Published in final edited form as:

Am J Kidney Dis. 2012 January ; 59(1): 4–8. doi:10.1053/j.ajkd.2011.11.011.

Genetic Risk Prediction for CKD: A Journey of a Thousand Miles

Jeffrey B. Kopp, MD and Cheryl A. Winkler

National Institutes of Health, Bethesda, Maryland

By targeting interventions to high-risk population subgroups, tools that provide quantitative estimates of the risk for particular clinical outcomes have the potential to improve clinical decision making and reduce morbidity and mortality. The paradigmatic example of a clinical risk score is the Framingham risk score, which incorporates demographic (age, race, sex) and clinical (smoking and diabetes status, serum total cholesterol, and blood pressure values) variables to estimate 10-year risk for myocardial infarction.¹ Similarly, several research groups have reported chronic kidney disease (CKD) risk scores.^{2–5} The receiver operating characteristic curve C statistic values ranged from 0.67 to 0.84, although the highest value did not derive from a replication cohort (Table 1). Others have reported risk scores for developing end-stage renal disease,^{6–8} and the role of renal risk scores has been reviewed.⁹

Will genetic risk scores provide additional information to clinical risk scores, by defining the risk profile more precisely, and thus find a place in personalized nephrology care? A journey of a thousand miles begins beneath one's feet, according to Lao-Tzu, a Chinese sage of the 6th century BCE. In this issue of the *American Journal of Kidney Diseases*, O'Seaghdha and colleagues¹⁰ have taken a first step toward the goal of assisting clinicians to use genetic information to supplement clinical scoring systems for prediction of CKD risk. The authors used data from the Framingham Heart Study Original and Offspring cohorts, involving 2,489 participants. CKD stage 3 was defined as estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m² based on serum creatinine using the 4-variable Modification of Diet in Renal Disease Study equation. Over a mean of 10.8 years of follow-up, 270 cases of CKD stage 3 developed. The authors used a *clinical* risk score, with the variables age and sex, and developed a *genetic* risk score, using 16 single-nucleotide polymorphisms, 1 from each of 16 loci that have previously been associated with eGFR <60 mL/min/1.73 m² in European descent populations.¹¹ Of these loci, 2 involved nonsynonymous variants (ie, were non-codon changing), 2 were upstream of the transcriptional start site, 9 were located within introns, and 3 were intergenic or of uncertain location. None of these variants has been shown to affect function but instead they likely track functional alleles.

In their analysis, O'Seaghdha and colleagues found similar genetic risk scores among participants with and without CKD stage 3 (17.5 ± 2.8 vs 17.3 ± 2.6 risk alleles, respectively, out of a maximum 32 risk alleles). Nevertheless, logistic regression analysis showed an age- and sex-adjusted odds ratio of 1.06 (95% confidence interval, 1.01–1.11, $P = 0.03$) for each risk allele, representing a 6% increase in risk. However, the C statistics for the clinical model (adjusted for age, sex, cohort status, baseline eGFR, hypertension, diabetes, and proteinuria) and the combined clinical/genetic model were 0.780 and 0.781, respectively ($P = 0.06$). Secondary analysis by age and by diabetes status did not alter the

Address correspondence to Jeffrey Kopp, MD, National Institutes of Health, 10 Center Dr, Bethesda, MD 20892. jbkopp@nih.gov.

Published by Elsevier Inc. on behalf of the National Kidney Foundation, Inc. This is a US Government Work. There are no restrictions on its use.

Financial Disclosure: The authors declare that they have no relevant financial interests.

outcome. Interestingly, a genetic risk score with risk alleles weighted by their β (regression) coefficients from previous meta-analyses showed more risk alleles in those who did and did not develop CKD stage 3 (18.2 and 17.7, respectively; $P = 0.01$), but this more informed genetic risk score still did not increase the C statistic in the combined clinical/genetic risk model.

The report by O'Seaghdha and colleagues has a number of limitations, as the authors point out. First, the study involved only European descent participants. As CKD risk is elevated in other ethnic groups, similar studies in diverse populations, using appropriately selected genetic variants, will be important. Second, the 16 genetic loci account for a small portion of the likely genetic contribution to CKD incidence. While certain common risk alleles may contribute to CKD risk, with the *MYH9/APOL1* locus being the most salient example,^{12,13} it appears likely that rare genetic variants at multiple loci may explain much of the missing heritability of common complex diseases, including CKD. Third, unless markers are in very strong linkage disequilibrium with the true causal variants, it may be difficult to infer mode of inheritance, and the allele model will generally be less accurate in quantifying risk than a genotype model informed by mode of inheritance.

