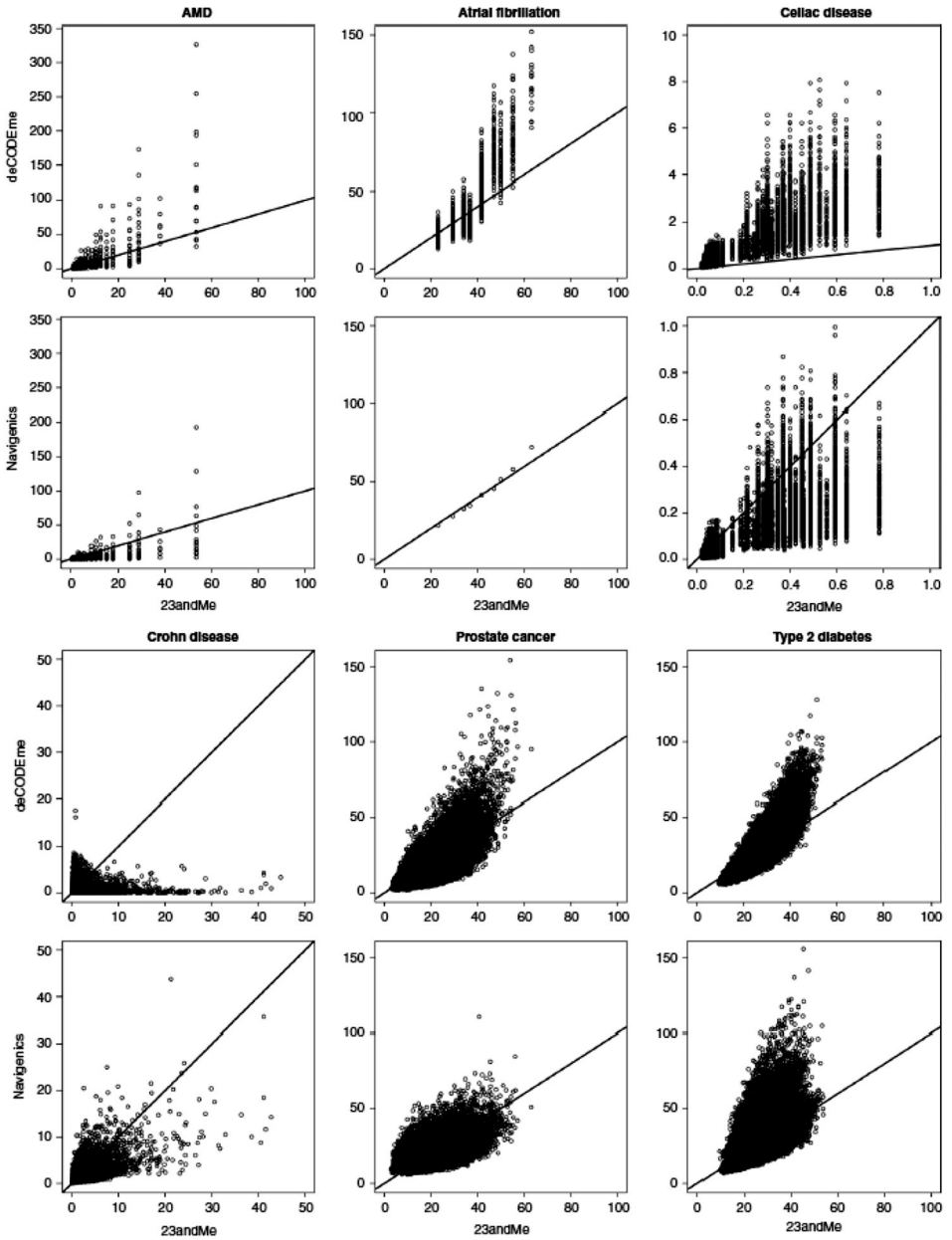
Table 3. Illustration of predictive ability using risk categories that 23andMe uses to classify disease risks.

		23andMe			deCODEme			Navigenics		
Diseases		↓	-	↑	↓	-	↑	↓	-	↑
Age-related macular	Patients	26.7	1.7	71.2	23.0	13.0	64.0	18.2	11.6	70.2
degeneration	Nonpatients	66.7	2.1	31.2	69.2	13.2	17.6	66.5	10.5	23.0
Atrial fibrillation	Patients	53.1	14.4	32.5	34.1	33.1	32.9	52.6	14.2	33.2
	Nonpatients	66.3	12.5	21.2	49.5	33.4	17.1	65.9	13.0	21.2
Celiac disease	Patients	42.1	0.0	57.9	22.9	2.5	74.6	24.6	4.9	70.5
	Nonpatients	80.6	1.1	18.3	75.4	2.8	21.8	75.7	5.8	18.4
Crohn disease	Patients	31.2	10.1	58.7	39.7	10.9	49.4	26.7[a]	15.1[a]	58.3[a]
	Nonpatients	69.7	10.9	19.4	83.8	5.5	10.7	65.6[a]	13.0[a]	21.5[a]
Prostate cancer	Patients	17.5	33.3	49.3	23.3	22.1	48.6	25.8	38.4	35.9
	Nonpatients	29.7	36.3	34.0	54.1	21.7	24.22	38.4	38.8	22.9
Type 2 diabetes	Patients	23.0	43.5	33.5	29.3	26.9	43.8	31.3[a]	31.2[a]	37.5[a]
	Nonpatients	33.5	43.7	22.8	48.4	27.2	24.5	48.4[a]	31.5[a]	20.1[a]

23andMe categorizes disease risks as decreased (↓), elevated (↑) and typical (-) risks if the risks of disease are lower than 20% below the average population risk, higher than 20% above the average population risk, and in between, respectively. Values are percentages.

[a]Calculated using an approximation of the formula described by Navigenics in their white paper (see **Supplementary Materials and Methods**).

Figure 1. Predicted risks by 23andMe, deCODEme and Navigenics for six multifactorial diseases.



The figure shows the predicted risks for a hypothetical population of 100,000 individuals (see Materials and Methods section). The solid line indicates when predicted risks by deCODEme or Navigenics are the same as predicted risks by 23andMe. Note that the ranges of the axes differ among the companies. AMD, age-related macular degeneration.

Table 4. Agreement between the three companies in assigning individual consumers to the same risk category, according to the risk categories used by 23andMe.

Diseases	Assigned to the same risk category by all three companies			Assigned to the same risk category by two companies			Assigned to different risk categories
	↓↓↓	---	↑↑↑	↑↑– ↓↓–	–-↑ –-↓	↑↑↓ ↓↓↑	↑–↓
Age-related macular degeneration	52.3	0.5	15.2	6.1	6.0	12.5	7.4
Atrial fibrillation	42.4	6.7	16.7	27.3	5.7	1.2	0.0
Celiac disease	75.3	0.0	13.8	9.0	0.4	1.3	0.3
Crohn disease	51.8	0.2	3.5	13.8	3.7	19.9	7.2
Prostate cancer	15.6	4.5	13.5	29.4	21.7	6.5	9.0
Type 2 diabetes	22.2	7.8	14.7	24.1	23.1	3.2	5.0

23andMe categorizes disease risks as decreased (↓), elevated (↑) and typical (-) risks if the risks of disease are lower than 20% below the average population risk, higher than 20% above the average population risk, and in between, respectively. Values are percentages. For example, ↓↓↓ indicates the percentage of individuals that were at decreased risk according to all three companies, and ↑–↓ indicates the percentage of individuals for which the three companies predicted risks in three different risk categories.

Figure 1 and **Supplementary Figure S1** show comparisons of predicted risks from the three companies for individual consumers. The strongest agreement in predicted risks was observed for atrial fibrillation, for which 23andMe and Navigenics predicted similar risks based on the same SNPs (see **Supplementary Table S1**), but many consumers received substantially different risk assessments from the companies for other diseases. For example, for Crohn disease, 23andMe used variants that had higher effect sizes than those used by Navigenics and variants that were not covered by deCODEme (see **Supplementary Table S1**); and for celiac disease, deCODEme predicted higher risks than 23andMe due to the higher average population risk that was used in the calculation (**Table 1**).

Figure 1 also shows that both deCODEme and Navigenics used formulas that allowed predicted risks to be >100%. The highest risks in our hypothetical population, 327% by deCODEme and 193% by Navigenics, were predicted for age-related macular degeneration. We examined the extent to which differences in the formulas could explain the prediction of risks >100% by applying the three formulas to the input data (average population risk, odds ratios, and allele frequencies of the SNPs) of 23andMe (see **Supplementary Figure S2**). **Supplementary Figure S2** shows that, in the range of higher predicted risks, the formulas of deCODEme and

Navigenics produced higher risks than those of 23andMe and that these risks could exceed 100%, as was shown for atrial fibrillation and prostate cancer.

Finally, again using the risk categories defined by 23andMe (see Materials and Methods section), we investigated the extent to which the three companies assigned individuals to the same risk category (**Table 4**). The highest concordance was observed for celiac disease, for which 89.0% of the individuals were assigned to the same risk category (75.3% as decreased risk and 13.8% as elevated risk), which is explained by the fact that all three companies test for the same variant that had a strong effect. For other diseases, concordance ranged from 33.6% (prostate cancer) to 68.0% (age-related macular degeneration). In most other instances, two companies assigned an individual to the same risk category and the third company predicted an average risk. Yet, for Crohn disease, age-related macular degeneration, and prostate cancer, 27.1%, 19.9%, and 15.5% of the individuals, respectively, were predicted opposing risks by at least two companies.

DISCUSSION

In 2008, 23andMe, deCODEme, and Navigenics, in collaboration with the Personalized Medicine Coalition, published a white paper in which they described the strategies they used for calculating genetic risks of disease (21). The companies explained and acknowledged that they use different SNPs, average population risks, and formulas to obtain predicted risks for consumers. Our analyses show that these differences in the SNPs, average population risks, and formulas yield substantial differences among the companies in the predictive ability for each disease and in predicted risks for individual consumers.

Before commenting on our results, three methodological issues may require further elaboration. First, we used simulated data to investigate the predictive ability of personal genome tests, because evidently empirical data were not available. On the basis of published genotype frequencies, we constructed genotype data for a hypothetical population of 100,000 individuals under the assumption that genetic variants inherit independently. Although this simulation method assumes perfect calibration of the risk models, which theoretically might lead to overestimation of AUC, we recently showed that this modeling approach was able to accurately replicate the AUC values of empirical prediction studies (15). We therefore believe it is reasonable to assume that the use of simulated data does not distort the results of this study. Second, we applied the risk categories utilized by 23andMe, which have relatively low thresholds to define risks as being decreased or increased. When individuals are easily classified in the very broad decreased or elevated risk categories, the agreement in assigning an individual to the same risk category, as presented in **Table 4**, is likely overestimated. And third, all companies in our

study provide regular updates of risk predictions to consumers when new SNPs are discovered or when better epidemiological data are available (22). We performed our analyses in January 2012 and verified all input data in December 2012. The most important change in that period was that Navigenics was acquired by Life Technologies and deCODEme by Amgen, and both no longer offer personal genome testing (2, 3). 23andMe had updated the prediction of age-related macular degeneration by the addition of two SNPs (1). Our results should therefore be interpreted as a historical comparison of direct-to-consumer personal genome testing and as an illustration of how differences in the sets of SNPs selected, the average population risks, and the formulas used for the calculation influence predicted risks and the predictive ability of personal genome tests.

The predictive ability of genetic tests as assessed by the AUC indicates the extent to which the test, at the "population" level, can discriminate between people who will develop the disease and those who will not. In contrast, a comparison of predicted risks indicates the extent to which "individual" consumers receive different predicted risks from the companies. Our study showed that the predictive ability differed among the companies for each of the diseases, and that differences in predicted risks were substantial even when tests had similar predictive ability. We also observed that, in exceptional cases, predicted risks of deCODEme and Navigenics could exceed 100%. We investigated three main factors that have an impact on predictive ability and predicted risks and that might explain these observations.

First, the companies included a different number of SNPs in their genetic risk models. For most diseases, the tests of deCODEme included the same SNPs as 23andMe and Navigenics, as well as additional SNPs that were not covered by the others. More SNPs generally implies more differentiation in predicted risks, as indicated by a higher AUC, and gives different risk predictions for individual consumers. For example, 23andMe and Navigenics predicted similar risks for atrial fibrillation (both AUC=0.58) because both considered the same two SNPs, whereas deCODEme considered four additional SNPs that introduced more variability in predicted risks and led to slightly higher predictive ability (AUC=0.62; **Table 2** and **Figure 1**). Note that tests with the same AUC do not necessarily predict the same risks at the individual level. Despite similar AUC values (0.61 and 0.60), 23andMe and Navigenics predicted markedly different risks for prostate cancer. In general, similar AUC values mean that the tests perform equally in identifying at-risk individuals at the population level, but individual consumers may be selected in the at-risk group on the basis of one test and not on the other when the predictive ability is not perfect, and the tests consider different risk factors. This may even occur when the AUC of the test is high, as was demonstrated for age-related macular degeneration, for which 20% of the consumers received risks in opposite risk categories (**Table 4**). Therefore, even tests

that have appreciable predictive ability at the population level may have contradictory results for individuals.

Second, all three companies used an estimate for the population disease risk as the starting point for their predictions. Some of these averages were relatively similar, but others were markedly different. For age-related macular degeneration, average risks were up to 2.5-fold higher, and for celiac disease up to 16-fold higher among the companies. Differences in average risks do not affect the predictive ability of the test, because they increase or decrease risks of the entire population to the same extent, but they do have an impact on actual values of predicted risks. This was most clearly demonstrated for celiac disease, for which almost all predicted risks by deCODEme were higher than those predicted by 23andMe and Navigenics, because their average population risk was up to 16-fold higher. The companies have likely used different epidemiological studies to obtain their estimates, but it is unlikely that differences in study population and design can explain the large differences in the average population that are used. It is more likely that some are prevalence and others are incidence estimates, or that the estimates are obtained from studies with different follow-up times, yielding different proxies for the lifetime risk. These inferences raise the question of whether the companies are calculating risks on the basis of information that is relevant to their consumers. Most genome-wide association studies are conducted in Caucasian populations, and the odds ratios from these studies may not be relevant for other ethnicities. Also, the companies used average estimates of lifetime risks and did not take age into account for the calculation of risk, but the remaining lifetime risks are not the same for 20- and 60-year-olds. And consumers might be more interested in short-term, e.g., 10-year, risks than lifetime risks, because these better reflect the risk of becoming ill at younger ages. A more in-depth reflection is needed on what risks are most appropriate to return in personal genome testing.

And third, the companies applied different formulas, which affected the exact prediction of risks. A difference among the formulas is that deCODEme multiplied the likelihood ratio of a genotype combination (genetic profile) by the average risk, Navigenics multiplied the relative risk by the lowest possible risk, where 23andMe multiplied the likelihood ratio by the average odds (16-18). These approaches yield similar predictions for lower risks, but the formulas of deCODEme and Navigenics appear to overestimate risks when predicted risks are higher. This difference in the calculation also results in scenarios in which predicted risks might become >100% for deCODEme and Navigenics (**Figure 1**), an observation that was previously made in a study on breast cancer risk (23). The strategy of 23andMe follows the widely accepted Bayes' theorem, which is in line with logistic regression and which prevents the resulting risks from exceeding 100%. DeCODEme multiplied likelihood ratios by the average risk, which is only

appropriate when risks are small (see **Supplementary Figure S2**). Finally, Navigenics multiplied relative risks by the lowest possible risk, a method that becomes computational infeasible on a standard computer for risk models that involve more than 14 SNPs. The question of which method is the most appropriate is difficult to answer, because it is unknown which model best reflects the underlying biological pathways to disease (24). Choosing the most appropriate computational method may improve calibration of risks, and potentially the predictive ability, but this improvement is likely minimal as compared with the improvement that could have been achieved if non-genetic risk factors were considered in the prediction of disease.

The differences in the selected SNPs, average disease risks, and formulas have different impacts on the predicted risks and the AUC values. They all determine the exact values of the predicted risks, but only the selected SNPs have an impact on AUC values. In general, the more SNPs included in the risk model and the higher their odds ratios and genotype frequencies, the higher the value of the AUC. Differences in average risks and in the formulas do not affect the AUC values because AUC is essentially a rank test, and these differences do not change the rank order of the predicted risks. The differences in allele frequencies and odds ratios, given that the companies used different sources to obtain this information, would seem to be a possible explanation for the observed differences in the AUC values. Yet AUC is known to be relatively insensitive and unable to detect minor improvements of risk models (25). The differences in odds ratios and allele frequencies were likely too minor to cause variation in the AUCs. The differences in the AUCs among the companies are predominantly explained by the selection of the SNPs.

In the absence of prospective empirical data, our study provided insight into the methodology and performance of risk estimation for personal genome tests. We showed that the predictive ability of personal genome tests and the predicted risks for individual consumers differed among the companies due to the differences in the SNPs selected, the average population risks, and the formulas. For six diseases, we showed that the personal genome tests of the three companies had limited predictive ability (atrial fibrillation, type 2 diabetes, and prostate cancer), a considerable (20-27%) probability of receiving "opposite" predictions (age-related macular degeneration and Crohn disease), or substantial differences in absolute risks at the individual level (celiac disease). These observations on the variation and pitfalls in disease risk predictions by personal genome tests provide insights into models of risk estimation and will inform the evolving discussion about the best use of genomic information in the consumer marketplace and in the practice of medicine.

REFERENCES

1. 23andMe. https://www.23andme.com/. Accessed 1 January 2012 and 1 December 2012.
2. deCODEme. http://www.decodeme.com/. Accessed 1 January 2012 and 1 December 2012.
3. Navigenics. http://www.navigenics.com/. Accessed 1 January 2012 and 1 December 2012.
4. Helgason A, Stefánsson K. **The past, present, and future of direct-to-consumer genetic tests.** *Dialogues Clin Neurosci* 2010;**12**:61–68.
5. Janssens AC, van Duijn CM. **Genome-based prediction of common diseases: advances and prospects.** *Hum Mol Genet* 2008;**17**(R2):R166–R173.
6. Visscher PM, Brown MA, McCarthy MI, Yang J. **Five years of GWAS discovery.** *Am J Hum Genet* 2012;**90**:7–24.
7. Hecker LA, Edwards AO, Ryu E, et al. **Genetic control of the alternative pathway of complement in humans and age-related macular degeneration.** *Hum Mol Genet* 2010;**19**:209-215.
8. Eleftherohorinou H, Wright V, Hoggart C, et al. **Pathway analysis of GWAS provides new insights into genetic susceptibility to 3 inflammatory diseases.** *PLoS ONE* 2009;**4:**e8068.
9. United States Government Accountability Office. **Direct-to-consumer genetic tests: Misleading test results are further complicated by deceptive marketing and other questionable practices.** 2010. www.gao.gov/assets/130/125079.pdf. Accessed 1 January 2012.
10. Ng PC, Murray SS, Levy S, Venter JC. An agenda for personalized medicine. Nature 2009;461:724-726.
11. Imai K, Kricka LJ, Fortina P. **Concordance study of 3 direct-to-consumer genetic-testing services.** *Clin Chem* 2011;**57**:518–521.
12. Swan M. **Multigenic condition risk assessment in direct-to-consumer genomic services.** *Genet Med* 2010;**12**:279-288.
13. Kuehn BM. **Risks and benefits of direct-to-consumer genetic testing remain unclear.** *JAMA* 2008;**300**:1503-1505.
14. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM. **Predictive testing for complex diseases using multiple genes: fact or fiction?** *Genet Med* 2006;**8**:395-400.
15. Kundu S, Karssen LC, Janssens AC. **Analytical and simulation methods for estimating the potential predictive ability of genetic profiling: a comparison of methods and results.** *Eur J Hum Genet* 2012;**20**:1270-1274.
16. Macpherson M, Naughton B, Hsu A, Mountain J. **White paper 23-01: estimating genotype-specific incidence for one or several loci. 2007.** https://23andme. https.internapcdn.net/res/pdf/HIC-SXIYiYqXreldAxO5yA_23-01_Estimating_Genotype_Specific_Incidence.pdf. Accessed 1 July 2012.
17. deCODEme. **SNP selection and risk calculation. demo.decodeme.com/healthwatch-information/risk-calculation.** Accessed 1 July 2012.
18. Navigenics. **The science behind the Navigenics service.** www.navigenics.com/static/pdf/Navigenics-TheScience.pdf. Accessed 1 July 2012.
19. Hanley JA, McNeil BJ. **The meaning and use of the area under a receiver operating characteristic (ROC) curve.** *Radiology* 1982;**143**:29–36.
20. Ihaka GR, Gentelman R. R: **A language for data analysis and graphics.** *J Comput Graph Stat* 1996;**5**:299-314.

21. **The Personalized Medicine Coalition. Personal genomics and industry standards: scientific validity.** http://www.personalizedmedicinecoalition.org/sites/all/themes/zen_pmc/documents/PMC%20personalgenomicsSci%20Valid15dec08.pdf. Accessed 1 December 2012.

22. Mihaescu R, van Hoek M, Sijbrands EJ, et al. **Evaluation of risk prediction updates from commercial genome-wide scans.** *Genet Med* 2009;**11**:588-594.

23. So HC, Sham PC. **Effect size measures in genetic association studies and age-conditional risk prediction.** *Hum Hered* 2010;**70**:205-218.

24. Moonesinghe R, Khoury MJ, Liu T, Janssens AC. **Discriminative accuracy of genomic profiling comparing multiplicative and additive risk models.** *Eur J Hum Genet* 2011;**19**:180–185.

25. Cook NR. **Use and misuse of the receiver operating characteristic curve in risk prediction.** *Circulation* 2007;**115**:928-935.

SUPPLEMENTARY MATERIALS AND METHODS

The Supplementary methods describe in more detail how the datasets were constructed and which sources of input data (odds ratios and genotype frequencies) were used for each research question. We also report our efforts to verify the risk calculations.

Simulated data

Construction of genotype data

Simulated datasets were constructed using a modeling procedure that has been validated and described in more detail elsewhere (1, 2) In short, this procedure creates genotypes for a hypothetical population of 100 000 individuals. For each SNP, genotypes are assigned randomly to individuals in such a way that the genotype or allele frequencies match pre-specified input values. The input values for the frequencies were obtained from Hapmap, from the scientific publications cited on the websites of the companies or directly from those websites, depending on the research question. Which source was used for each research question is specified below. When input values were allele frequencies, Hardy-Weinberg equilibrium was assumed to obtain genotype frequencies.

Calculation of predicted risks

Predicted risks were calculated using the methods of 23andMe, deCODEme and Navigenics, which were described on their websites or in downloadable white papers (3-5). To calculate disease risks, all three methods require information on the average 'population risk' and on the odds ratios and genotype or allele frequencies of the SNPs included in the test. The average population risks and SNPs were obtained from the websites of the companies and the odds ratios of the SNPs were extracted from the scientific studies referenced on the websites (accessed January 2012) (6-8) For the calculation of the likelihood ratios and relative risks that are needed to compute predicted risks, genotype and allele frequencies were obtained from Hapmap release #24 for 23andMe, from the cited scientific studies for deCODEme and from the company's website for Navigenics. For three SNPs used by 23andMe genotype frequencies were not available in Hapmap and hence were calculated from the odds ratios and likelihood ratios in available reports (see below). All risks were calculated for Caucasian men.

Although the methods of the three companies require the same input parameters, the formulas used for the exact calculation of risks had notable differences. 23andMe and deCODEme transformed genotype odds ratios of single SNPs into likelihood ratios, representing the odds of disease for each genotype relative to the average odds. To compute predicted risks, 23andMe then

multiplied the likelihood ratios of single genotypes by the average odds of disease and converted the odds into risks, whereas deCODEme multiplied the likelihood ratios by the average risk of disease. Navigenics did not transform the individual genotype odds ratios into likelihood ratios but into relative risks, and calculated the relative risks for all possible genotype combinations. To compute individual predicted risks, they divided the relative risk of each genotype combination by the average *relative* risk before multiplying these by the average *absolute* risk of disease. Given that for each SNP the reference genotype for the calculation of the relative risk was the genotype with the lowest risk of disease, this strategy is in essence similar to multiplying all relative risks by the theoretically lowest possible risk. Because Navigenics calculates the relative risks for all possible genotype combinations, the method requires substantial computer working memory. On a standard computer the method could run out of memory when the number of SNPs is higher than 14, which was the case for Crohn's disease and type 2 diabetes. For these diseases, we obtained an approximation of the lowest possible risk by dividing the population disease risk by the average of the relative risks for each genotype combination in our population.

Verification of risk calculations

To verify whether we were applying the methods of the companies accurately, we first attempted to reproduce the risks presented in the sample reports that are available on their websites as well as to reproduce the risks predicted for two researchers who had their DNA tested by each of these companies (RG and PdK). For the six diseases under study, we were able to verify a total of 48 predicted risks; 18 for 23andMe and deCODEme and only 12 for Navigenics because we only had information from the sample report and the report of one researcher. We exactly reproduced 31 of the 48 predicted risks, and found absolute differences smaller than 1% for 12 risks and differences up to 5% for 5 risks. These differences were mainly explained by the fact that for a few SNPs we could not retrieve the exact same odds ratios used by the companies.

Apart from verifying whether we accurately could reproduce risks from the available reports, we also verified for individual SNPs whether we could reproduce the likelihood ratios (23andMe and deCODEme) and odds ratios (Navigenics) that were presented in the reports and used in the calculations. The 44 SNPs tested by 23andMe, 97 SNPs by deCODEme and 72 SNPs by Navigenics generated 132, 302 and 216 likelihood ratios and odds ratios for included SNPs and haplotypes (for deCODEme). For 80 (61%), 174 (58%) and 174 (81%), at least one in each SNP, we knew the exact likelihood ratios used by 23andMe and deCODEme and the exact odds ratios that Navigenics used to calculate relative risks from the available reports. More information was available for Navigenics, because the odds ratios of the homozygous genotypes were given for all SNPs in each report. Using odds ratios and genotype frequencies from the

literature, we could reproduce 85% (68/80), 94% (163/174) and 98% (170/174) of the likelihood ratios and odds ratios with an absolute difference of 0.01 or smaller. For 23andMe, we were able to reproduce only 85% with an absolute difference of 0.01 or smaller, because for several SNPs the genotype frequencies were not in Hapmap release #24 and because, more often than for the other companies, the odds ratios found in the literature did not exactly reproduce the likelihood ratios reported in the available reports. Yet, absolute differences were 0.05 or larger for only three likelihood ratios of 23andMe, one of deCODEme and one odds ratio of Navigenics. Because we knew at least one likelihood ratio or odds ratio for each SNP, we assumed we could also correctly obtain the remaining likelihood ratios and odds ratios. For 23andMe and deCODEme, we could reproduce an additional 9% (7/80) and 3% (5/174) of the likelihood ratios by using likelihood ratios from the available reports and odds ratios from the literature. These efforts suggest that we managed to reconstruct the prediction methods of the companies. The minor differences in likelihood ratios and odds ratios might have affected the exact calculation of the predicted risks, but they were unlikely large enough to influence the main findings of the study.

Data analysis

Comparison of predicted risks

To compare predicted risks among the 3 companies, we constructed one large dataset with genotypes for all 113 SNPs tested by the three companies for all six diseases. These 113 SNPs remained from the total of 213 SNPs (44 for 23andMe, 97 for deCODEme and 72 for Navigenics) after excluding duplicates (n=73) and SNPs in linkage disequilibrium (LD; $r^2 \geq 0.6$; n=27). If SNPs were in LD, we selected the SNP that was 1) reported in the cited scientific studies; 2) used by two of the three companies; or 3) the first or most published in the GWAS catalog (9). Genotype frequencies were obtained from Hapmap release #28, except for two SNPs that were not available in Hapmap we used the frequencies reported on the website of Navigenics.

Comparison of predictive ability

To assess and compare the predictive ability, we used the genotype frequencies that each of the companies used for the calculation of the likelihood ratios or relative risks (see above). Hence, we constructed hypothetical populations for each company separately. The predictive ability was quantified by the area under the receiver operating characteristic curve (AUC) (10). AUC values range from 0.5 (random prediction) to 1.0 (perfect prediction). AUC represents the probability that a random individual who will develop the disease has a higher predicted risk than a random individual who will not develop the disease. For the calculation of AUC, disease status was randomly assigned to individuals based on their predicted risks in such a way

that for individuals with the same disease risk, the percentage of individuals who will develop the disease equals that risk when the subgroup of individuals with that risk would have been sufficiently large (1).

Note that to assess the predictive ability we constructed genotype datasets for each company separately based on the genotype frequencies they had used for the calculation of their predicted risks, instead of using a single set of frequencies for all companies, like we did for the first research question. Constructing separate datasets for each company based on their own input values ensures that all risk models are perfectly calibrated. Had we used e.g. Hapmap #24 to construct one large genotype dataset, then the risk models of 23andMe were expected to be better calibrated than those of deCODEme and Navigenics as 23andMe used Hapmap #24 genotype frequencies in the calculation of the disease risks. In that case, differences in AUC would have been in part due to differences in calibration. While this approach yields a valid comparison of the genetic tests, the exact AUC values should be interpreted with caution as they are likely overestimated. For all three companies, external validation of the risk models in an independent unselected population will likely show lower AUC values and suggest poorer predictive ability than presented in this study.

To illustrate the predictive ability, we obtained the distribution of predicted risks for people who will develop the disease and those who will not across the three risk categories that 23andMe distinguishes in the presentation of disease risks on the personal webpages of their consumers. The thresholds for these categories of decreased, typical and elevated risk are 20% below and above the average population risks (relative risks 0.83 and 1.2) (6).

Comparison of risk categories

Finally, we assessed the agreement between the companies in classifying each individual to the same risk category. We used the original large dataset, constructed for the comparison of predicted risks between the companies, to assess the agreement in classification across the three risk categories that 23andMe distinguishes.

All analyses were performed using R version 2.12.1 (11).

SUPPLEMENTAL REFERENCES:

1. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM. P**redictive testing for complex diseases using multiple genes: fact or fiction?** *Genet Med.* 2006;**8**(7):395-400.
2. Kundu S, Karssen LC, Janssens AC. **Analytical and simulation methods for estimating the potential predictive ability of genetic profiling: a comparison of methods and results.** *Eur J Hum Genet.* 2012;**20**(12):1270-1274.
3. Macpherson M, Naughton B, Hsu A, Mountain J. White paper 23-01: Estimating genotype-specific incidence for one or several loci. 2007; https://23andme.https.internapcdn.net/res/pdf/HIC-SXIYiYqXreldAxO5yA_23-01_Estimating_Genotype_Specific_Incidence.pdf. Accessed 1July 2012.
4. deCODEme. **SNP selection and risk calculation.** http://demo.decodeme.com/health-watch-information/risk-calculation. Accessed 1 July 2012.
5. Navigenics. **The science behind the Navigenics service.** http://www.navigenics.com/static/pdf/Navigenics-TheScience.pdf. Accessed 1 July 2012.
6. 23andMe. https://www.23andme.com/. Accessed 1 January 2012.
7. deCODEme. http://www.decodeme.com/. Accessed 1 January 2012.
8. Navigenics. http://www.navigenics.com/. Accessed 1 January 2012.
9. Hindorff LA, MacArthur J, Morales J, et al. **A Catalog of Published Genome-Wide Association Studies.** www.genome.gov/gwastudies. Accessed 1 July 2012.
10. Hanley JA, McNeil BJ. **The meaning and use of the area under a receiver operating characteristic (ROC) curve.** *Radiology.* 1982;**143**(1):29-36.
11. Ihaka GR, Gentelman R. R: **a language for data analysis and graphics.** *J Comput Graph Stat.* 1996;**5**:299-314.

Table S1. Overview of the single nucleotide polymorphisms and odds ratios used by 23andMe, deCODEme and Navigenics.

SNP	**Locus**	**Proxy SNP**	**Odds ratio**		
			23andMe	**deCODEme**	**Navigenics**
Age-related macular degeneration					
rs10490924	*10q26.13*	rs3750847	3.47*	3.47*	2.72/10.57
rs1061170	*1q31.3*	rs1061147	2.85*	Haplotype	3.10/6.30
rs1410996	*1q31.3*			Haplotype	3.16
rs2230199	*19p13.3*			1.69	1.70/2.60
rs547154	*6p21.33*	rs522162	1.88	1.88	3.13*
rs9332739	*6p21.33*			2.14	2.78
Atrial fibrillation					
rs10033464	*4q25*		1.39	Haplotype	1.39
rs13376333	*1q21.3*			1.13	
rs2200733	*4q25*		1.72	Haplotype	1.72
rs3807989	*7q31.2*			1.09	
rs3825214	*12q24.21*			1.14	
rs7193343	*16q22.3*			1.21	
Celiac disease					
rs1464510	*3q28*	rs9851967	1.22*		1.21
rs1738074	*6q25.3*			1.21	1.21
rs17810546	*3q25.33*			1.34	1.34
rs2187668	*6p21.32*		7.04	7.04	7.04
rs231779	*2q33.2*				1.24
rs2816316	*1q31.2*			1.41	1.41
rs3184504	*12q24.12*			1.19	1.19
rs6441961	*3p21.31*		1.32	1.21	
rs6822844	*4q27*	rs6840978	1.41	1.59	1.43*
rs917997	*2q12.1*			1.27	1.27
rs9811792	*3q25.33*				1.21
Crohn's disease					
rs1000113	*5q33.1*	rs7714584	1.33*		1.54/1.92
rs10045431	*5q33.3*			1.11	
rs10758669	*9p24.1*			1.12	
rs10761659	*10q21.2*		1.23/1.55		1.23/1.55
rs10883365	*10q24.2*	rs11190140	1.18*	1.18*	1.20/1.62
rs11175593	*12q12*				1.54
rs11209026	*1p31.3*		2.92	Haplotype	

SNP	Locus	Proxy SNP	Odds ratio		
			23andMe	deCODEme	Navigenics
rs11805303	*1p31.3*	rs1004819* rs10889677**	1.56*	Haplotype**	1.39/1.86
rs11584383	*1q32.1*	rs12122721		1.18	1.18*
rs12521868	*5q31.1*			1.21	
rs1456893	*7p12.2*			1.20	1.20
rs1551398	*8q24.13*			1.08	1.08
rs17234657	*5p13.1*		1.16	1.16	1.54/2.32
rs1736135	*21q21.1*			1.18	1.18
rs17582416	*10p11.21*	rs4934724		1.16*	1.16
rs1793004	*11p15.1*			1.24	
rs2066843	*16q12.1*			1.37	
rs2066844	*16q12.1*		2.97		1.97/3.29
rs2066845	*16q12.1*		6.32		3.05/12.13
rs2066847	*16q12.1*		6.68		1.97/3.29
rs224136	*10q21.2*			1.67	
rs2241880	*2q37.1*	rs10210302	1.45/1.77	1.45	1.19/1.85*
rs2274910	1q23.3			1.14	
rs2301436	*6q27*			1.21	1.21
rs2476601	*1p13.2*	rs6679677		1.31*	1.31
rs2542151	*18p11.21*	rs1893217	1.15*	1.35	1.30/2.01
rs2872507	*17q12*			1.12	1.12
rs3764147	*13q14.11*			1.25	1.25
rs4263839	*9q32*			1.22	1.22
rs4958847	*5q33.1*			1.36	
rs6908425	*6p22.3*			1.21	
rs744166	*17q21.2*			1.18	1.18
rs762421	*21q22.3*			1.13	1.13
rs7746082	*6q21*			1.17	1.17
rs7927894	*11q13.5*				1.16
rs9286879	*1q24.3*			1.19	1.19
rs9858542	*3p21.31*	rs3197999	1.20*	1.17	1.09/1.84
Prostate cancer					
rs10086908	8q24.21			1.15	
rs10486567	*7p15.2*		1.19/1.37	1.12/1.19	1.23/1.33
rs10505483	*8q24.21*	rs16901979	1.44/2.17	1.79	1.79*
rs10896449	*11q13.3*			1.19/1.47	1.19/1.47
rs10934853	*3q2.3*			1.12	
rs10993994	*10q11.23*		1.23	1.24	

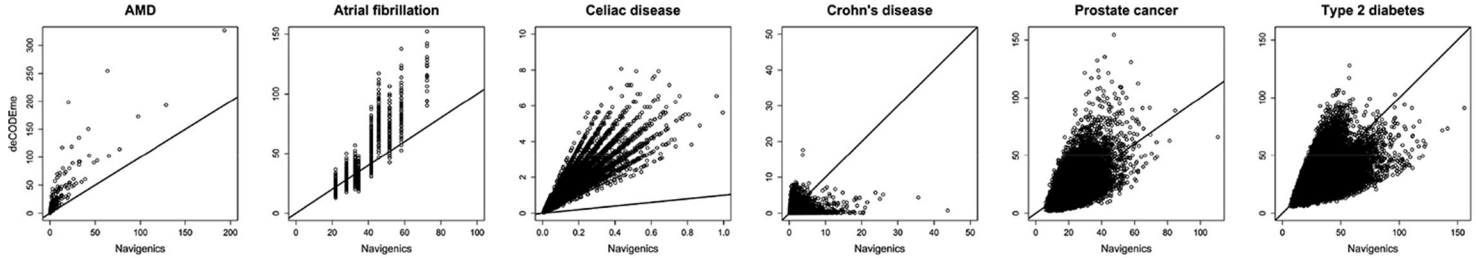
SNP	Locus	Proxy SNP	Odds ratio		
			23andMe	deCODEme	Navigenics
rs12621278	*2q31.1*	rs10207654	2.23	1.33*	
rs1447295	*8q24.21*		1.43/2.23	1.53	1.43/2.23
rs1512268	*8p21.2*		1.19/1.39	1.18	
rs16902104	*8q24.21*			1.21	
rs17021918	*4q22.3*		1.20	1.11	
rs1859962	*17q24.3*		1.20	1.20	1.20
rs2660753	*3p12.1*			1.08	
rs2735839	*19q13.33*			1.12	
rs401681	*5p15.33*			1.07	
rs4430796	*17q12*		1.17	1.22	1.24/1.48
rs4962416	*10q26.13*				1.16 /1.49
rs5759167	*22q13.2*			1.16	
rs5945572	Xp11.22			1.23	
rs620861	*8q24.21*			1.14	
rs6465657	*7q21.3*			1.12	
rs6983267	*8q24.21*		1.25	1.26/1.58	1.26/1.58
rs7127900	*11p15.5*		1.24/1.42	1.22	
rs721048	*2p15*	rs2710646		1.15*	1.15
rs7679673	*4q24*			1.10	
rs8102476	*19q13.2*		1.12	1.12	
rs9364554	*6q25.3*			1.14	
Type 2 diabetes					
rs10010131	*4p16.1*	rs10012946	1.15/1.23*	1.11	1.03/1.19
rs10244051	*7p21.2*			1.06	
rs10830963	*11q14.3*			1.09	
rs10923931	*1p12*	rs2793831		1.13*	1.13
rs1111875	*10q23.33*		1.13	1.17	1.11/1.21
rs13266634	*8q24.11*		1.12	1.14	1.18
rs1387153	*11q14.3*		1.09		
rs1801282	*3p25.2*		1.14	1.14	1.30/1.53
rs2237892	*11p15.4*	rs2283228	1.29	1.29	1.24*
rs2383208	*9p21.3*	rs10811661	1.19	1.20	1.20/1.44*
rs2877716	*3q21.1*			1.12	
rs340874	*1q32.3*			1.07	
rs4402960	*3q27.2*		1.14	1.14	1.14
rs4430796	*17q12*			1.10	1.08/1.19
rs4607103	*3p14.1*				1.09

SNP	**Locus**	**Proxy SNP**	**Odds ratio**		
			23andMe	**deCODEme**	**Navigenics**
rs4607517	*7p13*			1.07	
rs5215	*11p15.1*	rs5219	1.14*	1.15	1.11
rs7578597	*2p21*			1.15	
rs7756992	*6p22.3*	rs4712523	1.12*	1.20	1.15/1.50
rs780094	*2p23.3*			1.06	
rs7903146	*10q25.2*	rs4506565	1.37	1.37	1.36/1.88*
rs7961581	*12q21.1*			1.09	1.09
rs8050136	*16q12.2*				1.17
rs864745	*7p15.1*			1.10	1.10
rs9300039	11p12				1.80/2.61
rs9494266	*6q23.3*				2.31

Values are allelic or genotypic odds ratios (heterozygous/homozygous) for the single nucleotide polymorphisms (SNPs) that are used by the companies. SNPs that were in linkage disequilibrium ($r^2 \geq 0.6$; SNP annotation and proxy search of the Broad Institute) are mentioned on the same line as reference SNP and proxy SNP. We selected the reference SNP as the SNP that was reported in the cited scientific studies; used by two of the three companies; or the first or most published SNP in the GWAS catalog. Haplotype indicates that the SNP was used to construct a haplotype (odds ratios not reported). Locus names were obtained from Ensembl.

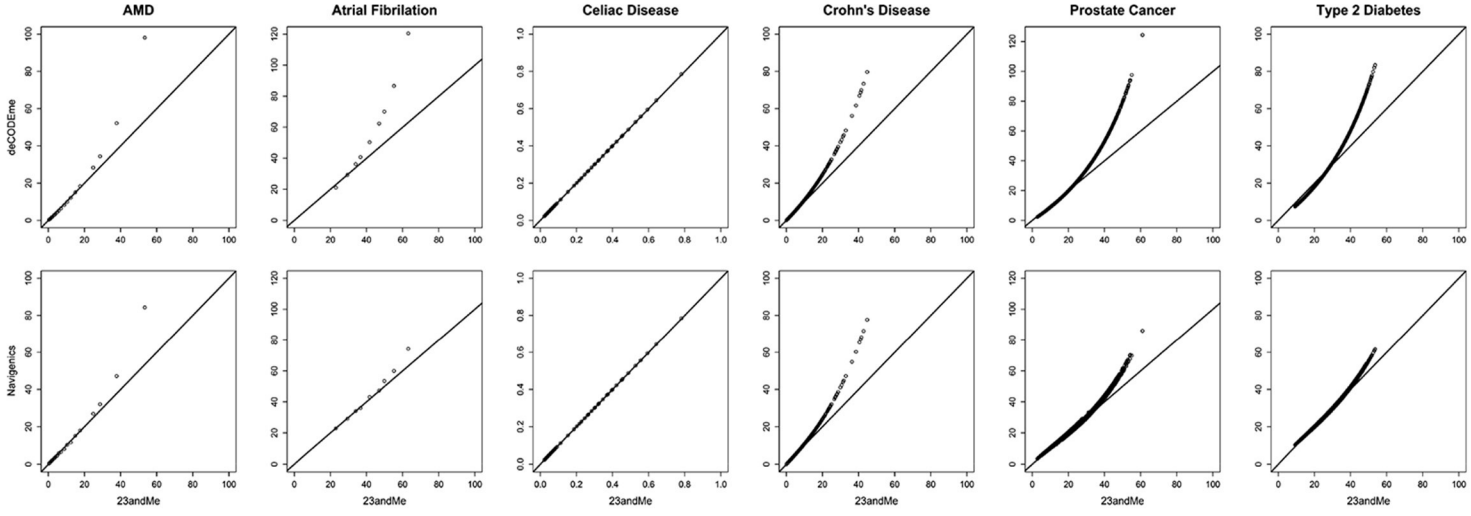
*/** Odds ratio of the proxy SNP.

Figure S1. Predicted risks by deCODEme and Navigenics for six multifactorial diseases.



Predicted risks for a hypothetical population of 100 000 individuals (see Methods). The solid line indicates when predicted risks by deCODEme are the same as predicted risks by Navigenics. Note that the ranges of the axes differ between the companies.

Figure S2. Predicted risks using the formulas by 23andMe, deCODEme and Navigenics when the same average population risks, odds ratios and allele frequencies were assumed.



Predicted risks for a hypothetical population of 100,000 individuals were calculated using the formulas of the companies applied to the average population risks, odds ratios, allele frequencies and number of SNPs used by 23andMe. The solid line indicates when predicted risks by respectively deCODEme and Navigenics were the same as predicted risks by 23andMe.

Chapter 10

A methodological perspective on genetic risk prediction studies in type 2 diabetes: recommendations for future research

Sara M. Willems, Raluca Mihaescu, Eric J. G. Sijbrands,
Cornelia M. van Duijn, A. Cecile J. W. Janssens

Current Diabetes Reports. 2011;11(6):511-8

ABSTRACT

Fueled by the successes of genome-wide association studies, numerous studies have investigated the predictive ability of genetic risk models in type 2 diabetes. In this paper, we review these studies from a methodological perspective, focusing on the variables included in the risk models as well as the study designs and populations investigated. We argue and show that differences in study design and characteristics of the study population have an impact on the observed predictive ability of risk models. This observation emphasizes that genetic risk prediction studies should be conducted in those populations in which the prediction models will ultimately be applied, if proven useful. Of all genetic risk prediction studies to date, only a few were conducted in populations that might be relevant for targeting preventive interventions.

INTRODUCTION

Type 2 diabetes (T2D) is a multifactorial disease, caused by a complex interplay between genetic and nongenetic risk factors. Compelling evidence has identified increasing age, higher body mass index (BMI), impaired fasting glucose, impaired glucose tolerance, higher glycated hemoglobin (HbA1c) level, and metabolic syndrome as important T2D risk factors (**Table 1**) (1-10). These nongenetic factors have a substantial impact on disease risk and are frequent. For example, metabolic syndrome poses an eight times higher T2D risk and is present in more than 40% of the individuals over 50 years of age. The high impact and frequency make these risk factors suitable candidates for targeting preventive interventions, such as medication, weight loss, and increased physical activity that can slow down or even reverse the disease process (11, 12).

In the past 5 years, genome-wide association studies have identified and replicated over 40 single nucleotide polymorphisms (SNPs) that predispose to T2D (13, 14). However, the effect sizes of the associated variants are very modest, with per allele odds ratios ranging from 1.05 to 1.35 (13). Even the strongest susceptibility variant, rs7903146 in the TCF7L2 gene, is a weaker predictor of T2D risk than most nongenetic risk factors. Evidently, the low effect sizes make single genetic risk factors unsuitable for targeting preventive interventions, but there is increasing interest in investigating the extent to which genetic risk factors combined can improve the prediction of the disease.

An improvement in the early identification of high-risk groups is warranted because T2D imposes a great burden on human health and health care systems (15, 16). An estimated 285 million people worldwide have diabetes (15) and this number is expected to increase by more than 50% in the next 20 years if no preventive strategies are implemented (15). To identify high-risk individuals, many risk prediction models have been proposed.

Guidelines for T2D prevention advocate the use of clinical risk scores as primary screening tools, followed by blood glucose measurements to detect individuals with impaired fasting glucose, impaired glucose tolerance, or metabolic syndrome (17). Examples of commonly used risk scores include the FINDRISC (Finnish Diabetes Risk Score) and the Diabetes Risk Calculator (18, 19). The FINDRISC score is based on age, BMI, waist circumference, use of antihypertensive medication, history of elevated blood glucose, daily physical activity and daily intake of fruits or vegetables, and the Diabetes Risk Calculator on age, waist circumference, gestational diabetes, height, race/ethnicity, hypertension, family history of diabetes, and exercise.

The predictive ability of these clinical risk scores is modest, but satisfactory. The area under the receiver operating characteristic curve (AUC) is a commonly used measure to indicate the predictive ability; the AUC indicates the discriminative accuracy of a prediction model. To generate the curve, on the x-axis 1-specificity is plotted, and on the y-axis sensitivity is plotted.