In Table 2, the results presented by O'Seaghdha and colleagues for the clinical risk score and the combined clinical risk and genetic risk scores are put in context with genetic risk score studies involving coronary heart disease and type 2 diabetes.^{10,14–17} In all 5 studies, the addition of the genetic risk score to the clinical risk score failed to improve the C statistic associated or the net reclassification improvement (NRI), although one study found gains in the clinical NRI and the integrated discrimination improvement.

Genetic risk scores attempt to combine data from multiple genetic variants to provide a risk estimate for complex genetic diseases, which typically involve multiple genetic variants, each of which has low penetrance, and environmental factors. This topic has been the subject of several recent reviews. Jostins and Barrett assessed the performance of genetic risk scores for 18 common diseases (not including CKD),¹⁸ while Drawz and Sedor discuss the role and assessment of genetic testing in personalized renal medicine.¹⁹ The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) working group reviewed the data for 29 genes in which variants have been implicated in cardiovascular risk and concluded that there was no evidence of net health benefit (benefits minus costs) to adding any of these genetic tests, singly or as panels, to the Framingham score.²⁰ A consortium meeting at the US Centers for Disease Control and Prevention in 2009 made recommendations for the generation and description of genetic risk scores, and noted that while many genetic risk scores have been published, the performance of most has been poor; exceptions noted include genetic risk scores for age-related macular degeneration, hypertriglyceridemia, and Crohn disease.^{21,22}

Is the improvement in the C statistic the best yardstick to measure the incremental benefit afforded by genetic risk score? Two aspects of model accuracy can be distinguished.²³ First, *discrimination* is a measure of how well a test distinguishes those with disease and those without disease, and is most commonly applied to assessing the sensitivity, specificity, and predictive value of diagnostic tests. The C statistic, which represents the area under the receiver operating characteristic curve, also is equivalent to the probability that an individual with disease has a higher test score than a healthy individual. Second, *calibration* is a measure of how well predicted probabilities agree with observed outcomes. It is important that the fit of predicted probabilities and observed outcomes is maintained across various subgroups (eg, deciles of risk); this can be assessed by various statistical tests, including the Hosmer-Lemeshow statistic.

Some statistical tests combine elements of calibration and discrimination, such as the likelihood ratio statistic, NRI, clinical NRI, and integrated discrimination improvement; this topic has been the subject of several recent reviews.^{24–27} As an example, the Framingham risk score classifies an individual's 10-year cardiovascular risk as low (<10%), moderate (10%–20%), and high (>20%), and the clinician could particularly benefit from knowing more about those in the moderate group. The NRI is calculated as follows:

$$\left[\frac{(\text{events reclassified higher} - \text{events reclassified lower})}{\text{total events}} \right] + \left[\frac{(\text{nonevents reclassified lower} - \text{nonevents reclassified higher})}{\text{total nonevents}} \right]$$

Both elements represent improvement and so the higher the sum, the more the model improvement. The clinical NRI applies NRI analysis to a particular subset, in this Framingham case, the moderate risk group (Table 2).

Although O'Seaghdha and colleagues found that the genetic risk score conferred no additional predictive information over a clinical risk score, this suggests that nongenetic factors, most of which are subject to modification by patients, are at present the only basis for kidney disease risk prediction in the general population. Nevertheless, this is the first step of an iterative process in which the growing understanding of the molecular landscape of CKD risk will almost certainly lead to integrated clinical and genetic risk scores. A similar genetic risk profile might be undertaken for diabetic nephropathy, given the growing number of associated single-nucleotide polymorphisms.²⁸

Acknowledgments

This work was supported by the Intramural Research Programs of the National Cancer Institute and the National Institute of Diabetes and Digestive and Kidney Diseases.

REFERENCES

1. Batsis JA, Lopez-Jimenez F. Cardiovascular risk assessment—from individual risk prediction to estimation of global risk and change in risk in the population. *BMC Med.* 2010; 8:29. [PubMed: 20500815]
2. Bang H, Vupputuri S, Shoham DA, et al. Screening for Occult Renal Disease (SCORED): a simple prediction model for chronic kidney disease. *Arch Intern Med.* 2007; 167(4):374–381. [PubMed: 17325299]
3. Chien KL, Lin HJ, Lee BC, Hsu HC, Lee YT, Chen MF. A prediction model for the risk of incident chronic kidney disease. *Am J Med.* 2010; 123(9):836–846.e2. [PubMed: 20800153]
4. Halbesma N, Jansen DF, Heymans MW, Stolk RP, de Jong PE, Gansevoort RT. Development and validation of a general population renal risk score. *Clin J Am Soc Nephrol.* 2011; 6(7):1731–1738. [PubMed: 21734089]
5. Kshirsagar AV, Bang H, Bombardieri AS, et al. A simple algorithm to predict incident kidney disease. *Arch Intern Med.* 2008; 168(22):2466–2473. [PubMed: 19064831]
6. Hippisley-Cox J, Coupland C. Predicting the risk of chronic kidney disease in men and women in England and Wales: prospective derivation and external validation of the QKidney Scores. *BMC Fam Pract.* 2010; 11:49. [PubMed: 20565929]
7. Johnson ES, Smith DH, Thorp ML, Yang X, Juhaeri J. Predicting the risk of end-stage renal disease in the population-based setting: a retrospective case-control study. *BMC Nephrol.* 2011; 12:17. [PubMed: 21545746]
8. Keane WF, Zhang Z, Lyle PA, et al. Risk scores for predicting outcomes in patients with type 2 diabetes and nephropathy: the RENAAL study. *Clin J Am Soc Nephrol.* 2006; 1(4):761–767. [PubMed: 17699284]
9. Taal MW, Brenner BM. Renal risk scores: progress and prospects. *Kidney Int.* 2008; 73(11):1216–1219. [PubMed: 18322541]