The AUC value represents the probability that the predicted risk of a random "patient" is higher than that of a random "nonpatient." When predicted risks of individuals who will develop the disease are always higher than the risks of those who will not develop the disease, the AUC is 1.0. When their risks are higher for 50% of the random pairs, the AUC is 0.50, equaling the predictive performance of tossing a coin (20).

The AUC was 0.65 in men and 0.66 in women for the FINDRISC score predicting impaired fasting glucose, impaired glucose tolerance, or undiagnosed diabetes, and 0.72 and 0.75 for detecting metabolic syndrome (18). The AUC of the Diabetes Risk Calculator was 0.70 for detecting impaired fasting glucose, impaired glucose tolerance, or undiagnosed diabetes (19). These modest AUC values indicate that many people who will develop T2D are not identified as being at increased risk by these risk scores, and that many that will not develop the disease are labeled as increased risk. Although offering lifestyle modification programs to individuals who will not develop T2D may do no harm and may even provide other benefits by reducing the risk of other diseases, not recognizing the many who will develop diabetes would clearly be missed opportunities to reduce the serious burden of disease (12). Some clinical risk models that include invasive measurements showed higher AUC values for detecting individuals who will develop T2D.

Table 1. Risk factors for type 2 diabetes.

Risk factor	Population	Frequency (%)	Diabetes risk (%)[a]	RR[b]
Age (y)				
0 to 44	General US population	61.3 [1]	1.7 [2]	1
45 to 64		25.9	12.2	7.2
65 to 74		6.8	19.9	11.7
75+		6.1	17.9	10.5
Sex				
Female	General US population	50.7 [1]	5.9 [3]	1
Male		49.3	6.6	1.1
BMI (kg/m^2)				
<25	US adults ages ≥20 years	32.0 [4, 5]	8 [6]	1
25 to <30		34.2	15	1.9
30 to <35		19.5	23	2.9
35 to <40		8.6	33	4.1
≥40		5.7	43	5.4
IFG/IGT				
Normoglycemic	Nondiabetic US adults ages≥18 years	65.4 [7]	NA [8]	1[d] [8]
IGT only	(frequency)	5.4	4.4–6.4[c]	5.5[d]
IFG only	Global cohorts (diabetes risk and RR)	19.4	6.1–9.2[c]	7.5[d]
IFG+IGT		9.8	10–15[c]	12.1[d]
HbA1c (%)				
<5.0	Nondiabetic middle-aged adults from 4 US communities	8.6 [9]	6[e] [9]	0.5[f] [9]
5.0 to <5.5		44.6	12[e]	1[f]
5.5 to <6.0		33.2	21[e]	1.9[f]
6.0 to<6.5		9.3	44[e]	4.5[f]
≥6.5		4.3	79[e]	16.5[f]
Metabolic syndrome				
No	US adults ages ≥50 years	56.5 [10]	4.1 [10]	1
Yes		43.5	34.0	8.3

[a]Values reported are prevalences unless otherwise indicated
[b]Unless referenced, values are calculated from the values depicted in the column "Diabetes risk"
[c]Annualized incidence of diabetes
[d]Annualized relative risk
[e]Cumulative 15-year incidence of diagnosed diabetes
[f]Multivariable adjusted hazard ratio of 15-year risk for each absolute increase in 1percentage point of glycated hemoglobin
BMI body mass index; *HbA_{1c}* glycated hemoglobin; *IFG* impaired fasting glucose; *IGT* impaired glucose tolerance; *NA* not available; *RR* relative risk

Table 2. Genetic risk prediction studies in T2D.

Study	No of polymorphisms	Clinical risk factors	AUC genetic	AUC clinical	AUC combined	Design	Age (mean, years)[a]	Sex (% men)[a]	BMI (mean, kg/m2)[a]
European									
Balkau et al. (23)									
Men	2	FPG, smoking status, WC, GGT	NA	0.85	0.85	Prospective cohort	50/47	100/100	27.5/25.1
Women	2	FPG, BMI, FH, TG	NA	0.92	0.91	Prospective cohort	52/47	0/0	29.2/23.7
Lyssenko et al. (24•, 25)	3	BMI, FPG	NA	0.68	0.68	Prospective cohort	45.1[b]	51/46	25.3[b]
Weedon et al. (26)	3	NA	0.58	NA	NA	Case–control	48.7/31.8	58/50	31.4/27.2
Vaxillaire et al. (27)	3	Age, sex, BMI	0.56	0.82	0.83	Prospective cohort	47.7[b]	50	24.3[b]
Cornelis et al. (28)	10	Age, sex, BMI, FH, smoking, alcohol intake, PA	NA	0.78	0.79	Nested case-control	49.0/48.1	43/38	27.7/24.4
Lyssenko et al. (29)									
Malmo	11	Age, sex, BMI, FH, BP, TG, FPG	0.63	0.74	0.75	Prospective cohort	45.5	64.9	24.3
Botnia	11	Age, sex, BMI, FH, BP, TG, FPG, HDL, WC	0.68	0.79	0.8	Prospective cohort	44.9	45.5	25.6
Cauchi et al. (30)	15	Age, sex, BMI	NA	NA	0.86	Case–control	62.9/54.7	62/42	29.0/24.7
Lin et al. (31)	15	Age, sex, FH, PA, WHR, triacylglycerol/HDL ratio	0.59	0.86	0.87	Cross-sectional	60.7/52.8	67/46	30.4/25.5
Fontaine-Bisson et al. (32)	17	Age, sex	NA	NA	0.59	Cross-sectional	53.6/53.1	58.4/50.2	29.5/25.8
van Hoek et al. (33)	18	Age, sex, BMI	0.6	0.66	0.68	Prospective cohort	68.2/69.0	44/40	28.0/26.0

Study	No of polymorphisms	Clinical risk factors	AUC genetic	AUC clinical	AUC combined	Design	Age (mean, years)[a]	Sex (% men)[a]	BMI (mean, kg/m2)[a]
Lango et al. (34)	18	Age, sex, BMI	0.6	0.78	0.8	Case–control	55.7/NA	56/51	31.5/26.9
Meigs et al. (35)	18	Age, sex, BMI, FH, FPG, SBP,HDL, TG	0.58[c]	0.9	0.9	Prospective cohort	42.1	47	25.6
Sparso et al. (36)	19	Age, sex, BMI	0.6	0.92	0.93	Case–control	60/47	59.3/46.3	30.6/25.6
Wang et al. (37)									
FINDRISC	19	Age, BMI, WC,PA, FH,diet, antihypertensive medication, previously known high glucose	0.55	0.73	0.73	Cross-sectional	45–74[d]	100/100	NA
FINDRISC+	19	FINDRISC, TG, HDL, adiponectin, ALT	0.55	0.77	0.77	Cross-sectional	45–74[d]	100/100	NA
Schulze et al. (38)	20	Age, WC, height, history of HT, PA, smoking, consumption of redmeat, whole-grain bread, coffee and alcohol, glucose, HbA_{1c}, TG, HDL, GGT, ALT, hs-CRP	NA	0.9	0.9	Prospective case-cohort	54.6/49.4	58.7/36.9	30.4/25.9
Talmud et al. (39•)									
Cambridge score	20	Age, sex, BMI, drug treatment, FH, smoking status	0.55	0.72	0.73	Prospective cohort	51.0/49.0	72.9/72.8	27.5/24.7
Framingham offspring score	20	Age, sex, BMI, parental history of T2D, HDL, TG, FPG	0.55	0.78	0.78	Prospective cohort	51.0/49.0	72.9/72.8	27.5/24.7
de Miguel-Yanes et al. (40)	40	Age, sex, FH, BMI, FPG, SBP,HDL, TG	0.61[c]	0.9	0.91	Prospective cohort	46	47	26

Study	No of polymorphisms	Clinical risk factors	AUC genetic	AUC clinical	AUC combined	Design	Age (mean, years)[a]	Sex (% men)[a]	BMI (mean, kg/m2)[a]
Asian									
Miyake et al. (41)	11	Age, sex, BMI	0.63	0.68	0.72	Case-control	61.3/67.5	58/46	23.6/23.3
Hu et al. (42)	11	Age, sex, BMI	0.62	0.61	0.67	Case-control	61.2/57.4	52/41	24.0/23.6

[a]Values provided are for participants with and without T2D, respectively, when two values are reported and for the total population when one value is reported. For prospective cohort studies, descriptive data from baseline examinations are given. Values are means unless otherwise indicated

[b]Median

[c]Adjusted for sex

[d]Range

ALT alanine aminotransferase; *AUC* area under the receiver operating characteristic curve; *BMI* body mass index; *BP* blood pressure; *FH* family history of type 2 diabetes; *FINDRISC* finnish diabetes risk score; *FPG* fasting plasma glucose; *GGT* γ-glutamyltransferase; *HbA1c* glycated hemoglobin; *HDL* high-density lipoprotein cholesterol; *hs-CRP* high-sensitivity C-reactive protein; *HT* hypertension; *NA* not available; *PA* physical activity; *SBP* systolic blood pressure; *TG* triglycerides; *T2D* type 2 diabetes; *WC* waist circumference; *WHR* waist-hip ratio

(Adapted from Mihaescu et al. (43•))

Figure 1. The area under the receiver operating characteristic curve (AUC) versus the number of single nucleotide polymorphisms included in the genetic risk models.

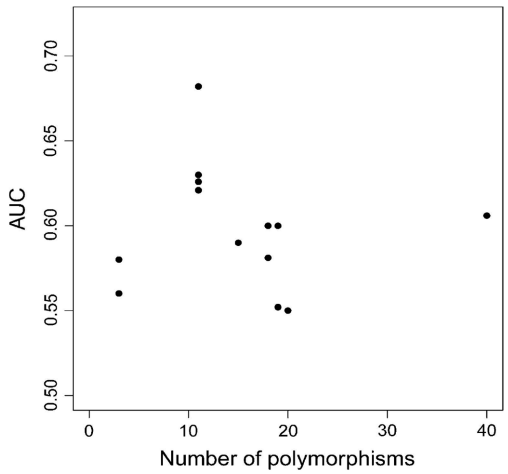
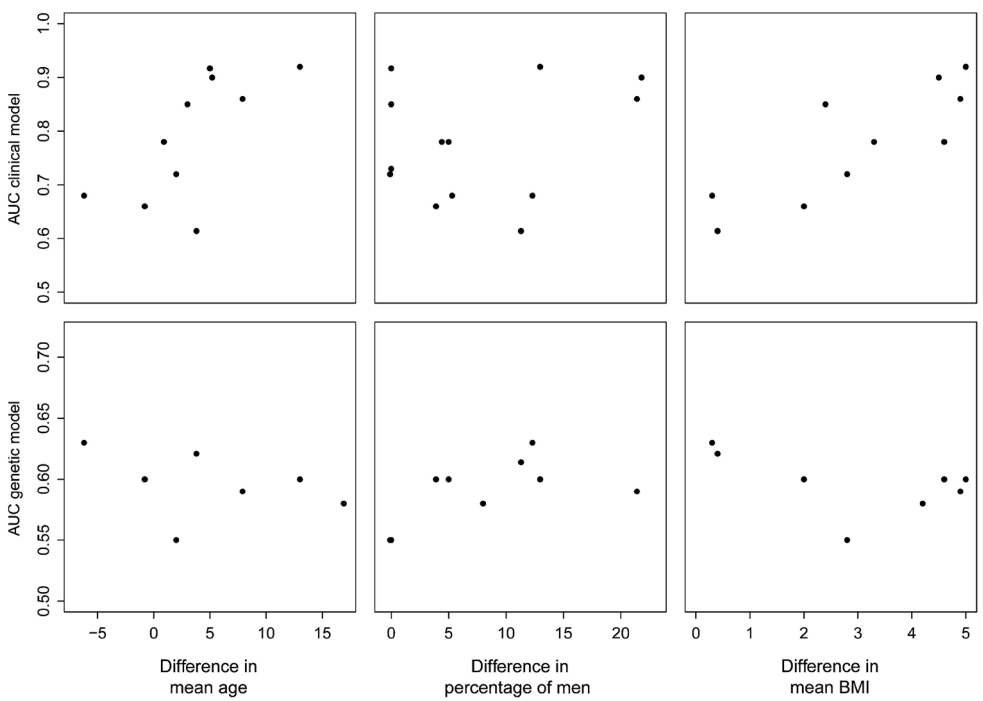


Figure 2. The area under the receiver operating characteristic curve (AUC) of the genetic and clinical models in relation to differences in mean age, percentage of men, and mean body mass index (BMI) between patients and controls.



An example is the Framingham Risk Score including age, sex, obesity, hypertension, parental history of diabetes, low levels of high-density lipoprotein cholesterol, elevated triglyceride levels, and impaired fasting glucose (21). The AUC of this risk model was 0.85 for predicting T2D in middle-aged adults (21). However, inclusion of invasive measurements that can change over time in clinical risk models might be inconvenient at the population level and these models still leave room for improvement.

Recent studies have investigated the predictive ability of risk models that include genetic variants only or genetic variants added to clinical risk factors. A study that investigated a genetic risk score based on 34 diabetes-associated variants showed a significant association of the risk score with risk of developing diabetes (22). This risk was attenuated by lifestyle interventions, also in individuals in the highest genetic risk quartile, suggesting that detecting individuals at high risk of developing T2D based on genetic variants and offering them lifestyle modification programs is useful. In this paper, we review genetic risk prediction studies from a methodologic perspective by focusing on factors in the choice of study design and population that may have impacted the observed predictive ability.

Genetic Risk Prediction Studies

The number of studies that investigate the predictive ability of genetic variants in T2D has increased rapidly (**Table 2**; (23, 24, 25–38, 39, 40–42, 43)). These studies assessed risk models that were based on genetic variants only or on a combination of both genetic and nongenetic variants. The table shows that the number of SNPs included in the genetic models has increased from 3 in 2005 to 40 in 2011. The models show considerable overlap in the genetic variants that were considered, but there also are many differences. Since its discovery, all but one of the studies had included TCF7L2 and the majority additionally investigated PPARG, CDKN2A/B, KCNJ11, IGF2BP2, SLC30A8, and HHEX-IDE-KIF11. Yet, most other SNPs were included in one or two models only (43). The same was observed for the clinical models. Most clinical models included at least age, sex, and BMI, but they differed in the other factors that were added, such as blood pressure, family history of T2D, and fasting plasma glucose level.

Table 2 shows that, almost without exception, the genetic risk models had lower AUC values than the clinical models. The AUC values for the genetic models ranged from 0.55 to 0.68 and for the clinical models from 0.61 to 0.92. **Table 2** also shows that the addition of genetic factors either did not or only marginally improved the AUC beyond that of the clinical risk models.

Predictive Ability of Clinical Risk Models

The differences in the predictive ability of clinical risk models are explained by how many and which risk factors are included in the model and by differences in study design and study population. This is nicely illustrated by three studies that had investigated largely the same 18 genetic variants. The AUCs of the genetic risks models in these studies were similar (0.58–0.60), but the AUCs of the clinical models were 0.66, 0.78, and 0.90 (33-35). The clinical models with AUC values of 0.66 and 0.78 included age, sex, and BMI, but the model with an AUC value of 0.90 also included T2D family history, fasting plasma glucose, systolic blood pressure, high-density lipoprotein cholesterol, and triglycerides. The excellent predictive ability was likely due to the inclusion of fasting plasma glucose, as individuals with impaired fasting glucose have a very high risk of developing T2D (**Table 1**). **Table 2** shows that AUC values tend to be higher when more risk factors are included in the model, particularly when fasting plasma glucose was included.

Yet, also the two studies that both investigated age, sex, and BMI in the clinical model had markedly different AUC values (0.66 and 0.78). The difference in these AUC values was likely explained by differences in the study design and population. The AUC of 0.66 was obtained in a prospective cohort study, the Rotterdam Study, and the AUC of 0.78 in a case-control study, consisting of case and control subjects from the GoDARTS (Genetics of Diabetes Audit and Research Tayside Study). Participants in the Rotterdam Study were older and less often men (**Table 2**), but the two populations predominantly differed in BMI. The mean BMI of the cases in the GoDARTS study was higher than the mean BMI of cases in the Rotterdam Study (31.5 vs 28.0 kg/m^2). Also, the difference in mean BMI between cases and controls was much larger in the GoDARTS study compared with the Rotterdam Study (4.6 vs 2.0 kg/m^2). In general and by definition, the predictive ability of risk models is higher when there are larger differences between cases and controls on the risk factors included in the risk model. Along the same lines, study design and population characteristics may have influenced the observed AUC values of the other clinical models, and also of AUC values of the genetic risk models.

Predictive Ability of Genetic Risk Models

The AUC values of the genetic risk models ranged from 0.55 to 0.68, a range that was much smaller than that of the clinical models. Similar as for the clinical risk models and given that all SNPs approximately have the same low effect size, one would expect better predictive ability for models that included a higher number of SNPs, but **Figure 1** shows that this was not observed for the studies listed in **Table 2**. The differences in the AUC values of the genetic risk scores cannot be explained by the number of polymorphisms included in the risk models. In fact, the highest

genetic AUC (0.68) was found for a model that included 11 SNPs, and the lowest for a model that included these exact 11 SNPs plus an additional 8 others. The explanation for the absence of this relationship is likely in the low effect sizes of the genetic variants. A higher number of SNPs only yields a slightly higher AUC, a combined effect that could easily be outweighed by the influence of other factors, such as study design and study population.

Genetic risk prediction models have been investigated in prospective cohort studies, in case-control studies and in cross-sectional studies, and in study populations that differed in age, sex, and BMI (**Table 2**). These have influenced the observed predictive ability of the genetic risk models. There are two ways in which these characteristics may impact the predictive ability: the clinical and demographic characteristics of the study population itself and the differences in these characteristics between patients and nonpatients.

Table 2 describes mean age and BMI and the percentage of men in published genetic risk prediction studies for T2D. Mean age varied from 42.1 to 68.9 years, mean BMI from 23.4 to 29.1 kg/m^2, and the percentage of men from 0% to 100%. It is often hypothesized that genetic risk factors may be more predictive in populations in which nongenetic T2D risk factors are not yet present (eg, in younger or normal weight cohorts), but AUC values of the genetic models were not markedly higher when populations were younger, had lower BMI, or had a lower percentage of men. However, because of the heterogeneity between the studies and their relatively small number, conclusions must be drawn with caution. Moreover, one study that had investigated the predictive performance in two age categories (<50 years vs ≥50 years) did find higher AUC values for the genetic risk score in younger people (AUC 0.66 vs 0.59) (40). The observation that a stratified analysis within a single study did show differences in predictive ability suggests that the absence of a clear relation of age, BMI, and sex with AUC values across studies is likely explained by the presence of other differences between the studies.

The other way in which clinical and demographic characteristics of the study population impact the predictive ability of risk models is through differences in these characteristics between patients and nonpatients. This specifically holds for characteristics that are included as risk factors in the prediction models, and for characteristics that are associated with these risk factors. Evidently and by definition, the presence of risk factors will differ between patients and nonpatients, but the difference can also be enlarged as a result of selection procedures. For example, patients who are recruited through hospitals may have more unfavorable risk profiles than patients randomly selected from the total patient population. Consequently, differences in risk factors between hospital-based cases and population-based controls will be larger and the impact of these risk factors on the predictive ability higher. For the studies listed in **Table 2**, differences in mean age ranged from -6.2 to 16.9 years,in mean BMI from 0.3 to 5.5 kg/m^2,

and differences in the percentage of men from -0.1% to 21.8%. **Figure 2** shows that larger differences in mean age and BMI between patients and nonpatients were associated with higher AUC values for the clinical risk models, and, although less apparent, lower AUC values for the genetic models. No relation was observed between clinical AUC values and the percentage of men included in the studies, but this may be because male sex only marginally increases T2D risk compared with age and BMI (**Table 1**).

A second methodological aspect that may impact the predictive ability of risk models is study design. Genetic risk prediction studies are preferably conducted in prospective follow-up studies, but cross-sectional and case-control studies have been used as well (**Table 2**). The impact of study design on AUC values of T2D risk prediction models is in part related to the impact of population characteristics. Selection procedures for cases and controls may affect differences in clinical and demographic characteristics between patients and nonpatients. Case-control studies may demonstrate AUC values that deviate from those observed in prospective cohort and cross-sectional studies when cases and controls are recruited from different sources.

Another way in which study design may impact the predictive ability of risk models is length of follow-up in prospective cohort studies. Longer follow-up increases the likelihood that clinical T2D risk factors change over time, and that as a result their baseline values will be less predictive for the development of disease, resulting in prediction models with lower AUC. The length of follow-up of the studies listed in **Table 2** varied from 6 to 25 years. Again, the number of prospective cohort studies was too small to investigate the impact of follow-up duration, but one study investigated the predictive ability in quintiles of follow-up time. This study demonstrated that the AUC of the clinical risk model decreased with increasing duration of follow-up, whereas the AUC of the genetic risk model increased (29). From the first to the fifth quintile, the clinical AUC value decreased from 0.75 to 0.67 and the genetic AUC value increased from 0.57 to 0.62 (29).

CONCLUSIONS

In this review, we showed that study design and population characteristics may have affected the observed predictive performance of risk models. AUC values of the clinical risk models were higher and, although weaker, AUC values of the genetic risk models were lower when there were larger differences in age and BMI between cases and controls. This observation has important implications for the design and health care relevance of genetic risk prediction studies.

The predictive ability of risk models is preferably investigated in prospective cohort studies, but in practice often only case-control or cross-sectional designs are available. Because clinical risk factors, particularly the difference in risk factors between cases and controls, impact

AUC values, it is expected that AUC values for genetic risk models obtained in case–control or cross-sectional studies may be valid when the distribution of these risk factors does not differ from prospective studies. For case–control studies, this means that the selection of cases and controls is not affected by these risk factors. In case of selection, transparency about the methods is important to enable a correct interpretation of the scientific and health care relevance of the results. For this reason, the GRIPS (Genetic Risk Prediction Studies) statement, a recently published guideline for the reporting of genetic risk prediction studies, recommends to describe eligibility criteria for participants, and sources and methods of selection of participants (44).

The observed impact of population characteristics implies that it is important to assess the predictive ability of risk scores in representative samples of the population in which the model is ultimately applied to get valid estimates of their performance in that population. The question then is: which populations do we want to target for the prevention of T2D? Evidently, these may include individuals with metabolic syndrome or overweight, but for genetic prediction this may particularly concern young individuals who have not developed clinical risk factors. To date none of the T2D risk prediction studies have been conducted in younger populations; all studies were conducted in populations which on average were older than 40 years of age, two even in populations over 60 years of age (33, 41). The study that best approximates the desired study population has been conducted in a population with a mean age of 42 years, a mean BMI of 25.6 kg/m^2, and an almost equal number of men and women (35). Given the observed differences in AUC values, we must conclude that we do not know whether genetic variants are useful in predicting T2D risk in younger populations. None of the studies so far has started from a health care perspective when investigating the predictive ability of T2D risk models.

There is increasing interest in investigating the value of genetic risk factors in the prediction of T2D risk. In this review, we demonstrated that the choice of study design and predominantly the choice of study population impact the observed predictive ability of risk models. For this reason it is important that the planning of future genetic risk prediction studies in T2D starts from a health care perspective by asking in which population we want to predict T2D risk. It is the answer to this question that determines the population in which the predictive ability should be assessed and that determines whether the results of the study ultimately can be informative and change health care practice.

REFERENCES

1. U.S. Census Bureau: General demographic characteristics, 2009 population estimates, United States. Available at http://factfinder. census.gov/servlet/QTTable?-ds_name=PEP_2009_EST&-qr_ name=PEP_2009_EST_DP1&-geo_id=01000US. Accessed June 2011.
2. Centers for Disease Control and Prevention: Percentage of civilian, noninstitutionalized population with diagnosed diabetes, by age, United States, 1980–2009. Available at http://www.cdc.gov/ diabetes/statistics/prev/national/figbyage.htm. Accessed June 2011.
3. Centers for Disease Control and Prevention: Age-adjusted percentage of civilian noninstitutionalized population with diagnosed diabetes, by sex, United States, 1980–2009. Available at http://www.cdc.gov/diabetes/statstics/prev/national/figbysex.htm. Accessed June 2011.
4. Centers for Disease Control and Prevention: Prevalence of overweight, obesity, and extreme obesity among adults: United States, trends 1976–1980 through 2007–2008. Available at http://www.cdc. gov/nchs/data/hestat/obesity_adult_07_08/obesity_adult_07_08.pdf. Accessed June 2011.
5. Flegal KM, Carroll MD, Ogden CL, et al. **Prevalence and trends in obesity among US adults, 1999–2008**. *JAMA*. 2010;**303**:235-41.
6. Nguyen NT, Nguyen XT, Lane J, et al. **Relationship between obesity and diabetes in a US adult population: findings from the National Health and Nutrition Examinaton Survey, 1999–2006**. *Obes Surg*. 2011;**21**:351-5.
7. Karve A, Hayward RA. **Prevalence, diagnosis, and treatment of impaired fasting glucose and impaired glucose tolerance in nondiabetic U.S. adults.** *Diabetes Care.* 2010;**33**:2355-9.
8. Gerstein HC, Santaguida P, Raina P, et al. **Annual incidence and relative risk of diabetes in people with various categories of dysglycemia: a systematic overview and meta-analysis of prospective studies.** *Diabetes Res Clin Pract.* 2007;**78**:305-12.
9. Selvin E, Steffes MW, Zhu H, et al. **Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults.** *N Engl J Med.* 2010;**362**:800-11.
10. Alexander CM, Landsman PB, Teutsch SM, et al. **NCEP-defined metabolic syndrome, diabetes, and prevalence of coronary heart disease among NHANES III participants age 50 years and older.** *Diabetes.* 2003;**52**:1210-4.
11. Gillies CL, Abrams KR, Lambert PC, et al. **Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and metaanalysis.** *BMJ.* 2007;**334**:299.
12. Colagiuri S. **Epidemiology of prediabetes.** *Med Clin North Am.* 2011;**95**:299–307.
13. McCarthy MI. From hype to hope? **A journey through the genetics of type 2 diabetes.** *Diabet Med.* 2011;**28**:132-40.
14. McCarthy MI. **Genomics, type 2 diabetes, and obesity.** *N Engl J Med.* 2010;**363**:2339-50.
15. International Diabetes Federation: Diabetes and impaired glucose tolerance. Global burden: prevalence and projections, 2010 and 2030. Available at: http://www.idf.org/diabetesatlas/diabetes-andimpaired-glucose-tolerance. Accessed June 2011.
16. Stumvoll M, Goldstein BJ, van Haeften TW. **Type 2 diabetes: principles of pathogenesis and therapy.** *Lancet.* 2005;**365**:1333-46.
17. Paulweber B, Valensi P, Lindström J, et al. **A European evidence-based guideline for the prevention of type 2 diabetes.** *Horm Metab Res.* 2010;**42 Suppl 1**:S3-36.

18. Saaristo T, Peltonen M, Lindström J, et al. **Cross-sectional evaluation of the Finnish Diabetes Risk Score: a tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome.** *Diab Vasc Dis Res.* 2005;**2**:67-72.

19. Heikes KE, Eddy DM, Arondekar B, et al. **Diabetes Risk Calculator: a simple tool for detecting undiagnosed diabetes and prediabetes.** *Diabetes Care.* 2008;**31**:1040-5.

20. Janssens ACJW, Khoury MJ. **Assessment of improved prediction beyond traditional risk factors – when does a difference make a difference?** *Circ Cardiovasc Genet.* 2010;**3**:3–5.

21. Wilson PWF, Meigs JB, Sullivan L, et al. **Prediction of incident diabetes mellitus in middle-aged adults.** *Arch Intern Med.* 2007;**167**:1068-74.

22. Hivert MF, Jablonski KA, Perreault L, et al. **Updated genetic score based on 34 confirmed type 2 diabetes loci is associated with diabetes incidence and regression to normoglycemia in the Diabetes Prevention Program.** *Diabetes* 2011;**60**:1340-8. This study investigated effectiveness of preventive interventions in individuals in all quartiles of genetic risk estimated by a genetic risk score. This is very interesting from a public health perspective.

23. Balkau B, Lange C, Fezeu L, et al. **Predicting diabetes: clinical, biological, and genetic approaches: data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR). Diabetes Care**. 2008;*31*:2056-61.

24. Lyssenko V, Almgren P, Anevski D, et al. **Genetic prediction of future type 2 diabetes.** *PLoS Med.* 2005;**2**:e345. This was the first paper in T2D that considered the combined effect of multiple susceptibility variants in the prediction of disease risk.

25. Janssens ACJW, Gwinn M, Subramonia-Iyer S, et al. **Does genetic testing really improve the prediction of future type 2 diabetes?** *PLoS Med.* 2006;**3**:e114.

26. Weedon MN, McCarthy MI, Hitman G, et al. **Combining information from common type 2 diabetes risk polymorphisms improves disease prediction.** *PLoS Med.* 2006;**3**:e374.

27. Vaxillaire M, Veslot J, Dina C, et al. **Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study.** *Diabetes.* 2008;**57**:244-54.

28. Cornelis MC, Qi L, Zhang C, et al. **Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry**. *Ann Intern Med*. 2009;**150**:541-50.

29. Lyssenko V, Jonsson A, Almgren P, et al. **Clinical risk factors, DNA variants, and the development of type 2 diabetes.** *N Engl J Med.* 2008;**359**:2220-32.

30. Cauchi S, Meyre D, Durand E, et al. **Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value. PLoS One.** 2008;*3*:e2031.

31. Lin X, Song K, Lim N, et al. **Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score–the CoLaus Study.** *Diabetologia.* 2009;**52**:600-8.

32. Fontaine-Bisson B, Renström F, Rolandsson O, et al. **Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population.** *Diabetologia.* 2010;**53**:2155-62.

33. van Hoek M, Dehghan A, Witteman JCM, et al. **Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study.** *Diabetes.* 2008;**57**:3122-8.

34. Lango H, Consortium UKTDG, Palmer CNA, et al. **Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk.** *Diabetes.* 2008;**57**:3129-35.

35. Meigs JB, Shrader P, Sullivan LM, et al. **Genotype score in addition to common risk factors for prediction of type 2 diabetes.** *N Engl J Med.* 2008;**359**:2208-19.

36. Sparso T, Grarup N, Andreasen C, et al. **Combined analysis of 19 common validated type 2 diabetes susceptibility gene variants shows moderate discriminative value and no evidence of gene-gene interaction.** *Diabetologia.* 2009;**52**:1308-14.

37. Wang J, Stancakova A, Kuusisto J, et al. **Identification of undiagnosed type 2 diabetic individuals by the Finnish diabetes risk score and biochemical and genetic markers: a population-based study of 7232 Finnish men.** *J Clin Endocrinol Metab.* 2010;**95**:3858-62.

38. Schulze MB, Weikert C, Pischon T, et al. **Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study.** *Diabetes Care*. 2009;**32**:2116-9.

39. Talmud PJ, Hingorani AD, Cooper JA, et al. **Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study.** *BMJ* 2010;**340**:b4838. This study is a very good example of a genetic risk prediction study, with a very complete description of the study population and a thorough assessment of the predictive ability.

40. de Miguel-Yanes JM, Shrader P, Pencina MJ, et al. **Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms.** *Diabetes Care.* 2011;*34*:121-5.

41. Miyake K, Yang W, Hara K, et al. **Construction of a prediction model for type 2 diabetes mellitus in the Japanese population based on 11 genes with strong evidence of the association.** *J Hum Genet.* 2009;**5**4:236-41.

42. Hu C, Zhang R, Wang C, et al. **PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population.** *PLoS One.* 2009;**4**:e7643.

43. Mihaescu R, Meigs J, Sijbrands E, et al. **Genetic risk profiling for type 2 diabetes.** *PloS Curr.* 2011;**3**:RRN1208. This paper is regularly updated and gives an overview of all genetic risk prediction studies conducted in T2D.

44. Janssens AC, Ioannidis JP, van Duijn CM, et al. **Strengthening the reporting of genetic risk prediction studies: the GRIPS statement.** *Eur J Epidemiol.* 2011;**26**:255-9. This paper describes reporting guidelines to enhance the transparency, quality, and completeness of the reporting of genetic risk prediction studies. A longer paper with examples and explanations was published simultaneously.

Chapter 11

Genetic risk profiling for prediction of type 2 diabetes

Raluca Mihaescu, James Meigs, Eric J.G. Sijbrands,

A. Cecile J.W. Janssens

PLoS Currents. 2011;3:RRN1208

ABSRACT

Type 2 diabetes (T2D) is a common disease caused by a complex interplay between many genetic and environmental factors. Candidate gene studies and recent collaborative genome-wide association efforts revealed at least 38 common single nucleotide polymorphisms (SNPs) associated with increased risk of T2D. Genetic testing of multiple SNPs is considered a potentially useful tool for early detection of individuals at high diabetes risk leading to improved targeting of preventive interventions.

CLINICAL SCENARIO

Both a population-based approach and a targeted high-risk approach are recommended as strategies for prevention of T2D. Several recent guidelines advocate screening for individuals at risk to develop T2D followed by blood glucose measurements to detect individuals with impaired fasting glucose (IFG) or impaired glucose tolerance (IGT) (1). Genetic testing of a panel of SNPs may be useful in detecting such groups of high-risk individuals in whom screening for T2D could be optimized.

Test Description

Genetic susceptibility testing for T2D is currently offered by several commercial companies that use genome-wide scans to deliver information about risk for many common complex diseases (see **Table 1**). For example, deCODEme offers predictions for 50 complex diseases and non-disease phenotypes that vary from breast cancer, atrial fibrillation, T2D or psoriasis, to eye color and bitter taste perception (2). Tests are available for purchase directly to the individual consumers, or through the request from a physician (see **Table 2**).

Direct-to-consumer risk companies sell risk profiles that differ in the number of genetic markers included and in the exact SNPs used. For example, deCODEme uses 21 SNPs from the genome-wide scan to calculate the risk of T2D for individuals with European descent, 9 SNPs for East Asians and 2 SNPs for African Americans (3). A test based on the same markers is also available as a separate T2D profile (4). Pathway Genomics offers also separate tests for individuals of African, Asian and Caucasian origin (5), 23andMe uses 9 SNPs to determine the risk of developing T2D (6), Navigenics tests18 SNPs (7), and GeneticHealth, an UK based company, calculates the risk for obesity, diabetes and weight loss using the same 7 SNPs (8).

Genetic risks are calculated on the basis of literature data. The companies take an average risk from some epidemiological study and multiply this with the odds ratios from published meta-analyses or large scale genome-wide association studies (9). Importantly, the companies do not use information about clinical risk factors when calculating the risk of disease. When available, some companies use sex, ethnicity and age matched population risks to depart from.

Table 1 Direct-to-consumer companies that sell genetic tests for Type 2 Diabetes risk.

Company name	Predict multiple diseases at the same time	URL
deCODEme	Yes	http://www.decodeme.com
	No	http://www.decodediagnostics.com
23andMe	Yes	https://www.23andme.com
Navigenics	Yes	http://www.navigenics.com
Pathway Genomics	Yes	http://www.pathway.com
GeneticHealth	No	http://www.genetic-health.co.uk
GHC Genetics	Yes	http://www.genscan.com/en/
CyGene Direct	Yes	http://www.cygenedirect.com

Table 2. Direct-to-consumer companies.

Company name	Delivery model	Accreditation
deCODEme	DTC	CLIA
23andMe	DTC	CLIA
Navigenics	Through physician or corporate wellness program	CLIA
Pathway Genomics	Through physician	CLIA
GeneticHealth	DTC	Not mentioned on website
GHC Genetics	DTC and through physician	ISO 9001:2001 ISO 27001:2006
CyGene Direct	DTC	CLIA

CLIA, Clinical Laboratory Improvement Amendments of 1988; DTC, direct-to-consumer.

Public Health Importance

T2D is a metabolic disorder characterized by hyperglycemia, insulin resistance and relative insulin deficiency. Diabetes is a leading cause of blindness, renal failure and limb amputation, and a major risk factor for cardiovascular morbidity and mortality (10). It is estimated that approximately 285 million people worldwide will have diabetes in 2010. This number is expected to increase by more than 50% in the next 20 years if no preventive strategies are implemented (11). Diabetes is responsible for almost four million deaths worldwide in the 20-79 age group in 2010, representing 6.8% of global all-cause mortality in this age group (11).

Preventive interventions for T2D, including medication, weight loss and increased physical activity, can slow or even reverse the disease process (12). For example, the United States Diabetes Prevention Program trial investigated the efficacy of intensive lifestyle interventions or metformin treatment compared to standard lifestyle recommendations (13). Lifestyle intervention resulted in 58% T2D risk reduction compared to the placebo arm, at 2.8 years of follow-up. For the same follow-up, metformin resulted in 31% T2D risk reduction (13). Genetic

tests are claimed by the DTC companies to improve risk prediction and increase adherence to preventive interventions (e.g., "Knowledge is self-empowering and it can motivate you towards taking steps that reduce other risk factors, which have been found to contribute to your genetic predisposition risk" (14)), thus helping to improve outcomes and reduce the costs and burden of disease for society (e.g., "The conditions included in Navigenics' analysis are those that are clinically actionable and those that contribute to the major burden of disease in the United States, such as myocardial infarction, cancer, and type 2 diabetes." (15))

Published Reviews, Recommendations and Guidelines

Systematic evidence reviews

None identified.

Recommendations by independent group

None identified.

Guidelines by professional groups

None identified.

A European multidisciplinary consortium developed an evidence-based guideline for the prevention of T2D. The consortium advocates the use of clinical risk scores as primary screening tools to identify high-risk groups in whom T2D screening may be targeted more efficiently. One such example is the Finnish risk test (FINDRISC) that provides ten-year risks to develop T2D. The FINDRISC score contains eight items: age, BMI, waist circumference, antihypertensive medication, history of elevated blood glucose, daily physical activity and daily intake of fruits or vegetables. In the context of targeted screening, the guideline includes the following recommendation about genetic testing**:** "despite the encouraging progress in our understanding of the genetic basis of T2DM, it is too early to use genetic information as a tool for targeting preventive efforts".[1] No other guidelines provide recommendations for or against the use of genetic testing for screening, prevention or treatment of T2D.

Evidence Overview

*Analytic Validity***:** Test accuracy and reliability in identifying multiple SNPs (analytic sensitivity and specificity).

Navigenics reports an analytic accuracy of 99% (15), deCODEme does not provide a measure of accuracy but describes the methods used to ensure good analytic validity (16), 23andMe does not disclose the methods used (17), and the same applies to Pathway Genomics (18). CyGene Direct briefly mentions the methods used to ensure good analytic validity (19). No information was available on the analytic validity of commercial tests for GeneticHealth and GHC Genetics.

Direct-to-consumer genetic testing services are not clearly regulated by governmental agencies. Their services may bypass healthcare providers who are typically responsible for appropriate ordering of lab tests and for discussing with patients the implications of test results. Not all companies explicitly report the analytic and clinical performance characteristics of their test systems. Following a recent Government Accountability Office investigation of companies providing direct-to-consumer genetic tests, the US Food and Drug Administration is considering premarket review of some laboratory-derived tests that pose higher clinical risks, assuring that the tests are evaluated for analytical and clinical validity (20).

Clinical Validity: Clinical validity refers to test accuracy and reliability in predicting risk of T2D (discrimination and calibration).

Discrimination shows how well the model can distinguish between individuals with and without disease. A commonly used measure of discrimination is the area under the receiver operating characteristic curve (AUC). AUC can vary from 0.5 (equal to tossing a coin) to 1 (perfect discrimination). AUC indicates the probability that, on average, an individual with the disease will be assigned a higher predicted risk than an individual without the disease. Calibration indicates how close the risks predicted by the model are to the actual observed risks. The Hosmer-Lemeshow (H-L) chi-square test is a commonly used summary measure of calibration. The H-L test compares the observed and predicted number of patients within specified risk groups, usually deciles of risk.

In most empirical studies, the genetic risk scores had lower discriminative accuracy than the clinical risk factors (21-23). Furthermore, addition of genetic factors to the clinical risk factors either did not change or only marginally improved the AUC beyond the clinical risk models. Like companies, all studies used multiplicative models or additive genetic effects (24-26), but whether this is correct has not been demonstrated. Besides, none of these studies investigated the same panel of SNPs as the companies do. Disagreement between results for identical DNA samples sent to 4 different companies reflects the use of different sets of markers to predict risk of disease and the use of different average risks to depart from (20, 27). **Table 3** presents an overview of the published studies conducted on T2D risk so far, mostly in European populations. **Table 4** shows the SNPs included in genetic risk scores in the studies summarized in **Table 1** and the SNPs used by three commercial companies to predict T2D risk. The other companies do not specify on their websites which SNPs they use for T2D risk prediction. For most companies, algorithms or criteria for interpreting SNP results are not made clear to the consumer. Even when this information is made available (24-26), it is sometimes difficult to know which effect sizes and genotype frequencies are used to calculate a composite risk (28).

Table 3. Genetic risk prediction studies in type 2 diabetes.

Study reference	Design	Sample size	Clinical risk factors	No of polymorphisms	AUC clinical (95% CI)	AUC genetic (95% CI)	AUC combined (95% CI)
European							
Weedon et al. (29)	Case-control	2409 T2D cases / 3668 controls	NA	3	NA	0.58	NA
Lyssenko et al. (30, 31)	Prospective cohort	2293	BMI, FPG	3	0.68 (0.63 to 0.73)	NA	0.68 (0.63 to 0.73)
Vaxillaire et al. (32)	Prospective cohort	3877	Age, sex, BMI	3	0.82	0.56	0.83
Cauchi et al. (33)	Case-control	4232 T2D cases / 4595 controls	Age, sex, BMI	15	NA	NA	0.86
van Hoek et al. (34)	Prospective cohort	6544	Age, sex, BMI	18	0.66 (0.63 to 0.68)	0.60 (0.57 to 0.63)	0.68 (0.66 to 0.71)
Lango et al. (35)	Case-control	2309 T2D cases / 2598 controls	Age, sex, BMI	18	0.78	0.60	0.80
Lyssenko et al. (36)	Prospective cohorts	1) Malmö study: 16,061	Age, sex, BMI, FH, BP, TG, FPG (in Malmö study), plus HDL and waist circumference (in Botnia study)	11	1) 0.743	1) 0.626	1) 0.753
		2) Botnia study: 2770			2) 0.786	2) 0.682	2) 0.801
Meigs et al. (37)	Prospective cohort	2377	Age, sex, BMI, FH, FPG, SBP, HDL, TG	18	0.900 (0.880 to 0.919)	0.581 (0.546 to 0.617) (adjusted for sex)	0.901 (0.881 to 0.920)
Balkau et al. (38)	Prospective cohort	1) Men: 1863	Current smoker (in men), waist circumference, HT, FPG, GGT, and FH and BMI (in women)	2	1)0.850	NA	1) 0.851
		2) Women: 1954			2) 0.917		2) 0.912

Study reference	Design	Sample size	Clinical risk factors	No of polymorphisms	AUC clinical (95% CI)	AUC genetic (95% CI)	AUC combined (95% CI)
European							
Lin et al. (39)	Cross-sectional	356 T2D cases / 5004 controls	Age, sex, FH, physical activity, triacylglycerol/HDL ratio, waist-hip ratio	15	0.86	0.59	0.87
Sparso et al. (40)	Case-control	4093 T2D cases / 5302 controls	Age, sex, BMI	19	0.92	0.60	0.93
Schulze et al. (41)	Prospective case-cohort	579 T2D cases / 1962 controls	Age, waist circumference, height, history of HT, physical activity, smoking, consumption of red meat, whole-grain bread, coffee and alcohol, glucose, HbA1c, TG, HDL, GGT, ALT, hs-CRP	20	0.90 (0.89 to 0.92)	NA	0.90 (0.89 to 0.91)
Cornelis et al. (42)	Nested case-control	2768 T2D cases / 3447 controls	Age, sex, BMI, FH, smoking, alcohol intake, physical activity	10	0.78 (0.77 to 0.79)	NA	0.79 (0.78 to 0.80)
Talmud et al. (43)	Prospective cohort	5135	1) Cambridge T2D risk score: age, sex, BMI, drug treatment, FH, smoking status	20	1) 0.72 (0.69 to 0.76)	0.55 (0.51 to 0.59)	1) 0.73 (0.69 to 0.76)
			2) Framingham offspring T2D risk score: age, sex, BMI, parental history of T2D, HDL, TG, FPG		2) 0.78 (0.75 to 0.82)		2) 0.78 (0.74 to 0.81)
Fontaine-Bisson et al. (44)	Cross-sectional	1327 T2D cases / 1424 controls	Age, sex	17	NA	NA	0.591

Study reference	Design	Sample size	Clinical risk factors	No of polymorphisms	AUC clinical (95% CI)	AUC genetic (95% CI)	AUC combined (95% CI)
European							
Wang et al. (45)	Cross-sectional	518 T2D cases / 6714 controls	1) Finnish Diabetes Risk Score (FINDRISC): age, BMI, waist circumference, antihypertensive medication, physical activity, previously known high glucose, FH, diet	19	1) 0.727 (0.705 to 0.749)	0.552 (0.526 to 0.578)	1) 0.730 (0.708 to 0.753)
			2) FINDRISC, TG, HDL, adiponectin, ALT		2) 0.772 (0.752 to 0.793)		2) 0.772 (0.751 to 0.793)
de Miguel-Yanes et al. (46)	Prospective cohort	3471 (11 358 person-observations)	1) Sex, FH, BMI, FPG, SBP, HDL, TG; stratified by age	40	1) <50 yrs: 0.908 (0.884 to 0.932) ≥50 yrs: 0.883 (0.863 to 0.903)	1) <50 yrs: 0.657 (0.611 to 0.703) ≥50 y: 0.590 (0.557 to 0.623)	1) <50 yrs: 0.911 (0.887 to 0.935) ≥50 yrs: 0.884 (0.865 to 0.904)
			2) Age, sex, FH, BMI, FPG, SBP, HDL, TG		2) 0.903 (0.889 to 0.917)	2) 0.606 (0.579 to 0.632) (adjusted for sex)	2) 0.906 (0.892 to 0.920)
Asian							
Miyake et al. (47)	Case-control	2316 cases / 2370 controls	Age, sex, BMI	11	0.68	0.63	0.72
Hu et al. (48)	Case-control	1849 T2D cases / 1785 controls	Age, sex, BMI	11	0.614 (0.595 to 0.632)	0.621 (0.604 to 0.639)	0.668 (0.650 to 0.685)

ALT, alanine aminotransferase; BMI, body mass index; BP, blood pressure; FH, family history of T2D; FPG, fasting plasma glucose; GGT, γ-glutamyltransferase; HDL, high-density lipoprotein cholesterol; hs-CRP, high sensitivity c-reactive protein; HT, hypertension; NA, not available; SBP, systolic blood pressure; TG, triglycerides.
Search strategy: We performed a search in PubMed and HuGE Navigator to identify relevant studies, scanned the reference lists from the retrieved articles to identify additional studies, and further used Web of Science to identify studies that cited the selected articles. The specific queries used are provided under the heading **Links**.

Table 4. Single nucleotide polymorphisms tested in risk prediction studies and used by commercial companies to predict type 2 diabetes risk.