10. O'Seaghdha CM, Yang Q, Wu H, Hwang S-J, Fox CS. Performance of a genetic risk score for CKD stage 3 in the general population. *Am J Kidney Dis.* 2012; 59(1):19–24. [PubMed: 21995970]
11. Kottgen A, Pattaro C, Boger CA, et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet.* 2010; 42(5):376–384. [PubMed: 20383146]
12. Genovese G, Friedman DJ, Ross MD, et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science.* 2010; 329(5993):841–845. [PubMed: 20647424]
13. Kopp JB, Smith MW, Nelson GW, et al. MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. *Nat Genet.* 2008; 40(10):1175–1184. [PubMed: 18794856]
14. 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. [PubMed: 19020323]
15. Paynter NP, Chasman DI, Pare G, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA.* 2010; 303(7):631–637. [PubMed: 20159871]
16. Ripatti S, Tikkanen E, Orho-Melander M, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet.* 2010; 376(9750):1393–1400. [PubMed: 20971364]
17. 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. [PubMed: 20075150]
18. Jostins L, Barrett JC. Genetic risk prediction in complex disease. *Hum Mol Genet.* 2011; 20(R2):R182–R188. [PubMed: 21873261]
19. Drawz PE, Sedor JR. The genetics of common kidney disease: a pathway toward clinical relevance. *Nat Rev Nephrol.* 2011; 7(48):458–468. [PubMed: 21712849]
20. EGAPP Working Group. Recommendations from the EGAPP Working Group: genomic profiling to assess cardiovascular risk to improve cardiovascular health. *Genet Med.* 2010; 12(12):839–843. [PubMed: 21042222]
21. Janssens AC, Ioannidis JP, van Duijn CM, Little J, Khoury MJ. Strengthening the reporting of Genetic Risk Prediction Studies: the GRIPS Statement. *PLoS Med.* 2011; 8(3):e1000420. [PubMed: 21423587]
22. 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; 19(5)
23. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation.* 2007; 115(17):928–935. [PubMed: 17309939]
24. Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration [published online ahead of print July 4, 2011]. *Eur J Clin Invest.*
25. 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. [PubMed: 21204120]
26. Romanens M, Ackermann F, Spence JD, et al. Improvement of cardiovascular risk prediction: time to review current knowledge, debates, and fundamentals on how to assess test characteristics. *Eur J Cardiovasc Prev Rehabil.* 2010; 17(1):18–23. [PubMed: 19952758]
27. Steyerberg EW, Vickers AJ. Decision curve analysis: a discussion. *Med Decis Making.* 2008; 28(1):146–149. [PubMed: 18263565]
28. McDonough CW, Palmer ND, Hicks PJ, et al. A genome-wide association study for diabetic nephropathy genes in African Americans. *Kidney Int.* 2011; 79(5):563–572. [PubMed: 21150874]

Table 1

Risk Scores for Screening and Prediction of Incident CKD Stage 3

Study	Design	Discovery Cohort	Validation Cohort	Variables in the Model	C Statistic ^a		
					Discovery Cohort	Validation Cohort	NRI and IDI
Bang et al, ² 2007	Prediction	NHANES (12 y)	ARIC	Age, sex, anemia, HTN, DM, CVD, HF, PVD, proteinuria (latter lacking in ARIC)	0.88 (0.86–0.90)	0.71	NR
Kshirsagar et al, ⁵ 2008	Prediction (up to 9 y)	ARIC, CVS (2/3 sample)	ARIC, CVS (1/3 sample)	Age, sex, anemia, HTN, DM, CVD, HF, PVD	0.69	0.68	NR
Chien et al, ³ 2010	Prediction	Taiwan University Hospital cohort (4 y)	Taiwan community cohort (2 y)	Age, BMI, DBP, T2DM, stroke	0.768 (0.748–0.798)	0.667 (0.631–0.703)	NS
Halbesma et al, ⁴ 2011	Prediction (median f/u 6.4 y)	PREVEND	PREVEND (bootstrap resampling)	Age, SBP, CRP, albuminuria	0.84 (0.82–0.86)	0.84	NR