		Direct-to-Consumer Companies									
Locus	SNPs	**deCODEme (European ancestry)**	**deCODEme (East Asian ancestry)**	**deCODEme (African ancestry)**	**23andme**	**Navigenics**	**Weedon et al.[29]**	**Lyssenko et al.[30, 31]**	**Vaxillaire et al.[32]**	**Cauchi et al.[33]**	**van Hoek et al.[34]**
TCF7L2	rs7903146	X	X	X	X	X[1]	X	-	X	X	X
TCF7L2	rs7901695[2]	-	-	-	-	-	-	-	-	-	-
TCF7L2	rs12255372[2]	-	-	-	-	-	-	-	-	-	-
PPARG	rs1801282	X	-	-	X	X	X	X	-	-	X
CDKN2A/B	rs10811661	X	X[1]	-	X[1]	X	-	-	-	X	X
CDKN2A/B	rs564398[4]	-	-	-	-	-	-	-	-	-	-
CDKN2A/B	rs1412829[4]	-	-	-	-	-	-	-	-	-	X
KCNJ11	rs5219	X[1]	X[1]	-	X	X[1]	X	-	-	-	X
IGF2BP2	rs4402960	X	X	-	X	X	-	-	-	X	X
SLC30A8	rs13266634	X	X	-	X	X	-	-	-	X	X
HHEX-IDE-KIF11	rs1111875	X	X	-	X	X	-	-	-	-	X
HHEX-IDE-KIF11	rs7923837[2]	-	-	-	-	-	-	-	-	X	-
HHEX-IDE-KIF11	rs10748582[2]	-	-	-	-	-	-	-	-	-	-
WFS1	rs10010131	X	-	-	X[1]	X	-	-	-	-	X[1]
NOTCH2/ADAM30	rs10923931	X[1]	-	-	-	X	-	-	-	-	X[1]
JAZF1	rs864745	X	-	-	-	X	-	-	-	-	X[1]
THADA	rs7578597	X	-	-	-	-	-	-	-	-	X
TSPAN8/LGR5	rs7961581	X	-	-	-	X	-	-	-	-	X[1]
TCF2/HNF1B	rs4430796	X	X	X	-	X	-	-	-	-	X
TCF2/HNF1B	rs757210[2]	-	-	-	-	-	-	-	-	-	-
TCF2/HNF1B	rs7501939[2]	-	-	-	-	-	-	-	-	-	-
CDKAL1	rs7756992	X	X	-	-	X	-	-	-	X	-
CDKAL1	rs7754840[2]	-	-	-	-	-	-	-	-	-	X
CDKAL1	rs10946398[2]	-	-	-	-	-	-	-	-	-	-
CDKAL1	rs4712523[2]	-	-	-	X	-	-	-	-	-	-
CDKAL1	rs17036101[4]	-	-	-	-	-	-		-	-	-
KCNQ1	rs2237892	X	X	-	-	X[1]	-	-	-	-	-
KCNQ1	rs2237895[4]	-	-	-	-	-	-	-	-	-	-
KCNQ1	rs231362[4]	-	-	-	-	-	-	-	-	-	-
MTNR1B	rs10830963	X	-	-	-	-	-	-	-	-	-
GCK	rs4607517	X	-	-	-	-	-	-	X[1]	-	-
GCKR	rs780094	X	-	-	-	-	-	-	-	-	-
DGKB/TMEM195	rs2191349	X[1]	-	-	-	-	-	-	-	-	-

Variants Evaluated													
Published Studies													
Lango et al.[35]	**Lyssenko et al.**[36]	**Meigs et al.**[37]	**Miyake et al.**[47]	**Balkau et al.**[38]	**Hu et al.**[48]	**Lin et al.**[39]	**Sparso et al.**[40]	**Schulze et al.**[41]	**Cornelis et al.**[42]	**Talmud et al.**[43]	**Fontaine-Bisson et al.**[44]	**Wang et al.**[45]	**Miguel-Yanes et al.**[46]
X	X	X	X	X	-	-	X	X	-	-	X	X	X
-	-	-	-	-	-	X	-	-	-	X	-	-	-
-	-	-	-	-	-	-	-	-	X	-	-	-	-
X	X	X	X	-	X	X	X	X	X	X	X	X	X
X	-	X	X	-	X	X	X	X	X	X	X	X	X
X	-	-	-	-	-	-	-	-	X	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
X	X	X	X	-	X	X[1]	X[1]	X	X	X[1]	X	X	X[1]
X	X	X[1]	X[1]	-	X[1]	X	X	X	X	X	-	X	X[1]
X	X	X	X[1]	-	X	-	X	X	X	X	X	X	X
X	X	X	-	-	-	X	X	X	X	X	X	X	X
-	-	-	X	-	-	-	-	-	-	-	X	-	-
-	-	-	-	-	X	-	-	-	-	-	-	-	-
X	X	-	-	-	X	X	X	X	X	-	X	X	X
X[1]	X	X	-	-	-	X	X	X	-	X	X	X	X
X	X	X	-	-	-	X	X	X	-	X	-	X	X
X	-	X	-	-	-	X	X	X	-	X	X	X	X
X	-	X	-	-	-	X	X	X	-	X	X	X	X
-	-	-	X	-	X	-	-	X	-	X	X	-	-
X	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	X	-	-	-	-	-	-
-	-	-	-	-	X	-	X	-	X	-	-	-	-
-	X	X	X	-	-	-	-	X	-	-	-	X	X
X	-	-	-	-	-	X	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	X	-	-	-
-	-	-	X	-	X		-	-	-	-	-	X[1]	X
-	-	-	-	-	-	-	X	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	X	-	-	-	-	X	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	X	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X

		Direct-to-Consumer Companies									
Locus	SNPs	deCODEme (European ancestry)	deCODEme (East Asian ancestry)	deCODEme (African ancestry)	23andme	Navigenics	Weedon et al.[29]	Lyssenko et al.[30, 31]	Vaxillaire et al.[32]	Cauchi et al.[33]	van Hoek et al.[34]
ADCY5	rs2877716	X	-	-	-	-	-	-	-	-	-
PROX1	rs340874	X	-	-	-	-	-	-	-	-	-
FTO	rs8050136	-	-	-	-	X	-	-	-	-	X
ADAMTS9	rs4607103	-	-	-	-	X	-	-	-	-	X[1]
CDC123/CAMK1D	rs12779790	-	-	-	-	-	-	-	-	-	-
CDC123/CAMK1D	rs11257622 [2]	-	-	-	-	-	-	-	-	-	X
VEGFA	rs9472138	-	-	-	-	-	-	-	-	-	-
DCD	rs1153188	-	-	-	-	-	-	-	-	-	-
BCL11A	rs10490072	-	-	-	-	-	-	-	-	-	-
BCL11A	rs243021 [4]	-	-	-	-	-	-	-	-	-	-
LOC387761	rs7480010	-	-	-	-	-	-	-	-	X	-
CAPN10	rs3792267	-	-	-	-	-	-	X	-	-	-
CAPN10	rs2975760 [3]	-	-	-	-	-	-	X	-	-	-
EXT2	rs3740878	-	-	-	-	-	-	-	-	X	-
IL6	rs1800795	-	-	-	-	-	-	-	X	-	-
chr11.41871942	rs9300039	-	-	-	-	X	-	-	-	-	-
LOC441171	rs9494266	-	-	-	-	X	-	-	-	-	-
CAMTA1	rs1193179	-	-	-	-	-	-	-	-	X	-
CXCR4	rs932206	-	-	-	-	-	-	-	-	X	-
INS	rs689	-	-	-	-	-	-	-	-	-	-
KCTD12	rs2876711	-	-	-	-	-	-	-	-	X	-
LDLR	rs6413504	-	-	-	-	-	-	-	-	X	-
LOC646279	rs1256517	-	-	-	-	-	-	-	-	X	-
MMP26	rs2499953	-	-	-	-	-	-	-	-	X	-
NGN3	rs10823406	-	-	-	-	-	-	-	-	X	-
HNF1A	rs1800574	-	-	-	-	-	-	-	-	-	-
CENTD2	rs1552224	-	-	-	-	-	-	-	-	-	-
HCCA2	rs2334499	-	-	-	-	-	-	-	-	-	-
HMGA2	rs1531343	-	-	-	-	-	-	-	-	-	-
KIAA1486	rs7578326	-	-	-	-	-	-	-	-	-	-
KLF14	rs972283	-	-	-	-	-	-	-	-	-	-
OASL/TCF1	rs7957197	-	-	-	-	-	-	-	-	-	-
PRC1	rs8042680	-	-	-	-	-	-	-	-	-	-
RBMS1/ITGB6	rs7593730	-	-	-	-	-	-	-	-	-	-

Variants Evaluated														
Published Studies														
Lango et al.[35]	**Lyssenko et al.[36]**	**Meigs et al.[37]**	**Miyake et al.[47]**	**Balkau et al.[38]**	**Hu et al.[48]**	**Lin et al.[39]**	**Sparso et al.[40]**	**Schulze et al.[41]**	**Cornelis et al.[42]**	**Talmud et al.[43]**	**Fontaine-Bisson et al.[44]**	**Wang et al.[45]**	**Miguel-Yanes et al.[46]**
-	-	-	-	-	-	-	-	-	-	-	-	-	X[1]
-	-	-	-	-	-	-	-	-	-	-	-	-	X
X	X[1]	-	-	-	X	X	X	X	-	X	-	X[1]	X[1]
X	-	X	-	-	-	X	X	X	-	X	-	X	X
X	-	X	-	-	-	X	X	X	-	X	X	X	X
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	X	-	-	-	-	-	X	-	X	X	-	X
-	-	X	-	-	-	-	-	X	-	-	X	-	X
-	-	X	-	-	-	-	-	X	-	X	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	X	X	-
-	-	-	-	-	-	-	-	-	-	X	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	X[1]	-	-
-	-	-	-	X	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	X	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	X	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X

		Direct-to-Consumer Companies									
Locus	SNPs	**deCODEme (European ancestry)**	**deCODEme (East Asian ancestry)**	**deCODEme (African ancestry)**	**23andme**	**Navigenics**	**Weedon et al.[29]**	**Lyssenko et al.[30, 31]**	**Vaxillaire et al.[32]**	**Cauchi et al.[33]**	**van Hoek et al.[34]**
TLE4	rs13292136	-	-	-	-	-	-	-	-	-	-
TP53INP1	rs896854	-	-	-	-	-	-	-	-	-	-
ZBED3	rs4457053	-	-	-	-	-	-	-	-	-	-
ZFAND6	rs11634397	-	-	-	-	-	-	-	-	-	-

SNPs, single nucleotide polymorphisms. [1]Another variant in perfect linkage disequilibrium ($R^2 > 0.90$) was used; [2]This variant is in high LD ($R^2 \leq 0.90$ and > 0.60) with the reference variant; [3]This variant is in low ($R^2 \leq 0.60$ and > 0.05) with the reference variant; [4]This variant has an $R^2 \leq 0.05$ with the reference variant. The first column lists all SNPs included in genetic tests for T2D, either used by DTC companies or available from published studies. "X" denotes that the SNP was included in the genetic risk model and "-" denotes that the SNP was not included.

Another important aspect when testing the performance of a prediction model is the model calibration. Measures of calibration were presented in some of the T2D risk prediction studies and generally showed sufficient model fit (36-39, 41, 43, 45, 46).

*Clinical Utility***:** Net benefit of test in improving health outcomes.

We assessed clinical utility as the added benefit of the test beyond traditional clinical predictors in improving health outcomes, and as the impact of genetic testing on attitudes, beliefs and health related behavior in individuals who receive genetic risk information.

First, clinical utility is reflected in the impact of a risk prediction model on the classification of individuals in risk groups for which the preventive interventions differ. Percentage of reclassification and the net reclassification improvement (NRI) are recently developed measures that assess this aspect of clinical utility. Reclassification is the percentage of individuals that change from one risk category based on the original prediction model to a different risk category based on the updated model. NRI separately considers the reclassification in cases and non-cases. Cases are correctly classified when they move to a higher risk category and wrongly classified when they move to a lower category. Non-cases move correctly to a lower category and wrongly to a higher. NRI is the sum of the net correct moves: the proportion of cases moving up minus the proportion of cases moving down, plus the proportion of non-cases moving down minus the proportion of non-cases moving up (49). **Table 5** shows the amount of reclassification resulted from the addition of genetic information to clinical data in T2D risk prediction, either directly

Variants Evaluated													
Published Studies													
Lango et al.[35]	**Lyssenko et al.[36]**	**Meigs et al.[37]**	**Miyake et al.[47]**	**Balkau et al.[38]**	**Hu et al.[48]**	**Lin et al.[39]**	**Sparso et al.[40]**	**Schulze et al.[41]**	**Cornelis et al.[42]**	**Talmud et al.[43]**	**Fontaine-Bisson et al.[44]**	**Wang et al.[45]**	**Miguel-Yanes et al.[46]**
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	X

reported in the original studies or calculated from reclassification tables available from original papers (23). Since most genetic risk prediction studies in T2D have been performed in European populations (see **Table 3**) it is impossible to generalize the performance of the genetic tests to populations with different ancestry. Furthermore, the incidence rates of T2D vary even within European ancestry groups. As a result, no clinically defined risk categories exist that can be applied across different populations where the underlying risk of T2D varies and, therefore, the cut-off values chosen to define the risk groups differ among studies. This is an important aspect in the interpretation of reclassification measures, as the choice of cut-off has a high impact on the percentage of reclassification observed (23). In consequence, the assessment of NRI in the absence of clinically estimated cut-offs is of limited value.

Table 5. Reclassification measures from genetic risk prediction studies in type 2 diabetes.

Study reference	Updated model	Cut-off values	NRI, % (P value)	Reclassification, %
Meigs et al. (23, 37)	Most complete clinical model plus 18 SNPs	2 and 8%	2.13 (0.17)	6.28
Lyssenko et al.(23, 36)	Most complete clinical model plus 11 SNPs	10 and 20%		
1)Malmö study			1) 4.5 ($2.5x10^{-5}$)	1) 15.59
2)Botnia study			2) 8.79 (0.13)	2) 9.44
Talmud et al. (43)	1) Cambridge risk score plus 20 SNPs	5, 10 and 15%	1) 4.6 (0.17)	1) 16
	2) Framingham risk score plus 20 SNPs		2) -3.2 (0.35)	2) 12.3
de Miguel-Yanes et al. (46)	Most complete clinical model plus 40 SNPs	2 and 8%	1) <50 yrs: 10.2 (0.001) 2) ≥50 yrs: 0.4 (0.7) 3) All: 1.8 (0.2)	NA

NA, not assessed.

Second, when the impact on outcome prediction is not available, clinical utility is reflected in the public interest and health care provider interest in genetic testing, the uptake of the tests and the effect of testing on outcomes such as adherence to lifestyle changes or to medication for prevention and treatment of disease.

A survey conducted among primary care physicians and endocrinologists (n=304) and patients (152 non-diabetic and 89 with T2D) assessed beliefs regarding the clinical use of genetic testing for T2D. Subjects answered questions related to three domains: testing for risk prediction, testing to motivate behavior change and testing to guide medication prescription. Most physicians (88%) and patients (79%) were in favor of genetic testing in general. However, patients were more likely than physicians to request genetic testing for risk prediction and treatment guidance. Patients, and to a lesser extent physicians, expressed expectations that knowledge of genetic risk would motivate adoption of preventive lifestyle recommendations and increase adherence to treatment (50).

We identified four registered clinical trials (see **Links** for search strategy) that aim to assess the impact of genetic testing on risk perception and behavior change in patients with T2D:

- **Genetic Counseling and Lifestyle Change for Diabetes Prevention (GC/LC):** "This study will examine the impact of diabetes genetic counseling on patient motivation and disease prevention behaviors among subjects with pre-diabetes. Intervention subjects will be provided with their individual diabetes genotype risk score derived from aggregating the combined results of 37 diabetes risk-associated genetic loci. Controls will not be tested. All subjects will be enrolled in a 12-week diabetes prevention program." (ClinicalTrials.gov identifier: NCT01034319)
- **The Impact of Genetic Testing for Type 2 Diabetes on Health Behaviors:** "We will evaluate the impact of genetic testing for type 2 diabetes on psychological, health behavior, and clinical outcomes." (ClinicalTrials.gov Identifier NCT01060540) The genetic test consists of SNPs in the *TCF7L2*, *PPARG* and *KCNJ11* genes.
- **Effect of Type 2 Diabetes Genetic Risk Information on Health Behaviors and Outcomes (TDE):** "The primary objective of the study is to assess the clinical utility of a genetic test for Type 2 diabetes risk in combination with standardized risk assessment compared with standardized risk assessment alone." (ClinicalTrials.gov Identifier NCT00849563) Variants not specified.
- **Predictive Genetic Risk Assessment Trial (PGT):** "A critical goal of this clinical trial is to understand how individual patients and their doctors perceive and respond to genetic risk information that is largely uncertain." (ClinicalTrials.gov Identifier NCT00782366) Variants not specified.

METHODS

To identify published reviews, recommendations and guidelines on genetic testing for T2D risk prediction we searched: the Agency for Healthcare Research and Quality (AHRQ), the Cochrane Collaboration, the US Preventive Task Force, the Evaluation of Genomic Applications in Practice and Prevention Working Group, the National Institute for Health and Clinical Excellence, the NHS Evidence – National Library of Guidelines; the Canadian Medical Association Infobase: Clinical Practice Guidelines, the European Society for Human Genetics. To retrieve information about companies that offer DTC genetic testing for T2D risk prediction we performed a search in Google, followed the list of companies from a published review on DTC genomic companies (27) and collected additional information from discussions with other researchers.

Links

- **PubMed:** type 2 diabetes AND (genetic markers OR risk polymorphisms OR genetic score* OR susceptibility variants OR genetic risk factors OR genetic testing OR genotype score) AND (risk assessment OR disease prediction OR risk prediction OR discriminative value OR ROC curve)
- **HuGE Navigator:** type 2 diabetes(Text+MeSH)>>Gene-gene interactions, Genetic testing(Category)
- **ClinicalTrials.gov:** type 2 diabetes AND genetic tests | type 2 diabetes
- **U.S. Food and Drug Administration:** No information indentified

REFERENCES

1. Paulweber B, Valensi P, Lindstrom J, Lalic NM, Greaves CJ, McKee M, et al. **A European evidence-based guideline for the prevention of type 2 diabetes.** *Horm Metab Res* 2010;**42 Suppl 1:**S3-36.
2. deCODEme. (cited 2010 November 16); Available from: http://www.decodeme.com
3. deCODEme. Type 2 Diabetes. (cited 2010 November 16); Available from: http://demo.decodeme.com/health-watch/details/T2D
4. deCODEhealth. Type 2 Diabetes. (cited 2010 November 16); Available from: http://health.decode.is/type_2_diabetes.php
5. Pathway Genomics. (cited 2010 November 16); Available from: http://www.pathway.com/dna-reports/full-list-of-conditions
6. 23andMe. Type 2 Diabetes. (cited 2010 November 16); Available from: https://www.23andme.com/health/Type-2-Diabetes/
7. Navigenics. Type 2 diabetes. (cited 2010 December 20); Available from: http://www.navigenics.com/demo/your_dna/d/type_2_diabetes
8. GeneticHealth. Obesity test. (cited 2010 November 16); Available from: http://www.genetic-health.co.uk/dna-test-services/obesity-test.htm
9. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. **Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis.** *Nat Genet* 2010;**42**(7)**:**579-89.
10. **Screening for type 2 diabetes mellitus in adults: U.S. Preventive Services Task Force recommendation statement.** *Ann Intern Med* 2008;**148**(11)**:**846-54.
11. International Diabetes Federation. (cited 2010 September 1); Available from: http://www.eatlas.idf.org
12. Gillies CL, Abrams KR, Lambert PC, Cooper NJ, Sutton AJ, Hsu RT, et al. **Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis.** *BMJ* 2007;**334**(7588)**:**299.
13. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. **Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin.** *N Engl J Med* 2002;**346**(6)**:**393-403.
14. CyGeneDirect. Knowing your risks. (cited 2010 December 20); Available from: http://www.cygenedirect.com/browse-11802/Knowing-Your-Risks.html
15. Navigenics. Physician whitepaper. (cited 2010 December 20); Available from: http://www.navigenics.com/static/pdf/physician/physician-whitepaper.pdf
16. deCODEme. SNP measurement. (cited 2010 December 20); Available from: http://demo.decodeme.com/health-watch-information/snp-measurement
17. 23andMe. Science. (cited 2010 December 20); Available from: https://www.23andme.com/more/science/
18. PathwayGenomics. Terms of service. (cited 2010 December 20); Available from: http://www.pathway.com/about-us/terms-of-service
19. CyGeneDirect. DNA standards. (cited 2010 December 20); Available from: http://www.cygenedirect.com/browse-10494/Dna-Standards.html
20. Kuehn BM. **Inconsistent results, inaccurate claims plague direct-to-consumer gene tests.** *JAMA* 2010; **304**(12)**:**1313-5.

21. Ruchat SM, Vohl MC, Weisnagel SJ, Rankinen T, Bouchard C, Perusse L. **Combining genetic markers and clinical risk factors improves the risk assessment of impaired glucose metabolism.** *Ann Med* 2010;**42**(3)**:** 196-206.

22. Grant RW, Moore AF, Florez JC. **Genetic architecture of type 2 diabetes: recent progress and clinical implications.** *Diabetes Care* 2009;**32**(6)**:**1107-14.

23. Mihaescu R, van Zitteren M, van Hoek M, Sijbrands EJ, Uitterlinden AG, Witteman JC, et al. **Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve.** *Am J Epidemiol* 2010;**172**(3)**:**353-61.

24. 23andMe. How it works. (cited 2010 December 20); Available from: https://www.23andme.com/howitworks/

25. deCODEme. Risk calculation. (cited 2010 December 20); Available from: http://demo.decodeme.com/health-watch-information/risk-calculation

26. Navigenics. The science. (cited 2010 December 20); Available from: http://www.navigenics.com/static/pdf/Navigenics-TheScience.pdf

27. Swan M. **Multigenic condition risk assessment in direct-to-consumer genomic services.** *Genet Med* 2010; **12**(5)**:**279-88.

28. 23andMe. Type 2 Diabetes, technical report. (cited 2010 December 20); Available from: https://www.23andme.com/health/Type-2-Diabetes/techreport/

29. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, et al. **Combining information from common type 2 diabetes risk polymorphisms improves disease prediction.** *PLoS medicine* 2006;**3**(10)**:**e374.

30. Lyssenko V, Almgren P, Anevski D, Orho-Melander M, Sjogren M, Saloranta C, et al. **Genetic prediction of future type 2 diabetes.** *PLoS medicine* 2005;**2**(12)**:**e345.

31. Janssens AC, Gwinn M, Khoury MJ, Subramonia-Iyer S. **Does genetic testing really improve the prediction of future type 2 diabetes?** *PLoS medicine* 2006;**3**(2)**:**e114; author reply e27.

32. Vaxillaire M, Veslot J, Dina C, Proenca C, Cauchi S, Charpentier G, et al. **Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study.** *Diabetes* 2008;**57**(1)**:**244-54.

33. Cauchi S, Meyre D, Durand E, Proenca C, Marre M, Hadjadj S, et al. **Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value.** *PloS one* 2008;**3**(5)**:**e2031.

34. van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, et al. **Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study.** *Diabetes* 2008;**57**(11)**:**3122-8.

35. Lango H, Consortium UKTDG, Palmer CN, Morris AD, Zeggini E, Hattersley AT, et al. **Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk.** *Diabetes* 2008;**57**(11)**:**3129-35.

36. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, et al. **Clinical risk factors, DNA variants, and the development of type 2 diabetes.** *N Engl J Med* 2008;**359**(21)**:**2220-32.

37. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. **Genotype score in addition to common risk factors for prediction of type 2 diabetes.** *N Engl J Med* 2008;**359**(21)**:**2208-19.

38. Balkau B, Lange C, Fezeu L, Tichet J, de Lauzon-Guillain B, Czernichow S, et al. **Predicting diabetes: clinical, biological, and genetic approaches: data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR).** *Diabetes Care* 2008;**31**(10)**:**2056-61.

39. Lin X, Song K, Lim N, Yuan X, Johnson T, Abderrahmani A, et al. **Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score–the CoLaus Study.** *Diabetologia* 2009;**52**(4)**:**600-8.

40. Sparso T, Grarup N, Andreasen C, Albrechtsen A, Holmkvist J, Andersen G, et al. **Combined analysis of 19 common validated type 2 diabetes susceptibility gene variants shows moderate discriminative value and no evidence of gene-gene interaction.** *Diabetologia* 2009;**52**(7)**:**1308-14.

41. Schulze MB, Weikert C, Pischon T, Bergmann MM, Al-Hasani H, Schleicher E, et al. **Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study.** *Diabetes Care* 2009;**32**(11)**:**2116-9.

42. Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, et al. **Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry.** *Ann Intern Med* 2009;**150**(8)**:**541-50.

43. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, et al. **Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study.** *BMJ* 2010; **340:**b4838.

44. Fontaine-Bisson B, Renstrom F, Rolandsson O, Payne F, Hallmans G, Barroso I, et al. **Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population.** *Diabetologia* 2010.

45. Wang J, Stancakova A, Kuusisto J, Laakso M. **Identification of undiagnosed type 2 diabetic individuals by the Finnish diabetes risk score and biochemical and genetic markers: a population-based study of 7232 Finnish men.** *J Clin Endocrinol Metab* 2010;**95**(8)**:**3858-62.

46. de Miguel-Yanes JM, Shrader P, Pencina MJ, Fox CS, Manning AK, Grant RW, et al. **Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms.** *Diabetes Care* 2010.

47. Miyake K, Yang W, Hara K, Yasuda K, Horikawa Y, Osawa H, et al. **Construction of a prediction model for type 2 diabetes mellitus in the Japanese population based on 11 genes with strong evidence of the association.** *J Hum Genet* 2009;**54**(4)**:**236-41.

48. Hu C, Zhang R, Wang C, Wang J, Ma X, Lu J, et al. **PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population.** *PloS one* 2009;**4**(10)**:**e7643.

49. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. **Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond.** *Stat Med* 2008;**27**(2)**:**157-72; discussion 207-12.

50. Grant RW, Hivert M, Pandiscio JC, Florez JC, Nathan DM, Meigs JB. **The clinical application of genetic testing in type 2 diabetes: a patient and physician survey.** *Diabetologia* 2009;**52**(11)**:**2299-305.

Chapter 12

Summary and general discussion

SUMMARY

This thesis describes methodological and empirical studies of genetic risk prediction of common diseases. The *methodological* studies involved the evaluation of traditional and new methods of model performance, the evaluation of rare variants for risk prediction of common diseases, the assessment of simulation strategies for replication of empirical data, and the evaluation of reporting in empirical risk prediction studies. The *empirical* studies involved the evaluation of direct-to-consumer (DTC) genetic tests, the external validation of genetic risk prediction models across European samples, and the evaluation of differences in prediction performance in published genetic risk prediction models.

Both simulated and empirical data were used to conduct these studies. By means of simulation, a population distribution of genetic variants was created starting from the effect size and frequency of individual variants and assuming that genetic variants are inherited independently and that their joint effects follow a multiplicative risk model. The empirical data came from the Rotterdam study, a prospective, population-based, cohort study among 7,983 inhabitants of a Rotterdam suburb, designed to investigate determinants of chronic diseases; and from a community-based cohort i.e., the combination of Atherosclerosis Risk in Communities Study, Cardiovascular Health Study and Framingham Heart Study.

SUMMARY OF STUDIES PRESENTED IN THIS THESIS

Reclassification measures versus the area under the ROC curve (Chapter 2)

Reclassification measures that assess improvement in risk prediction model performance have recently received much attention. They are considered more sensitive in detecting small increments in model improvement than traditional measures. However, it was not determined yet whether these measures should be used as an additional step after the traditional measures of model improvement, like the area under the ROC curve (AUC), or should replace them. To answer this question, percentages of total reclassification, net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were calculated for different improvements in the AUC, starting from the same baseline AUC. This study revealed three characteristics of reclassification measures.

First, it demonstrated that the percentages of reclassification varied with the cutoff values used to define the risk categories. For example, for a small increase in the AUC, when data from the Rotterdam study was used, the percentage of total reclassification ranged from 0 to 23%, depending on what cutoff value was chosen.

Second, and related to the previous, more reclassification was observed when the threshold that defined the risk categories was set around the center of the risk distribution. Thus, studies

using arbitrary selection of risk categories can show significant reclassification simply because many individuals have a predicted risk around the threshold used to define the risk categories.

Third, this study also demonstrated that reclassification observed in the absence of any AUC improvement implies that the updated model did not make fewer errors, but simply made different errors than the initial model.

Because reclassification varies with the cutoff thresholds, reclassification should only be used if meaningful categories exist. Selection of the cutoff point should not be based on the maximum amount of reclassification. The choice of the optimal cutoff value should ultimately be determined by weighing the harms and benefits of false positive and false negative decisions. Furthermore, the number of risk categories should be determined by the number of available preventive treatment strategies. To conclude, the AUC and reclassification measures provide complementary information about the improvement of risk prediction; the AUC is a measure of discrimination and reclassification is a measure of clinical usefulness.

Selection of high-risk groups based on genetic data (Chapter 3)

Classification of individuals in high-risk groups based on their genetic profiles is considered a potential novel strategy for targeting preventive intervention for common diseases (1). Because in many instances no thresholds for the definition of the risk groups exist, in the research setting arbitrary risk thresholds are used to evaluate the clinical utility of genetic tests, as discussed in Chapter 2 of this thesis. How the choice of risk thresholds determines the performance of risk classification, i.e. the sensitivity, specificity, and positive and negative predictive value of the test, and what other epidemiological parameters are relevant is currently unknown. Using simulated data, illustrated with an example on age-related macular degeneration (AMD) risk, this study assessed how sensitivity, specificity, positive and negative predictive values of genetic risk models change along the range of threshold values that can be chosen to define at-risk groups. This study showed that for low risk thresholds the sensitivity was high and the positive predictive value low, while for high risk thresholds the sensitivity was low and the positive predictive value was high. Low risk thresholds corresponded to risk-groups with a frequency higher than the disease frequency, and high risk thresholds to risk-groups with a frequency lower than the disease frequency. Only when the frequency of the at-risk group and the disease frequency were equal, positive predictive value and sensitivity were equal and they increased with increasing AUC of the risk model. To conclude, the frequency of the at-risk group relative to the frequency of disease in the population has a strong influence on the performance of the risk classification strategy. This relationship dictates if the selection of individuals at high-risk will be an effective public health strategy. By replicating the results in the AMD risk prediction

example, where the risks were calculated from logistic regression analyses, this study showed that this relationship is generalizable beyond the simple count scores to more common genetic risk prediction models. These findings are relevant for research focusing on the assessment of novel biomarkers in addition to established risk factors, a topic that currently receives much attention, as they provide insight into the potential clinical utility of the biomarkers.

Prediction based on meta-analysis data using simulations strategies (Chapter 4)

The conduction of empirical studies on the predictive ability of genetic risk models, as assessed by the AUC, is costly and their results have often shown moderate predictive ability so far. It is of interest to investigate whether the AUC of empirical studies can be replicated by simulation studies using published odds ratios (ORs) and genotype frequencies from genome-wide association studies (GWASs), published prior to the empirical studies. When simulations can accurately replicate these AUC values, simulations studies can be used to indicate whether empirical evaluation of a genetic risk model is warranted. Simulations may also be used when no data is available, for example to evaluate DTC genetic tests. By comparing the AUC values estimated in simulation studies with those reported in the empirical prediction studies this study showed that, in general, the simulation method estimated similar AUC values. The simulation method replicated the published AUC values most accurately when empirical prediction studies used logistic regression models, but less accurate when prediction studies used risk scores (weighted and unweighted). Given that the AUC is less sensitive to the minor differences in ORs or frequencies of the genetic variants and that the differences in ORs across studies generally are small, the AUC values remained similar despite using different parameters for the SNPs. In conclusion, this simulation method has potential since it can provide an estimate of the predictive ability of genetic risk models based on published GWAS results and thus helps allocating money and time more efficiently.

Role of rare variants for risk prediction of common diseases (Chapter 5)

Next generation sequencing technologies are envisioned to enable the discovery of rare genetic variants for common diseases. Given that they are likely to be functional, i.e. to affect protein structure, and to show large effects on disease risk, rare variants are believed to improve the predictive ability of available risk prediction models for common diseases. The aim of this study was to investigate this hypothesis using simulation methods. Baseline risk models were built and their AUC was varied between 0.50 and 0.95. To these models, a single variant representing the cumulative frequency and OR of several rare variants was added. The benefit of adding the rare variants was assessed by measuring the increment in the AUC (ΔAUC), IDI and NRI.

The addition of rare variants showed minimal improvement in the AUC and reclassification measures. All measures increased with the frequency and OR of the variant. These results were replicated in a scenario in which a clinical risk model for atrial fibrillation was simulated. For this example, the maximum increment in the AUC was 0.02 for the variant with the largest simulated frequency (i.e. 0.01) and OR (i.e. 10). Because rare variants have by definition a low sensitivity when disease is common, their predictive ability at the population level will remain low. Variants with a larger effect and higher frequency than those simulated in this study will be needed to observe an improved predictive ability over available risk models.

Review of genetic risk prediction studies based on the GRIPS statement (Chapter 6)

The recently published Genetic RIsk Prediction Studies (GRIPS) statement seeks to improve the transparency, completeness and quality of reporting of genetic risk prediction studies. The aim of this study was to assess the quality of reporting prior to the guideline publication. Studies were selected if they were published in 2010 and if they developed or validated a prediction model based on multiple DNA variants. Forty-two studies met these inclusion criteria. The quality of reporting was good for the general items, but for the relevant and specific items the reporting was poor. Only 12% of the articles reported all essential items needed to perform external validation of the risk model. When interpreting these results, it is important to stress that this study focused on the quality of reporting and did not assess the likelihood of bias or the methodological quality of genetic risk prediction studies. These results do not show whether the authors correctly applied the measures of model performance. This study showed that the quality of reporting in genetic risk prediction studies was suboptimal prior to publication of the guidelines and indicated that so far genetic risk prediction studies provide insufficient information to replicate their models and to perform external validation.

Risk updates from direct-to-consumer genetic tests (Chapter 7)

Increasing number of companies sell online genome-wide scans for prediction of risk for common diseases directly to consumers. Because gene discovery research is still ongoing, these companies offer to update risk predictions when new susceptibility genes are discovered. However, the usefulness of receiving risk updates with every new discovery is unknown. Using empirical data on type 2 diabetes (T2D) as an example, this study assessed the change in disease risks when predictions based on one gene (TCF7L2) were consequently updated to18 genes and 18 genes plus clinical factors. The risk was stratified into two categories, that is below and above average risk. Overall, 39% of the individuals shifted once from above average to below average risk, or vice versa, and 11% shifted twice, i.e. back to their original category. To conclude,

updating risk predictions produces contradictory information about an individual's risk status over time, which is undesirable and may additionally be accompanied by varying lifestyle and nutritional recommendations.

External validation of genetic risk prediction models across European samples (Chapter 8)

Genetic variants are increasingly tested, alone or in combination with acknowledged risk factors, for their ability to predict risk of common diseases. Although the literature abounds of genetic risk prediction studies, they have rarely been validated in external samples. Using empirical data from nine European samples, the predictive performance of 37 susceptibility variants for T2D was investigated by constructing in each sample a risk model and testing them in all other samples. This study showed that the study design and characteristics of the populations greatly influence the predictive performance of genetic tests. The AUC values varied largely between the samples included in the study. Studies with small number of participants, like some of the genetic isolates, had genetic models with very large AUC values at model development, reflecting an overestimation of the true predictive ability of the model. High AUC values were also seen in cases-control studies in which individuals are selected from the extremes of the risk distribution. These high AUC values largely decreased when the models were tested in the other samples (maximum difference −0.33). This decrease in the AUC may be explained by the differences in the beta coefficients between the samples. Differences between beta coefficients estimates probably simply reflect sampling variation or low power to detect associations at common variant loci with modest effects in studies with small sample size. The lowest AUC at model development and validation was seen in the prospective cohort study. This study underscores the importance of developing genetic risk prediction models in samples with similar characteristics to the real clinical and public health settings for which genetic testing is envisioned.

Comparison of genetic risk predictions between commercial companies (Chapter 9)

In this study, the predictive ability and predicted risks of personal genome tests offered by 23andMe, deCODEme and Navigenics was assessed using the simulation strategies evaluated in Chapter 4 of this thesis. Genotype data was simulated based on published genotype frequencies and disease risks were predicted using the formulas of the companies. The predictive ability was assessed by the AUC. Predicted risks differed substantially between companies due to differences in the average population risks of disease, the number of SNPs and the computational methods used to predict individual risks. Per disease, the AUC values were similar between the companies. However, the predictive ability of the genetic tests varied substantially between diseases. The

AUC ranged from 0.58 for atrial fibrillation to 0.82 for age-related macular degeneration, celiac disease and Crohn's disease. The higher AUC values were determined by the availability of at least one SNP with a substantial impact on disease risk as indicated by a high odds ratio. A limitation of this study is that the AUC was calculated using genotype datasets based on the frequencies that the companies use to calculate the risks. This strategy gives the AUC for the best case scenario. In a random external population the AUC would be lower and therefore personal genome tests are likely less accurate than this simulation study suggests. Additionally, these analyses showed that predicted risks differed substantially between the three companies. The differences in predicted risks were based on the differences between the average risks of disease considered by each company, the number of SNPs and between the methods used to calculate them. DeCODEme and Navigenics used methods that allowed for the risk to be higher than 100%. Remarkably, companies provide different predicted risks even when personal genome tests have a similar predictive ability between companies. Thus individuals may receive contradictory information about genetic risks depending on which company they choose for testing. This study shows that the results provided by DTC genetic tests are not informative and the public should be educated about the limitations of these tests.

Methodological evaluation of genetic risk prediction studies for Type 2 Diabetes (Chapters 10 & 11)

Genome-wide association studies have been successful in identifying many polymorphisms associated with increased risk of T2D. Fueled by these results, many studies have investigated the predictive ability of genetic risk models in T2D. Genetic susceptibility testing for T2D was, at the time of writing the studies presented in this thesis, also offered by several commercial companies. In Chapters 10 and 11, the genetic risk prediction studies and the commercial tests were reviewed from a methodological perspective. In most empirical studies, the genetic risk scores had lower discriminative accuracy than the clinical risk factors. Furthermore, addition of genetic factors either did not change or only marginally improved the AUC beyond the clinical risk models. The distribution of clinical risk factors, particularly the difference in risk factors between cases and controls, and the study design largely influenced the AUC of the clinical models and to a smaller degree also the AUC of genetic models, as empirically demonstrated in Chapter 8 of this thesis. A conclusion of this research is that if case-control or cross-sectional studies are used to test the predictive ability of a genetic risk model, the AUC will only be valid if the distribution of risk factors is similar to the prospective studies. Furthermore and again (see Chapter 10), the predictive ability of T2D risk models should be investigated from the

perspective of a realistic health care scenario, i.e. by testing the models in populations relevant for targeting preventive interventions.

Direct-to-consumer risk companies sell risk profiles that differ in the number of genetic markers included, in the exact SNPs used and in the methods used to combine the SNPs in a risk model, as shown in Chapter 9 of this thesis. These differences, together with the differences in average population risks used as the baseline risk to depart from when computing predicted risks, are factors that lead to the disagreement in test results. Although both companies and scientific studies used multiplicative models or additive genetic effects, whether this is correct has not been demonstrated. Remarkably, none of the studies reviewed in Chapters 10 and 11 investigated the same panel of SNPs as the companies do. This, together with the lack of agreement on the true genetic model is further indication for the premature introduction of commercial genetic test based on multiple SNPs.

GENERAL DISCUSSION

Implications for methodology

Results presented in this thesis have important implications for the methodology of genetic risk prediction model construction and for the evaluation of genetic risk prediction models for common diseases. Both implications are discussed below. First, the methods of genetic risk model construction, including the statistical models, the model assumptions and the selection of high-risk groups based on genetic risk profiles, are discussed. Second, the methods used to evaluate genetic models for risk prediction of common diseases, the traditional and the new measures of model performance (i.e. AUC and reclassification), are discussed. This section concludes with a comment on the role of simulations in the field of genetic risk prediction for common diseases.

Methods of genetic risk model construction

Genetic risk prediction models usually include many genetic risk factors, alone or in combination with clinical risk factors. The large number of variables included in these models makes genetic risk prediction a challenging task from a methodological perspective. The assumptions taken and the population used to construct and test the genetic risk prediction models will ultimately influence the reported predictive ability of the models.

First, the study design and the clinical characteristics of the population included in the study have a large impact on the observed predictive ability of the genetic risk models (see Chapters 8 & 10). The highest predictive ability was obtained in studies using over-selected cases and controls. However the predictive ability of these models largely decreased at validation

on a prospective cohort study (see Chapter 8, this thesis). Case-control studies are excellent designs for the identification of genetic risk variants, but for evaluation of the predictive ability of genetic risk models cohort studies in populations for which the test is intended should be employed.

Second, no general agreement exists on the statistical models to be used in the construction of genetic risk models (see Chapter 6, this thesis). Although the most used statistical models are logistic regression and construction of risk scores, it is not known if these are the correct models. Other approaches have been tried, including Bayesian models and support vector machines (2, 3). However, these models have not shown a better predictive value than the commonly used models logistical regression and risk scores (3). Furthermore, no general agreement exists either about the weights incorporated in risk scores. If these should be derived from the own data, should be effect sizes from meta-analyses or should be effect sizes weighted by variability in the data it is still a matter of debate.

Third, the impact of the model assumptions on the predictive ability is not fully understood yet. Often the assumption of additive genetic effects is used in the construction of empirical genetic risk models (see Chapters 10 & 11, this thesis). This assumes the independence of effects but there is no definitive evidence that this simplistic approach is correct. A number of factors, if proven to be relevant, could impact model construction if their effect is strong and if they are common (see Chapter 5, this thesis). Genes interact with each other and environmental factors alter the expression of genes through various mechanisms (4). Although such interactions have been identified, these discoveries need yet to be validated across various samples. There are also other mechanisms that operate on the genomic level, for example methilation or the parent of origin effect (5, 6). In future, genetic risk prediction models should include higher levels of interaction between genetic factors, such as the multilocus effect, i.e. specific combinations of genetic risk variants that increase the risk of disease.

Fourth, there is a lot of debate in the literature on the selection of high-risk groups based on genetic risk profiles (see Chapter 3, this thesis). As explained above, common disorders are caused by an intricate interplay between genetic and environmental risk factors that make it impossible to reveal the complete causal pathways. Each factor has a minor contribution to disease risk, and many combinations of risk factors exist that can lead to development of disease. Large numbers of genes will unlikely have substantial added predictive value over traditional risk factors if these variants predispose for the risk factors. Thus, recent research has been directed towards assessing the predictive ability of genetic tests in a selected subgroup of individuals at higher risk in which genetic factors may operate more prominently. One example is a selection based on age, i.e. at a younger age genetic risk factors would account for a larger proportion of

risk for disease than at older ages, when the accumulation of environmental factors would be a more important risk factor compared to genetic risk (7). The identification of risk groups in which genetic tests may have a higher predictive ability needs to be more systematically explored. Nonetheless, when the predictive value of a model that includes traditional risk factors slightly improves by the addition of genetic variants it indicates that genetic variants have an effect beyond intermediate phenotypes (8). This finding is useful both for risk prediction and for the information it brings about the pathways through which genetic factors operate on disease risk.

Methods for evaluation of genetic risk prediction models

Beyond the debate on the most accurate representation of the genetic architecture of disease when constructing risk models for common diseases, there is an abundant discussion about the methods that should be used to evaluate the performance of genetic risk models (see Chapter 2, this thesis).

First, the AUC is one of the most commonly used measures of predictive accuracy in genetic risk prediction research. The advantage of the AUC is that it does not need pre-defined risk thresholds. That is, the AUC can be used for any common disease and can be compared across diseases. Generally the AUC of clinical risk models for common diseases did not significantly increase with the addition of genetic information indicating that genetic factors have currently a limited use as risk predictors of common diseases (9). The AUC may improve if more risk variants with both higher ORs and genotype frequencies will be found, but if such risk variants would exist, we would probably have found them by now in GWASs. However, some researchers have considered the minimal improvement in the AUC as an indication of the limitations of this measure. The AUC has been criticized because it is insensitive to risk factors that have a small effect on disease risk (10), does not provide information about the risk at the individual level and does not provide information on clinical consequences (11). Thus, factors need to have a very large effect size in order to determine an increment in the AUC. Additionally, the AUC is based on ranking of risks between cases and controls and it does not take into account the absolute value of the risk difference. For example a high AUC can be obtained for a model that assigns a slightly higher risk to all cases compared to the controls, but according to the detractors of the AUC, in clinical practice the utility of such a risk model would be of limited value.

Second, partly in response to the criticism about the AUC more focus has been recently directed towards reclassification of individuals into risk groups (12-14). Researchers have rapidly adopted these new measures and indications of a possible role for genetic factors in risk prediction have been based on positive reclassification values. Research presented in this thesis draws attention over the cautious use and interpretation of these new measures. If they assume

risk categories they are only useful for certain diseases for which clinically defined risk groups are available. It has been shown in this thesis that even in absence of improvement in the AUC reclassification occurs (see Chapter 2, this thesis). This reflects the fact that the updated model still makes mistakes in risk classification, but assigns incorrect risks to other people than the original model. Reclassification measures that do not take into account risk categories generally do not show improvement when the AUC remains unchanged after the addition of genetic variants.

This thesis shows that only when the AUC is considered sufficient for a specific application, should the clinical utility of the model be evaluated (see Chapter 2, this thesis). Reclassification measures should be seen as a subsequent step in the assessment of a risk model. First, the general measures of model performance should be used to indicate if the model is accurate, well calibrated and has a good discriminative ability. Besides the reclassification measures, the decision curve analysis (11) and formal decision analytic methods can be used to incorporate the value of benefits and costs into the assessment of genetic risk models. Furthermore, the full potential of the AUC has not been explored. Some research has been done investigating the relationship between the AUC and various other measures of model performance (15), but these results have not yet been translated to genetic risk prediction research.

Simulation methods in the field of genetic risk prediction

Simulations are useful tools to evaluate the above mentioned factors: the effect of the model assumptions on the predictive performance, the value of genetic factors for risk prediction before large and expensive empirical studies are performed, the characteristics of the new measures and the relation between new and old measures of model performance. Furthermore, simulations are useful for the evaluation of DTC genetic test offered by commercial companies (see Chapters 7 & 9, this thesis). The simulation models used in this thesis assumed independence of the genetic variants and this may not reflect the true relationship between genetic variants and other genetic or clinical risk factors. These simulation models need to be expanded in the future to include these relationships, if proven. Nevertheless, it has been shown in this thesis that the results of empirical studies can generally be replicated by the simulation method (see Chapter 4, this thesis).

Clinical and public health implications

Genetic information is investigated for its role in clinical and public health practice (16). Genetic testing is envisioned to make personalized medicine possible through identification of people who are at increased risk, for targeting screening interventions and prevention strategies.

By preventing diseases comorbidity will be prevented as well. This thesis focuses on both the population and the individual level. For both applications it is important that genetic risk models are calibrated and can discriminate between individuals who will and those who will not develop the disease.

Prediction at population level

Genetic prediction at population level needs high discriminative accuracy, especially if the interventions are invasive. For most diseases, up to now genetic factors show insufficient discriminative accuracy. A high AUC can be observed in case-control studies with highly selected cases and controls, but this is an overestimation of the accuracy that the model would show in practice (see Chapter 8, this thesis). Genetic testing for common diseases in fact can never achieve perfect predictive ability, since there are always other factors beside the genetic risk variants that influence the disease risk. This is explained by the relationship between heritability of disease and the influence of genetic factors, i.e. for lower heritability estimates, like those seen for most common diseases (17), only a small part of the variation in the phenotype is caused by the variation in genetic factors (18). Ultimately, whether the AUCs will be considered to be high enough depends on the goal of testing, the burden of disease, the costs of disease, the availability of (preventive) treatment and the adverse effects of false-positive and false-negative test results.