Note: Three studies defined the outcome as eGFR <60 mL/min/1.73m², while for the study by Halbesma et al, the outcome was rapid progression (decline rate in the top quintile) combined with eGFR < 60 mL/min/1.73 m². The receiver operating characteristic C statistic (a C score of 0.5 is equivalent to random selection) is shown, with the comparison between discovery and replication cohorts.

Abbreviations: ARIC, Atherosclerosis Risk in Communities; BMI, body mass index; CKD, chronic kidney disease; CRP, C-reactive protein; CVD, cardiovascular disease; CVS, Cardiovascular Health Study; DBP, diastolic blood pressure; DM, diabetes mellitus; HF, heart failure; HTN, hypertension; IDI, integrated discrimination improvement; NHANES, National Health and Nutrition Examination Survey; NR, not reported; NRI, net reclassification improvement; NS, not significant; PREVEND, Prevention of Renal and Vascular End-stage Disease Study; PVD, peripheral vascular disease; SBP, systolic blood pressure; T2DM, type 2 diabetes mellitus.

^aWhere given, the numbers in parentheses indicate the 95% confidence interval.

Table 2
Comparative Performance of Clinical, Genetic, and Combined Risk Scores in Predicting Complex Disease Incidence

Disease	Study	No. in Cohort (Cohort Name)	Replication Cohort (No.)	Outcome	Variables	Clinical Risk Score			Genetic Risk Score			Model Combining Clinical/Genetic Risk Scores	
						C Statistic	No. SNPs	C Statistic	OR (95% CI) per Risk Allele	C Statistic	Effect of Adding Genetic Risk Score	C Statistic	Effect of Adding Genetic Risk Score
CHD	Paynter et al. ¹⁵ 2010	19,313 (Women's Genome Health Study)	None	10-y MI risk	Age	0.701	12	0.517	1.05 (1.01–1.09)	0.705	NRI, 0.6%		
	Ripatti et al. ¹⁶ 2010	3,829 (Finnish case-control study)	Finnish/Swedish prospective cohorts (30,720)	10-y MI risk	Age, sex, LDL, HDL, smoking, BMI, BP, DM	0.860	13	ND		0.881	NRI, 2.2%; clinical NRI, 9.7%; IDI, 3%		
T2DM	Meigs et al. ¹⁴ 2008	2,377 (Framingham Offspring Study)	None	Incident DM	Age, sex, family history, BMI, fasting glucose, SBP, HDL, TG	0.900	18	ND	1.12 (1.07–1.17)	0.901	NRI, 4%		
	Talmud et al. ¹⁷ 2010	5,355 (UK Whitehall II study)	None	Incident DM	Same	0.78	20	0.54	ND	0.78	NRI, 0.2%		
CKD	O'Seaghdha et al. ¹⁰ 2011	2,489 (Framingham Heart Study)	None	eGFR < 60 after a mean of 10.8 y or Offspring), eGFR, HTN, DM, proteinuria	Age, sex, cohort (Original or Offspring), eGFR, HTN, DM, proteinuria	0.780	16	ND	1.06 (1.01–1.11) ^a	0.781	ND		

Note: For CHD and T2DM, 2 studies were selected that each that provided both clinical and genetic risk scores. For CKD, O'Seaghdha et al represent the only report to date that provides both clinical risk score and genetic risk score. All studies assumed that each allele had an equivalent effect on the primary outcome. The C statistic is shown for the clinical risk score, the genetic risk score (where available), and the combined clinical and genetic risk scores. Evaluating changes in C statistic, genetic risk scores published to date have in general not improved on clinical risk scores in predicting complex disease phenotypes. On the other hand, analytic approaches that took into account reclassification from one clinical category to another reached significance only in the report by Ripatti et al.

Abbreviations: BMI, body mass index; BP, blood pressure; CHD, coronary heart disease; CI, confidence interval; CKD, chronic kidney disease; CRP, C-reactive protein; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; HDL, high-density lipoprotein; HTN, hypertension; IDI, integrated discrimination improvement; LDL, low-density lipoprotein; MI, myocardial infarction; ND, not

determined; NRI, net reclassification improvement; OR, odds ratio; SBP, systolic blood pressure; SNPs, single nucleotide polymorphisms; T2DM, type 2 diabetes mellitus; TG, triglycerides; UK, United Kingdom.

^aAge and sex adjusted.