Prediction at individual level

This thesis covers also the aspects of genetic risk prediction at the individual level, with a focus on the differences between risk prediction based on common and rare genetic variants. Simultaneous testing of multiple genes is characterized by the fact that all individuals in a population carry at least one or more risk genotypes, even those persons with a lower than average risk of disease. Due to the overall small effect sizes of the genetic variants, most individuals have profiles that are associated with disease risks that are only slightly higher or lower than the average risk in the population (see Chapter 7, this thesis). As a result, the value of genetic testing for risk prediction based on common variants has up to now a limited value at the individual level. Rare variants on the other hand are found in only a few individuals for whom the predicted risk is very large compared to the average risk of disease in the population (see Chapter 5, this thesis). However, the impact of the false-positive findings, i.e. carriers of rare variants that eventually will not develop the disease, can have serious clinical consequences if one considers for example the use of a genetic test to direct an invasive intervention. Testing for rare variants in correlation with family history may be more specific, i.e. would decrease the number of false-positive findings, and should be explored in future research.

Genetic risk prediction in scientific literature

The literature on genetic risk prediction is rapidly expanding. Easy access to the published articles however is hampered by the variability in terminology used to describe the aim of genetic risk prediction. There is also a large variability in the model assumptions and the parameters of the genetic variants. To allow the comparison of risk models, it is thus very important that researchers provide detailed information about these relevant factors. Reporting so far, as shown in Chapter 6 of this thesis, is insufficient: there is a wide variability in the parameters used to construct the genetic risk models, the populations used, the methods used to assess the models and the reporting of all these steps. Furthermore, the validation of the risk models is rarely performed as we have shown in our review of the literature on genetic risk prediction models (see Chapter 6, this thesis). An improved and regulated reporting, as described in the GRIPS guideline (19), is necessary to allow comparability of results and to further the replication of the published genetic risk models.

Direct-to-consumer genetic tests

Despite the limited applicability of genetic information for risk prediction in clinical practice, commercial companies provide genome scans based risk predictions directly to consumers (20-26). Transparency regarding the variants, parameters and computational methods used are essential for a proper evaluation of the tests offered by these companies. It has been shown in this thesis, using the simulation method, that the predicted risks for the same individual based on the information provided by three such companies largely differed depending on which company's data was used to calculate the risk (see Chapter 9, this thesis). It has also been shown that these predictions are unstable in time (see Chapter 7, this thesis) and the impact of such changes in prediction on the health related behavior of the consumers remains unknown. Due to their limited transparency and the unstable results the commercial use of genetic tests remains a premature application. It is of interest to mention here that two of the three companies in our studies have already gone out of business since the completion of this thesis.

Ethical implications of genetic risk prediction

The use of genetic information for risk prediction gives rise to specific ethical and legal implications. Examples include biobanks and the indefinite storage of DNA for future retesting. General ethical aspects are related to informed consent, privacy, confidentiality, discrimination and access to results. More research is needed to assess the impact of genetic testing on changes in the behavior of the individuals, e.g. the adoption of a healthier lifestyle or increased adherence to preventive strategies. Another aspect of the ethical implications of genetic testing is, as mentioned

above, the commercialization of genetic tests. Direct-to-consumer genetic tests are clear about the variants they use to predict risk of disease but less transparent about the parameters for the genetic variants and the methods used to calculate predictive risks (see Chapter 9, this thesis). Scientific evaluation of these tests, as reported in Chapters 7 & 9 of this thesis, are necessary to inform the public and to support more rigorous regulations of DTC genetic tests.

Future research

The work presented in this thesis opens the way to exciting new research areas: the impact of genetic testing on health care, the improvement of genetic risk prediction models and the further development of the simulation strategies. These aspects are discussed below in more detail.

As stressed in this thesis, to be useful for the prediction of disease, genetic profiles should discriminate between subjects who will develop the disease and those who will not. The discriminative accuracy depends, as expected, on the number of genes involved, the frequency of the risk alleles and the risks associated with the genotypes (see Chapter 5, this thesis). A good discriminative accuracy is also needed to achieve meaningful improvement in reclassification and high positive predictive values of genetic tests (see Chapters 2 & 3, this thesis). Next, it should be assessed whether the predictive value is sufficient for example to improve population health or to improve the efficiency or quality of health care.

Characteristics of the existing measures for the evaluation of risk prediction models as well as the relationship between the measures should be further explored. In the future, the genetic models take into account multilocus genetic effects, in other words the specific combination of genetic variants should be evaluated and not only their independent effects. Family history should also be incorporated in genetic risk prediction models, especially to evaluate if it can improve the predictive ability of rare genetic variants (see Chapter 5, this thesis). Overall, the mindset should change when constructing and testing genetic risk prediction models: the development and validation of genetic risk prediction models should be performed only from the clinical or public health perspective that represents the intended use of such models.

Simulation methods, which now focus solely on genetic risk variants, should be improved by modeling also the distribution of non-genetic risk factors and the correlations between the various risk factors. Simulations will be essential for the development and testing of new measures of model performance and for the evaluation of DTC tests. They should be more often used as a first step before conducting time consuming and expensive empirical studies. Costs should be included in such models and proper decision models should be constructed for the genetic models that show promising predictive accuracy to indicate if genetic tests are efficient strategies in public health practice.

We are still far from completely understanding the forces operating at the genomic level in determining the predisposition for common diseases, but have gained an important advance: the development of genetic risk prediction as a field in its own and the acknowledgment that it needs a standardized approach in the evaluation of genetic tests and in reporting the results. The doors are now open.

REFERENCES

1. Humphries SE, Yiannakouris N, Talmud PJ. **Cardiovascular disease risk prediction using genetic information (gene scores): is it really informative?** *Curr Opin Lipidol*. 2008;**19**(2):128-32.
2. Davies RW, Dandona S, Stewart AF, et al. **Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies.** *Circ Cardiovasc Genet*. 2010;**3**(5):468-74.
3. Makowsky R, Pajewski NM, Klimentidis YC, et al. **Beyond missing heritability: prediction of complex traits.** *PLoS Genet*. 2011;**7**(4):e1002051.
4. Moore JH, Williams SM. **Epistasis and its implications for personal genetics.** *Am J Hum Genet*. 2009;**85**(3):309-20.
5. Schnabel RB, Baccarelli A, Lin H, et al. **Next steps in cardiovascular disease genomic research–sequencing, epigenetics, and transcriptomics.** *Clin Chem*. 2012;**58**(1):113-26.
6. Guilmatre A, Sharp AJ. **Parent of origin effects.** *Clin Genet*. 2012;**81**(3):201-9.
7. de Miguel-Yanes JM, Shrader P, Pencina MJ, et al. **Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms.** *Diabetes Care*. 2011;**34**(1):121-5.
8. Kathiresan S, Melander O, Anevski D, et al. **Polymorphisms associated with cholesterol and risk of cardiovascular events.** *N Engl J Med*. 2008;**358**(12):1240-9.
9. Salari K, Watkins H, Ashley EA. **Personalized medicine: hope or hype?** *Eur Heart J*. 2012.
10. Cook NR. **Use and misuse of the receiver operating characteristic curve in risk prediction.** *Circulation*. 2007;**115**(7):928-35.
11. Vickers AJ, Elkin EB. **Decision curve analysis: a novel method for evaluating prediction models.** *Med Decis Making*. 2006;**26**(6):565-74.
12. Cook NR, Ridker PM. **Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures.** *Ann Intern Med*. 2009;**150**(11):795-802.
13. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., et al. **Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond.** *Stat Med*. 2008;**27**(2):157-72; discussion 207-12.
14. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. **Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers.** *Stat Med*. 2011;**30**(1):11-21.
15. Kraemer HC. **Reconsidering the odds ratio as a measure of 2x2 association in a population.** *Stat Med*. 2004;**23**(2):257-70.
16. Jostins L, Barrett JC. **Genetic risk prediction in complex disease.** *Hum Mol Genet*. 2011;**20**(R2):R182-8.
17. Rowe SJ, Tenesa A. **Human complex trait genetics: lifting the lid of the genomics toolbox - from pathways to prediction.** *Curr Genomics*. 2012;**13**(3):213-24.
18. Visscher PM, Hill WG, Wray NR. **Heritability in the genomics era–concepts and misconceptions.** *Nat Rev Genet*. 2008;**9**(4):255-66.
19. Janssens AC, Ioannidis JP, van Duijn CM, et al. **Strengthening the reporting of genetic risk prediction studies: the GRIPS statement.** *Eur J Hum Genet*. 2011;**19**(8):833-6.
20. Pathway Genomics. (http://www.pathway.com/dna-reports/full-list-of-conditions). (Accessed November 16, 2010).

21. 23andMe. Science. (https://www.23andme.com/more/science/). (Accessed December 20, 2010).

22. CyGeneDirect. Knowing your risks. (http://www.cygenedirect.com/browse-11802/Knowing-Your-Risks.html). (Accessed December 20, 2010).

23. deCODEme. (http://www.decodeme.com). (Accessed November 16, 2010).

24. GeneticHealth. Obesity test. (http://www.genetic-health.co.uk/dna-test-services/obesity-test.htm). (Accessed November 16, 2010).

25. Navigenics. The science. (http://www.navigenics.com/static/pdf/Navigenics-TheScience.pdf). (Accessed December 20, 2010).

26. PathwayGenomics. Terms of service. (http://www.pathway.com/about-us/terms-of-service). (Accessed December 20, 2010).

Appendix

Table 1. ACCE Model List of 44 Targeted Questions Aimed at a Comprehensive Review of Genetic Testing.

Element	Component	Specific Question
Disorder/Setting		1. What is the specific clinical disorder to be studied? 2. What are the clinical findings defining this disorder? 3. What is the clinical setting in which the test is to be performed? 4. What DNA test(s) are associated with this disorder? 5. Are preliminary screening questions employed? 6. Is it a stand-alone test or is it one of a series of tests? 7. If it is part of a series of screening tests, are all tests performed in all instances (parallel) or are only some tests performed on the basis of other results (series)?
Analytic Validity		8. Is the test qualitative or quantitative?
	Sensitivity	9. How often is the test positive when a mutation is present?
	Specificity	10. How often is the test negative when a mutation is not present?
		11. Is an internal QC program defined and externally monitored? 12. Have repeated measurements been made on specimens? 13. What is the within- and between-laboratory precision? 14. If appropriate, how is confirmatory testing performed to resolve false positive results in a timely manner? 15. What range of patient specimens have been tested? 16. How often does the test fail to give a useable result? 17. How similar are results obtained in multiple laboratories using the same, or different technology?
Clinical Validity	Sensitivity	18. How often is the test positive when the disorder is present?
	Specificity	19. How often is the test negative when a disorder is not present?
		20. Are there methods to resolve clinical false positive results in a timely manner?
	Prevalence	21. What is the prevalence of the disorder in this setting?
		22. Has the test been adequately validated on all populations to which it may be offered? 23. What are the positive and negative predictive values? 24. What are the genotype/phenotype relationships? 25. What are the genetic, environmental or other modifiers?
Clinical Utility	Intervention	26. What is the natural history of the disorder?
	Intervention	27. What is the impact of a positive (or negative) test on patient care?
	Intervention	28. If applicable, are diagnostic tests available?
	Intervention	29. Is there an effective remedy, acceptable action, or other measurable benefit?
	Intervention	30. Is there general access to that remedy or action?
		31. Is the test being offered to a socially vulnerable population?
	Quality Assurance	32. What quality assurance measures are in place?
	Pilot Trials	33. What are the results of pilot trials?

Element	Component	Specific Question
Clinical Utility	Health Risks	34. What health risks can be identified for follow-up testing and/or intervention?
		35. What are the financial costs associated with testing?
	Economic	36. What are the economic benefits associated with actions resulting from testing?
	Facilities	37. What facilities/personnel are available or easily put in place?
	Education	38. What educational materials have been developed and validated and which of these are available?
		39. Are there informed consent requirements?
	Monitoring	40. What methods exist for long term monitoring?
		41. What guidelines have been developed for evaluating program performance?
ELSI	Impediments	42. What is known about stigmatization, discrimination, privacy/confidentiality and personal/family social issues?
		43. Are there legal issues regarding consent, ownership of data and/or samples, patents, licensing, proprietary testing, obligation to disclose, or reporting requirements?
	Safeguards	44. What safeguards have been described and are these safeguards in place and effective?

http://www.cdc.gov/genomics/gtesting/ACCE/acce_proj.htm; accessed on 17-03-2013 (1)

REFERENCES

1. Haddow JE, Palomaki GE. ACCE: **A Model Process for Evaluating Data on Emerging Genetic Tests. In: Khoury M LJ, Burke W, ed. Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease.** : *Oxford University Press*, 2003:217-33.

Samenvatting

SAMENVATTING

In dit proefschrift wordt methodologisch en empirisch onderzoek naar het voorspellen van genetisch risico op veelvoorkomende ziekten beschreven. Het methodologisch onderzoek betrof de toepassing van traditionele en nieuwe methoden voor het bepalen van de voorspellende waarde en het klinisch nut van predictiemodellen, het gebruik van zeldzame varianten voor risicopredictie van veelvoorkomende ziekten, het gebruik van simulatiemodellen om empirisch onderzoek te repliceren, en de kwaliteit en volledigheid van rapportage in empirische risicopredictie-onderzoeken. Het empirisch onderzoek omvatte de evaluatie van DTC-genetische-tests (direct-to-consumer = direct aan de consument), de externe validatie van genetisch risico predictiemodellen in Europese onderzoeken, en de analyse van kwaliteitsverschillen in gepubliceerde genetisch risico predictiemodellen.

Voor deze onderzoeken werd gebruik gemaakt van zowel gesimuleerde als empirische gegevens. De gesimuleerde gegevens werden verkregen door met behulp van simulatiestudies populatiegegevens te genereren die representatief waren voor die van populaties in empirische onderzoeken. De empirische gegevens waren afkomstig uit de Rotterdam Studie, een prospectieve cohortstudie onder 7983 inwoners van een Rotterdamse wijk, waarin determinanten van chronische ziekten werden onderzocht, en uit Amerikaans bevolkingsonderzoek, te weten de Atherosclerosis Risk in Communities Study, de Cardiovascular Health Study en de Framingham Heart Study.

Reclassificatiematen versus de oppervlakte onder de ROC-curve (Hoofdstuk 2)

Er is de laatste tijd veel aandacht voor reclassificatiematen die verbeteringen in de kwaliteit van risico predictiemodellen uitdrukken. Ze zouden nauwkeuriger zijn in het detecteren van kleine verbeteringen in modellen dan traditionele maten. Het was echter nog niet duidelijk of deze maten gebruikt zouden moeten worden als een additionele stap naast de traditionele maten van modelverbetering, zoals de AUC (area under the curve = oppervlakte onder de ROC-curve), of dat ze deze zouden moeten vervangen. Om deze vraag te kunnen beantwoorden werden percentages berekend van totale reclassificatie, NRI (net reclassification improvement = netto reclassificatieverbetering) en IDI (integrated discrimination improvement = geïntegreerde discriminatieverbetering) voor verschillende verbeteringen in de AUC, met een gelijke AUC als uitgangspunt. Dit onderzoek heeft drie eigenschappen van de reclassificatiematen aan het licht gebracht.

Ten eerste bleek dat de percentages van reclassificatie varieerden, in samenhang met de afkapwaarden die gebruikt werden om de risicocategorieën te definiëren. Met gebruik van de

data van de Rotterdam Studie bijvoorbeeld, varieerde bij een kleine verhoging van de AUC het percentage reclassificatie tussen 0 en 23 procent, afhankelijk van de gekozen afkapwaarde.

Ten tweede, en gerelateerd aan het voorgaande, werd er meer reclassificatie gezien als de afkapwaarde die de risicocategorieën definieerde rond het centrum van de risicodistributie werd gekozen. Onderzoeken met willekeurig gekozen risicocategorieën kunnen een significante hoeveelheid reclassificatie laten zien, simpelweg omdat veel individuen een voorspeld risico hebben in de buurt van de afkapwaarde die gebruikt is om de risicocategorieën te definiëren.

Ten derde kwam naar voren dat reclassificatie waarbij geen verbetering van de AUC te zien is, impliceert dat het geüpdate model niet minder fouten maar andere fouten maakt dan het oorspronkelijke model.

Omdat reclassificatie samenhangt met de afkapwaarden heeft reclassificatie alleen zin als er gebruik wordt gemaakt van gefundeerde afkapwaarden. De keuze van het afkappunt dient niet gebaseerd te worden op de maximum hoeveelheid reclassificatie, maar dient uiteindelijk bepaald te worden door het wegen van de nadelen van fout positieve en de nadelen van fout negatieve beslissingen. Verder dient het aantal risicocategorieën bepaald te worden door het aantal preventieve behandelingsstrategieën.

Concluderend: de AUC en de reclassificatiemaat geven aanvullende informatie over het verbeteren van risicovoorspelling; de AUC is een maat voor discriminatie en reclassificatie is een maat voor klinische bruikbaarheid.

Selectie van hoogrisicogroepen gebaseerd op genetische data (Hoofdstuk 3)

Classificatie van individuen in hoogrisicogroepen gebaseerd op hun genetische profielen wordt beschouwd als een potentieel nieuwe strategie voor gerichte preventieve interventie van veelvoorkomende ziekten (1). Omdat er in veel gevallen geen afkapwaarden voor de definitie van de risicogroepen bestaan, worden er in de onderzoeksomgeving arbitraire afkapwaarden gebruikt om de klinische bruikbaarheid van genetische tests te evalueren, zoals besproken in Hoofdstuk 2 van dit proefschrift. Op dit moment is nog onbekend welke invloed de keuze van risicodrempelwaarden heeft op de kwaliteit van de risicoclassificatie, i.e. de sensitiviteit, specificiteit, de positief en negatief voorspellende waarde van de test. Ook is nog niet bekend welke andere epidemiologische parameters relevant zijn. Met gebruik van gesimuleerde data, geïllustreerd met een voorbeeld van het risico op AMD (age-related macular degeneration = leeftijdgerelateerde maculaire degeneratie), toetst dit onderzoek hoe de sensitiviteit, specificiteit en de positief en negatief voorspellende waarden van de genetisch risicomodellen afhangen van de te kiezen afkapwaarden om risicogroepen te definiëren. Dit onderzoek laat zien dat voor lage risicodrempelwaarden de sensitiviteit hoog was en de positief voorspellende waarde laag

terwijl voor hoge risicodrempelwaarden de sensitiviteit laag was en de positief voorspellende waarde hoog. Lage risicodrempelwaarden corresponderen met risicogroepen met een frequentie hoger dan de ziektefrequentie, en hoge risicodrempelwaarden met risicogroepen met een frequentie lager dan de ziektefrequentie. Alleen wanneer de frequentie van de risicogroep en de ziektefrequentie gelijk waren, waren positief voorspellende waarde en sensitiviteit gelijk, en oplopend met een stijging van de AUC van het risicomodel.

Concluderend: de frequentie van de risicogroep ten opzichte van de frequentie van de ziekte in de populatie heeft een grote invloed op de kwaliteit van de risicoclassificatiestrategie. De relatie tussen deze twee factoren bepaalt of de selectie van individuen met een hoog risico een effectieve strategie is voor de publieke gezondheidszorg. Met de replicatie van de resultaten in het AMD-risicopredictievoorbeeld, waar de risico's werden berekend door logistische regressieanalyse, liet dit onderzoek zien dat deze relatie generaliseerbaar is, en niet alleen geldt voor scores gebaseerd op aantallen, maar ook voor meer gebruikelijke genetisch risico predictiemodellen. Deze bevindingen zijn relevant voor onderzoek naar de bruikbaarheid van nieuwe biomarkers in aanvulling op bekende risicofactoren, een onderwerp dat op dit moment veel aandacht krijgt, omdat het inzicht geeft in de mogelijke klinische bruikbaarheid van de biomarkers.

Voorspelling gebaseerd op meta-analysedata met gebruik van simulatiestrategieën (Hoofdstuk 4)

Het uitvoeren van empirisch onderzoek naar het voorspellend vermogen van genetisch risicomodellen, zoals uitgedrukt in de AUC, is kostbaar en de resultaten hebben tot nu toe vaak een matig voorspellend vermogen laten zien. Daarom is het interessant te onderzoeken of de AUC van empirisch onderzoek gerepliceerd kan worden door middel van simulatiestudies met gebruik van gepubliceerde OR's (odds ratio's) en genotypefrequenties van GWAS's (genome wide association studies = genoombrede associatiestudies), die al eerder dan de empirische onderzoeken gepubliceerd zijn. Als simulaties de AUC-waarden accuraat kunnen repliceren, dan kunnen simulatiestudies gebruikt worden om te bepalen of empirische evaluatie van een genetisch risicomodel zinvol is. Simulaties kunnen ook gebruikt worden wanneer er geen data beschikbaar zijn, bijvoorbeeld om DTC-genetische-tests te evalueren. Voor dit onderzoek zijn AUC-waarden die geschat zijn in simulatiestudies vergeleken met gepubliceerde waarden uit empirische predictieonderzoeken, met als uitkomst dat in het algemeen, de simulatiemethode op dezelfde AUC-waarden uitkomt. De simulatiemethode repliceert de gepubliceerde AUC-waarden zeer accuraat wanneer empirische predictieonderzoeken logistische regressiemodellen gebruikten, maar minder accuraat wanneer predictieonderzoeken risicoscores (gewogen en

ongewogen) gebruikten. Gegeven dat de AUC minder gevoelig is voor kleine verschillen in OR's of frequenties van de genetische varianten en dat de verschillen in OR's tussen onderzoeken in het algemeen klein zijn, blijven de AUC-waarden gelijk ondanks het gebruik van verschillende parameters voor de genetische varianten.

Concluderend: deze simulatiemethode heeft potentieel omdat de methode een schatting kan geven van het voorspellend vermogen van genetisch risicomodellen gebaseerd op gepubliceerde GWAS-resultaten en daardoor van waarde kan zijn bij het efficiënter toewijzen van gelden en tijd.

De rol van zeldzame varianten voor risicovoorspelling van veelvoorkomende ziekten (Hoofdstuk 5)

Van de next-generation-sequencingtechnologieën wordt verwacht dat ze het mogelijk maken om zeldzame genetische varianten van veelvoorkomende ziekten te ontdekken. Gegeven dat ze waarschijnlijk functioneel zijn, i.e. ze kunnen proteïnestructuren beïnvloeden, en dat ze grote effecten hebben op het risico op ziekte, wordt van zeldzame varianten gedacht dat ze het voorspellende vermogen van beschikbare risico predictiemodellen voor veelvoorkomende ziekten kunnen verbeteren. Het doel van dit onderzoek was om deze hypothese te testen met simulatiemethoden. Er zijn baseline-risicomodellen gebouwd met een AUC variërend van 0,50 tot 0,95. Aan deze modellen werd een enkele variant, die de cumulatieve frequentie en OR van verschillende zeldzame varianten voorstelde, toegevoegd. Het nut van het toevoegen van zeldzame varianten werd getoetst door de verhoging van AUC te meten en de IDI en NRI te meten. De toevoeging van zeldzame varianten liet een minimale verbetering in de AUC en reclassificatiematen zien. Alle maten stegen met de frequentie en OR van de variant. Deze resultaten werden gerepliceerd in een scenario waarin een klinisch risicomodel voor atriumfibrilleren werd gesimuleerd. Voor dit voorbeeld was het maximum in stijging van de AUC 0,02 voor de variant met de hoogste gesimuleerde frequentie (i.e. 0,01) en OR (i.e. 10). Omdat wanneer een ziekte vaak voorkomt, zeldzame varianten per definitie een klein deel van de ziektegevallen verklaren, zal het voorspellend vermogen op populatieniveau laag blijven. Er zullen varianten met een groter effect en hogere frequentie dan gesimuleerd in dit onderzoek nodig zijn om een verbetering van voorspellend vermogen te zien ten opzichte van de beschikbare risicomodellen.

Beoordeling van genetisch risico predictie-onderzoeken gebaseerd op de GRIPS-richtlijn (Hoofdstuk 6)

De recent gepubliceerde GRIPS-richtlijn (Genetic Risk Prediction Studies statement) beoogt de verbetering van de transparantie, volledigheid en kwaliteit van de rapportage van genetisch risico predictie-onderzoeken. Het doel van dit onderzoek was het beoordelen van de kwaliteit van rapportage voordat de richtlijn werd gepubliceerd. Er werden onderzoeken geselecteerd die in 2010 werden gepubliceerd en waarin een predictiemodel gebaseerd op meerdere DNA-varianten werd ontwikkeld of gevalideerd. Tweeënveertig onderzoeken voldeden aan deze criteria. De kwaliteit van rapportage was goed wat betreft de algemene onderdelen, maar wat betreft de relevante specifieke onderdelen was de rapportage zwak. In slechts 12 procent van de artikelen werden alle essentiële onderdelen die nodig zijn voor externe validatie van het risicomodel gerapporteerd. Het is van belang te benadrukken dat bij de interpretatie van deze resultaten het onderzoek gericht was op de kwaliteit van rapportage en dat de aannemelijkheid van bias of de methodologische kwaliteit van de onderzoeken niet werden getoetst. Niet onderzocht is of de auteurs de modelprestatiematen correct hebben toegepast. Dit onderzoek liet zien dat de kwaliteit van rapportage in genetisch risico predictie-onderzoeken vóór de publicatie van de richtlijn niet optimaal was en dat de onderzoeken ontoereikende informatie verschaften voor het repliceren van de toegepaste modellen en het uitvoeren van externe validatie.

Risico-updates van DTC-genetische-tests (Hoofdstuk 7)

Steeds meer bedrijven verkopen online genoombrede scans voor de predictie van het risico op veelvoorkomende ziekten direct aan de consument. Omdat het bij het ontdekken van nieuwe genen om lopend onderzoek gaat, bieden die bedrijven aan om de risicopredicties te updaten als er nieuwe genen worden ontdekt die in verband worden gebracht met de vatbaarheid voor ziekte. Het nut van dergelijke updates bij elk nieuw ontdekt gen is echter onduidelijk. Gebruikmakend van empirische gegevens van type 2-diabetes als voorbeeld, werd in dit onderzoek de verandering in het geschatte risico onderzocht als de predictie gebaseerd op een enkel gen (TCF2G2) werd geüpdatet naar een predictie gebaseerd op 18 genen of 18 genen plus relevante klinische factoren. Daarbij werden twee risicocategorieën onderscheiden: groter dan gemiddeld en kleiner dan gemiddeld. Door de updates veranderde de risicocategorie bij 39 procent van de individuen één maal (van hoog naar laag of andersom) en bij elf van de individuen twee maal, i.e. terug naar de oorspronkelijke categorie.

Concluderend: het updaten van risicoschattingen kan leiden tot tegenstrijdige informatie met betrekking tot iemands risicostatus in het verloop van de tijd. Dit is niet wenselijk, ook omdat het kan leiden tot veranderende adviezen op gebied van leefstijl en/of dieet.

Externe validatie van risico predictiemodellen genetisch risico predictiemodellen aan de hand van Europese steekproeven (Hoofdstuk 8)

Genetische varianten worden steeds vaker getest, alleen of in combinatie met vastgestelde risicofactoren, op hun vermogen om het risico van veelvoorkomende ziekten te voorspellen. Hoewel de literatuur rijk is aan genetisch risico predictie-onderzoeken zijn er maar weinig gevalideerd met externe gegevens. Met gebruik van empirische gegevens van negen Europese steekproeven is het voorspellend vermogen van 37 genetische risicovarianten voor type 2-diabetes onderzocht door per steekproef een risicomodel op te stellen en dat vervolgens in alle andere steekproeven te toetsen. Dit onderzoek laat zien dat de onderzoeksopzet en de eigenschappen van de populaties van grote invloed zijn op het voorspellend vermogen van genetische tests. De AUC-waarden varieerden sterk tussen de verschillende steekproeven. Onderzoeken met een klein aantal deelnemers, bijvoorbeeld bij genetisch geïsoleerde populaties, leverden bij de modelontwikkeling modellen op met zeer hoge AUC-waarden, hetgeen een overschatting van het werkelijke voorspellende vermogen is. In case-controlstudies waarin individuen geselecteerd werden uit de extremen van de risicodistributie werden eveneens hoge AUC-waarden gezien. Deze hoge AUC-waarden daalden sterk wanneer de modellen getest werden in de andere steekproeven (maximum verschil -0,33). Deze daling in de AUC kan verklaard worden door een verschil in de bèta-coëfficiënten in de gebruikte steekproeven. Verschillen tussen bèta-coëfficiëntschattingen zijn waarschijnlijk het gevolg van steekproefvariatie, of van het gering vermogen van onderzoeken met kleine steekproefgrootte om associaties bij veelvoorkomende variante loci met gematigde effecten te detecteren. De laagste AUC bij modelontwikkeling en -validatie werd gezien in de prospectieve cohortstudie. Dit onderzoek onderstreept het belang van een steekproefpopulatie die representatief is voor de populatie waar de risico predictiemodellen genetisch risico predictiemodellen voor bedoeld zijn. Hiermee dient dus bij de ontwikkeling van de modellen rekening gehouden te worden.

Vergelijking van genetischrisicopredicties van verschillende commerciële bedrijven (Hoofdstuk 9)

In dit onderzoek zijn het voorspellend vermogen en de voorspelde risico's van persoonlijke genoomtests, 23andMe, deCODEme en Navigenics, getoetst met gebruik van de simulatie-strategieën zoals geëvalueerd in Hoofdstuk 4 van dit proefschrift. Genotypen werden gesimuleerd op basis van gepubliceerde genotypefrequenties en ziekterisico's werden voorspeld met gebruik van de formules van de bedrijven. Het voorspellend vermogen werd getoetst met behulp van de AUC. De voorspellingen van de drie bedrijven weken aanzienlijk van elkaar af ten gevolge van verschillen in het gemiddelde risico op ziekte in de populatie, het aantal genetische variantenen

de rekenmethoden gebruikt om het individuele risico te voorspellen. Per ziekte waren de AUC-waarden van de drie bedrijven vergelijkbaar. Maar het voorspellend vermogen van de genetische tests varieerde aanzienlijk tussen verschillende ziektes. De AUC varieerde tussen 0,58 voor atriumfibrilleren tot 0,82 voor leeftijdgerelateerde maculaire degeneratie, coeliakie en ziekte van Crohn. De hogere AUC-waarden waren het gevolg van de aanwezigheid van minstens één genetische variant met een substantiële invloed op het ziekterisico, uitgedrukt in een hoge OR. Een beperking van dit onderzoek is dat de AUC werd berekend aan de hand van genotypedatasets gebaseerd op dezelfde frequenties die de bedrijven gebruikten om de risico's te berekenen. Deze methode leidt tot een optimistische AUC-waarde. In een willekeurige externe populatie zou de AUC lager zijn en daarom zijn persoonlijke genoomtests waarschijnlijk minder accuraat dan dit onderzoek suggereert. Bovendien laat deze analyse zien dat de risicovoorspellingen van de drie bedrijven onderling substantieel kunnen verschillen. Deze verschillen zijn het gevolg van afwijkingen in de gemiddelde risico's zoals aangenomen door elk van de bedrijven, het aantal genetische variantenen de methoden gebruikt om de risico's te berekenen. Bij de door deCODEme en Navigenics gebruikte methoden mag het risico hoger dan 100 procent zijn. Opmerkelijk is dat de bedrijven verschillende risico's voorspellen zelfs als persoonlijke genoomtests een vergelijkbaar voorspellend vermogen hebben. Hierdoor kan de informatie die mensen over hun genetische risico's ontvangen, afhankelijk zijn van het bedrijf waar ze voor kiezen. Dit onderzoek laat zien dat de resultaten van DTC-genetische-tests weinigzeggend zijn, en dat het publiek moet worden voorgelicht over de beperkingen van deze tests.

Methodologische evaluatie van genetisch risico predictie-onderzoek voor type 2-diabetes (Hoofdstuk 10 & 11)

Door het uitvoeren van genoombrede associatieonderzoeken is men erin geslaagd veel polymorphismen die gerelateerd zijn aan een verhoogd risico op type 2-diabetes te identificeren. Vanwege deze resultaten is er veel onderzoek gedaan naar het voorspellend vermogen van genetischrisicomodellen voor type 2-diabetes. Op het moment van schrijven van dit proefschrift werden genetischrisicotests voor deze ziekte al door verschillende commerciële bedrijven geleverd. In Hoofdstuk 10 en 11 zijn de genetisch risico predictie-onderzoeken en de commerciële tests beoordeeld vanuit een methodologisch perspectief. In de meeste empirische onderzoeken bleken de genetischrisicoscores een lagere onderscheidende nauwkeurigheid te hebben dan de klinische risicofactoren. Daarnaast bleek een toevoeging van genetische factoren niets of weinig aan de AUC te verbeteren ten opzichte van de klinische risicomodellen. De verdeling van klinische risicofactoren, met name het verschil in risicofactoren tussen cases en controls, en de onderzoeksopzet waren sterk van invloed op de AUC van de klinische modellen

en in mindere mate ook op de AUC van de genetische modellen, zoals empirisch aangetoond in Hoofdstuk 8 van dit proefschrift. Een van de conclusies uit dit onderzoek is dat bij gebruik van casecontrolonderzoek of cross-sectioneel onderzoek om het voorspellend vermogen van een genetischrisicomodel te testen, de AUC alleen valide is als de verdeling van risicofactoren vergelijkbaar is met die van de prospectieve studies. Ook hier geldt dat het voorspellend vermogen van type 2-diabetesrisicomodellen onderzocht moet worden vanuit het perspectief van een realistisch zorgscenario, i.e. de modellen moeten getest worden met populaties die relevant zijn voor het doel van preventieve interventies (zie Hoofdstuk 10).

Risicoprofielen verkocht door DTC-bedrijven verschillen in het aantal genetische markers, het exacte aantal genetische variantenen in de rekenmethoden die worden gebruikt om de genetische variantenin één model te combineren, zoals aangetoond in Hoofdstuk 9 van dit proefschrift. Samen met de verschillen in gemiddelde populatierisico's die als uitgangspunt zijn gebruikt voor het berekenen van de individuele risico's, leiden deze verschillen tot tegenstrijdige testresultaten. Multiplicatieve modellen of additieve modellen voor genetische effecten worden door zowel bedrijven als in wetenschappelijke onderzoeken gebruikt, hoewel niet is aangetoond dat het gebruik van deze modellen terecht is. Opmerkelijk is dat in geen van de in Hoofdstuk 10 en 11 beoordeelde onderzoeken hetzelfde panel van genetische variantenis onderzocht als door de bedrijven gebruikt werd. Ook dit gegeven, samen met het gebrek aan overeenstemming over het werkelijke genetische model bevestigt de conclusie dat de introductie van commerciële genetische tests gebaseerd op meerdere genetische variantenvoorbarig is.

REFERENTIES

1. Humphries SE, Yiannakouris N, Talmud PJ. **Cardiovascular disease risk prediction using genetic information (gene scores): is it really informative?** *Curr Opin Lipidol.* 2008;**19**(2):128-32.

Acknowledgements

The completion of this thesis was only possible with the continuous support and dedication of many wonderful people. My gratitude goes to those that helped and inspired me along the past years.

First and foremost I would like to thank Prof. Janssens for her trust and for providing me with the opportunity to work at the department of Epidemiology at Erasmus Medical Center, for her inspirational motivation and passion for research and for being a great mentor. Dear Cecile, I could not have asked for a better supervisor to guide me through this incredible journey in scientific research and writing. Thank you for making it all possible!

Dear Prof. Hunnink, dear Myriam, thank you for your guidance, for the book clubs and for stimulating the communication between the members of our group. You always made me feel "at home" in the group.

My gratitude goes to the members of the reading committee, Prof. Tiemeier, Prof. Steyerberg, Prof. Cornel and the members of the larger committee, Prof. Veltman, Prof. Knoers, Dr. van Rossum, and Prof. Borry for taking the time to read my thesis and for accepting to act as opponents at my PhD defense.

My special thanks go to all the co-authors of the papers presented in this thesis, for their contributions and commitment! Dear Mandy, it was a pleasure working with you on a number of projects. Thank you for your to-the-point comments!

Dear Prof. Hofman, it was a pleasure and an honor to learn the foundations of epidemiology from you! I have enjoyed the courses, the Monday seminars and the stimulating discussions that emerged. Thank you for your support and encouragement during my work in the department of Epidemiology.

There are of course many other teachers that shaped my evolution and paved my way during the last 5 years. My gratitude goes to Prof. van Duin, Prof. Lesaffre, Prof. Uiterlinden, Prof. Oostra, Dr. Rizoupulos.

Erica, your warmth, smile and readiness to help were greatly appreciated. Nano, you always assisted me promptly whenever I had problems with a program and knew how to release the stress by making a joke. Thank you!

As part of my PhD I had the extraordinary chance to work for three months in the department of Biostatistics of the Boston University, in Boston, USA. I would like to thank first and foremost Prof. Benjamin who facilitated this fellowship. Dear Emelia, you helped me in every possible way, from joining the regular meetings, arranging discussions about my project and providing the data, to offering my husband and I a dinner and receiving us as guests in your home. We have both appreciated very much your hospitality and we warmly and sincerely thank you! Dear Dr. Pencina, dear Michael, thank you for patiently guiding me while I struggled with

the simulations. I am grateful to you also for giving me the opportunity to present my work at the Congress on Mathematical Modeling and Computational Simulation of Cardiovascular and Cardiopulmonary Dynamics in Williamsburg. I have enjoyed very much working with you! Prof. Lunetta and Prof. Sullivan, thank you for the eye-opening discussions!

My dear colleagues of the ART group, now the Clinical Epidemiology group, Tessa, Nathalie, Rachel, Ineke, Suman, Bart, Bob, Farzin, John, Sandra, Minghui, Felisia, Aletta, Ruben, Loes, Eline, Edwin I had a wonderful time with you and the memories from these last five years will remain forever. I enjoyed the diversity of our group and I have admired your talent and determination. It was a great pleasure to have met you!

Dear colleagues from genetic epidemiology: Najaf and Ayse (what a great time we had in Greece!), Sara, Linda, Adriana (thank you for your contagious smile and for the excellent Colombian coffee!), Natalia (my dear friend, I will miss our heart to heart talks), Aaron and Maxim; from epidemiology: Nese (thank you for your friendship), Rachel, Annemarie, Abbas and Maryam; from statistics: Karolina, Magdalena, Marek, Sten. I enjoyed the time we spent together, whether it was during the department events, after-work drinks, Dutch courses or just a chat at the coffee machine!

Dear Karin en Hille, we shared our struggles, dreams and happy moments and always found a place to meet whether it was in Rotterdam, Utrecht or Eindhoven. Thank you for standing by me during this important moment in my life!

My dear friends from home, Mihaela, Adriana, Oana you have always had time to listen and to encourage me. There are no words to describe how much I value your friendship!

A special thank you goes to my sister, Oana, for the unconditional love and support and for being a source of inspiration with her talent and creativity.

Dear parents, Paula and Traian, without your dedication, abnegation and love I would have never reached this moment! You have encouraged us to use every opportunity to broaden our horizons through knowledge and education. But more importantly you taught us never to forget where we came from. I dedicate this thesis to you, my source of motivation and inspiration.

Dear Roxana and Corneliu, my parents-in-law, you have been so wonderful to me and received me in your family with open hearts! You had so much patience during these years and understood when I was not able to join Mircea in his trips home. I am so lucky to have you, thank you! Dear Eleonora, Mircea's grandmother, you have always showed great care for my wellbeing and did all you could to make us feel at home when we visited you. A big and warm thank you goes to you!

My beloved husband, Mircea, it is wonderful that we have each other! You were there when I could barely move my feet to reach the top of Mont Blanc and during many other beautiful

and challenging adventures. Now I have summited another peak, but I know that, once again, this trip was not a lonely journey. You stood behind every step and you provided me with love, strength and support. I embrace you in my heart.

About the author

CURRICULUM VITAE

Raluca Mihăescu was born on September 6, 1977 in Iaşi, Romania. In 1995 she graduated from Mihai Eminescu high-school, Iaşi and started medical school at the University of Medicine and Pharmacy Grigore T. Popa in Iaşi. In 2001 she obtained her MD title and followed one year of postgraduate medical practice in internal medicine and surgery. In June 2002 she was a research fellow at the Smokers' Clinic, Barts and The London, London, UK in the group of Prof. Peter Hajek. Between 2004 and 2006 she followed a Master of Science in Psychosocial Intervention and Psychotherapy, University Alexandru Ioan Cuza, Faculty of Psychology and Educational Sciences in Iaşi and a Master of Science in Bioethics and Access to Health Care, University of Medicine and Pharmacy Grigore T. Popa in Iaşi while working as a general doctor. In 2006 she started the training in adult psychiatry at the University Hospital Socola in Iaşi.

In May 2008 she started the work presented in this thesis at the department of Epidemiology of the Erasmus MC under the supervision of Prof. A. Cecile J.W. Janssens and Prof. M.G. Myriam Hunink. As part of this PhD project she obtained a Master of Science degree in Health Sciences (Epidemiology) in August 2010 from the Netherlands Institute for Health Sciences (NIHES). Her MSc-thesis was awarded the NIHES prize for best paper written under the guidance of a NIHES tutor during the academic year of 2009-2010. As part of her PhD project she spent three months in 2011 as a research fellow at the department of Biostatistics, Boston University School of Public Health and The Framingham Heart Study, in the group of Prof. Emelia Benjamin.

Currently, she follows the residency training in psychiatry at the department of Psychiatry at Erasmus Medical Center, in Rotterdam.

PhD Portfolio

Name PhD student:	Raluca Mihaescu	PhD period:	2008 – 2012
Erasmus MC Department:	Epidemiology	Promotors:	Prof. Dr. A.C.J.W. Janssens
Research School:	NIHES Master of Genetic Epidemiology		Prof. Dr. M.G.M. Hunink

1. PhD training

	Year	Workload (Hours/ECTS)
General academic skills		
– Literature search and End Note course	2008	2 days
– Introduction to SPSS	2008	0.15
– Introduction to R programming language	2008	5 days
Research skills		
– Principles of Research in Medicine	2008	0.7
– Clinical Decision Analysis	2009	0.7
– Methods of Clinical Research	2008	0.7
– Topics in Meta-analysis	2010	0.7
– Health Economics	2009	0.7
– Genome Wide Association Analysis	2008	1.4
– Conceptual Foundation of Epidemiologic Study Design	2009	0.7
– Principles of Genetic Epidemiology	2008	0.7
– Primary and Secondary Prevention Research	2009	0.7
– Introduction to Decision-making in Medicine	2009	0.7
– Genomics in Molecular Medicine	2008	1.4
– Large-scale Multicenter Studies	2008	0.4
– Study Design	2008	4.3
– Classical Methods for Data-analysis	2009	5.7
– Modern Statistical Methods	2009	4.3
– Introduction to Clinical Research	2010	0.9

	Year	Workload (Hours/ECTS)
In-depth courses		
(e.g. Research school, Medical Training)		
– Genetic-Epidemiologic Research Methods	2009	5.7
– SNP's and Human Diseases	2008	1.4
– Courses of the Quantitative Researcher	2009	1.4
– Intervention Research and Clinical Trials	2010	0.9
– Genetic Analysis in Clinical Research	2009	1.9
– Advances in Population-based Studies of Complex Genetic Disorders	2009	1.4
– Genetic Linkage Analysis: Model-free Analysis	2009	1.4
– Ethical, Legal and Societal Aspects of Genomics	2010	1.4
Presentations		
– Invited lecture at the 2011 NHLBI-VCU-W&M World Congress on Mathematical Modeling and Computational Simulation of Cardiovascular and Cardiopulmonary Dynamics, College of William & Mary, Williamsburg, VA	2011	0.5
– Poster presentation at the European Society of Human Genetics Conference, Amsterdam	2011	0.5
– Poster presentation at the XVIIIth World Congress on Psychiatric Genetics, Athens, Greece	2010	0.5
– Oral presentation at 6th annual Centre for Medical Systems Biology symposium, Rotterdam	2009	0.2
– Oral presentation at the European Society of Human Genetics Conference, Vienna, Austria	2009	0.5
– Oral presentations at the Genetic Epidemiology seminars, Erasmus MC	2008-2010	
– Oral presentations at the Assessment of Radiological Technology (ART) meetings, Erasmus MC	2009-2012	

	Year	Workload (Hours/ECTS)
International conferences		
– Annual Centre for Medical Systems Biology Symposium	2009	0.3
– Annual Nederlandse Associatie voor Community Genetics en Public Health Genomics Meeting	2010-2011	0.6
– Research Institute of Diseases in the elderly (RIDE) symposium, Amsterdam	2008-2009	0.6
– Federation of Dutch Medical Scientific Societies (FEDERA), "Age brings infirmity: controversies and diagnosis", Leiden	2009	0.3
– Cardiology and vascular medicine, Rotterdam	2009	0.5
Seminars and workshops		
– Regular seminars at the department of epidemiology	2008-2012	4.0
– Mixed effects models working group at Erasmus MC		
Other		
– Research fellow (three months) at the group of Prof. Emelia Benjamin at the Department of Biostatistics, Boston University School of Public Health and The Framingham Heart Study	2011	3.0
– Reviewer for American Journal of Epidemiology and BMC Medical Genetics	2011-2012	
Awards		
– NIHES Award for best article written under the guidance of a NIHES tutor, Rotterdam	2010	

	Year	Workload (Hours/ECTS)
2. Teaching activities		
– Teaching assistant to fourth year medical students (Methods of epidemiological and clinical research) at Erasmus MC	2010	
– Invited speaker at the NIHES winter programme course "Introduction to clinical and public health genomics"	2012	

1 ECTS (European Credit Transfer System) equals a workload of 28 hours.

Publications

Mihaescu R, Pencina MJ, Alonso A, Lunetta KL, Heckbert SR, Benjamin EJ, Janssens AC. Incremental value of rare genetic variants for the prediction of multifactorial diseases. Genome Medicine. 2013;5(8):76.

Kalf RR, **Mihaescu R**, Kundu S, de Knijff P, Green RC, Janssens AC. Variations in predicted risks in personal genome testing for common complex diseases. Genetics in Medicine. 2013; doi:10.1038/gim.2013.80.

Mihaescu R, Moonesinghe R, Khoury MJ, Janssens AC. Predictive genetic testing for the identification of high-risk groups: a simulation study on the impact of predictive ability. Genome Medicine. 2011;3(7):51.

Willems SM, **Mihaescu R**, Sijbrands EJ, van Duijn CM, Janssens AC. A methodological perspective on genetic risk prediction studies in type 2 diabetes: recommendations for future research. Current Diabetes Reports. 2011;11(6):511-8.

Mihaescu R, Meigs J, Sijbrands E, Janssens AC. Genetic risk profiling for prediction of type 2 diabetes. PLoS Currents. 2011;3:RRN1208.

Mihaescu R, van Zitteren M, van Hoek M, Sijbrands EJ, Uitterlinden AG, Witteman JC, Hofman A, Hunink MG, van Duijn CM, Janssens AC. Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve. American Journal of Epidemiology. 2010;172(3):353-61.

Mihaescu R, Detmar SB, Cornel MC, van der Flier WM, Heutink P, Hol EM, Rikkert MG, van Duijn CM, Janssens AC. Translational research in genomics of Alzheimer's disease: a review of current practice and future perspectives. Journal of Alzheimer's Disease. 2010; 20(4)**:** 967-80.

Mihaescu R, van Hoek M, Sijbrands EJ, Uitterlinden AG, Witteman JC, Hofman A, van Duijn CM, Janssens AC. Evaluation of risk prediction updates from commercial genome-wide scans. Genetics in Medicine. 2009;11(8):588-94.