ELSEVIER

Journal of Clinical Epidemiology ■ (2015) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

A simple-to-use method incorporating genomic markers into prostate cancer risk prediction tools facilitated future validation

Sonja Grill[a,*], Mahdi Fallah[b], Robin J. Leach[c,d], Ian M. Thompson[c], Kari Hemminki[b,e], Donna P. Ankerst[a,c,f,g]

[a]*Department of Life Sciences of the Technical University Munich, Liesel-Beckmann-Str. 2, 85354 Freising, Germany*
[b]*Division of Molecular Genetic Epidemiology, German Cancer Research Centre, Im Neuenheimer Feld 580, Im Technologiepark, 69120 Heidelberg, Germany*
[c]*Department of Urology of the University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX, 78229, USA*
[d]*Department of Cellular and Structural Biology of the University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229, USA*
[e]*Center for Primary Health Care Research, Lund University, Box 117, 221 00 LUND, Sweden*
[f]*Department of Mathematics of the Technical University Munich, Boltzmannstr. 3, 85748 Garching b. München, Germany*
[g]*Department of Epidemiology and Biostatistics of the University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229, USA*

Accepted 9 January 2015; Published online xxxx

Abstract

Objectives: To incorporate single-nucleotide polymorphisms (SNPs) into the Prostate Cancer Prevention Trial Risk Calculator (PCPTRC).

Study Design and Setting: A multivariate random-effects meta-analysis of likelihood ratios (LRs) for 30 validated SNPs was performed, allowing the incorporation of linkage disequilibrium. LRs for an SNP were defined as the ratio of the probability of observing the SNP in prostate cancer cases relative to controls and estimated by published allele or genotype frequencies. LRs were multiplied by the PCPTRC prior odds of prostate cancer to provide updated posterior odds.

Results: In the meta-analysis (prostate cancer cases/controls = 386,538/985,968), all but two of the SNPs had at least one statistically significant allele LR ($P < 0.05$). The two SNPs with the largest LRs were rs16901979 [LR = 1.575 for one risk allele, 2.552 for two risk alleles (homozygous)] and rs1447295 (LR = 1.307 and 1.887, respectively).

Conclusion: The substantial investment in genome-wide association studies to discover SNPs associated with prostate cancer risk and the ability to integrate these findings into the PCPTRC allows investigators to validate these observations, to determine the clinical impact, and to ultimately improve clinical practice in the early detection of the most common cancer in men.

Keywords: Single-nucleotide polymorphism; Prostate cancer; Risk prediction; Likelihood ratio; Genome-wide association study; Meta-analysis

1. Introduction

As clinical practice increasingly focuses on personalized medicine and the capitalization of data from rapidly expanding data warehouses, so too must commonly used clinical risk prediction tools evolve. Currently, there are hundreds of clinical risk prediction tools available online, with objectives ranging from the prediction of onset of disease for use in screening to prognosis of outcomes after treatment. The most commonly used risk prediction tools, such as the Framingham risk calculator for cardiovascular events and the Breast Cancer Risk Assessment Tool, summarize previous analyses of very large observational or clinical trial populations. They serve two audiences—the clinician/patient duo trying to make a medical decision but also researchers concerned with the evaluation of new markers of disease. Variables typically included in risk calculators are the established factors routinely collected in practice: blood serum markers, bio-measures, such as blood

Funding: This work was sponsored by U.S. National Institutes of Health grants 1R01CA179115-01A1, U01CA86402, U24CA115102, and 5P30 CA0541474-18. The study was also supported by the EU Transcan program through German Ministry of Education and Science (01KT1402) (K.H.). No authors have a financial conflict of interest with this manuscript.

* Corresponding author. Technische Universitaet Muenchen, Zentrum Mathematik, Boltzmannstr. 3, 85748 Garching b. München, Germany. Tel.: +49-89-289-18390; fax: +49-89-289-18435.

E-mail address: sonja.grill@tum.de (S. Grill).

http://dx.doi.org/10.1016/j.jclinepi.2015.01.006
0895-4356/

What is new?

Key findings

- A simple and intuitive statistical method to update existing disease risk calculators for single-nucleotide polymorphisms (SNPs) based on published frequencies, with the option to incorporate external information on linkage disequilibrium, has been developed. This method has been applied to incorporate 30 SNPs that have been validated in multiple genomewide association studies (GWAS) into the online Prostate Cancer Prevention Trial Risk Calculator (PCPTRC).

What this adds to what was known?

- GWAS continue to discover novel SNPs for disease and validate previously discovered SNPs. Concurrently, advances in clinical risk prediction tools based on the established risk factors continue to evolve. This study provides the link between the two fields, by providing a method to incorporate the latest SNPs into risk calculators.

What is the implication and what should change now?

- The online PCPTRC now allows optional inclusion of up to 30 SNPs to facilitate external validation in other populations. Other established disease risk calculators could similarly incorporate information from newly discovered SNPs.

pressure, and demographic or behavioral factors, such as age and smoking history.

As an example to be used in this article, the Prostate Cancer Prevention Trial Risk Calculator (PCPTRC) predicts the likelihood of detecting prostate cancer if a prostate biopsy was to be performed based on the risk factors, prostate-specific antigen (PSA), digital rectal examination (DRE) findings, age, race, family history of prostate cancer, and prior biopsy history [1]. In addition to assisting clinicians and their patients around the world, posting of the calculator online brought the additional advantage of facilitating dozens of external validations across a range of international populations, thereby rapidly gaining evidence regarding its appropriateness for these populations [2–13].

Since its establishment, the PCPTRC has been modified to incorporate newly discovered and FDA-approved markers for prostate cancer, including the urine marker PCA3, serum marker percent free PSA, and detailed family history of prostate and breast cancer, through a Bayesian technique for updating a risk tool, called the likelihood ratio (LR) applied to data from external case–control studies [14–16]. These external case–control studies comprised a different patient population from that used to build the PCPTRC. In particular, the PCPTRC was built on a primarily older and healthy Caucasian US population undergoing screening for prostate cancer, whereas the external case–control studies were gathered from symptomatic patients presenting to local medical providers. These two populations were used to build separate parts of the updated risk calculator, which was a necessity because the new markers could not be retrospectively measured on the original participants of the PCPTRC. External validation of the updated PCPTRC on completely independent cohorts where individual patients had both the original PCPTRC risk factors and new markers measured was the proof-of-principle that the potential bias incurred would not outweigh the increase in predictive accuracy. To expedite external validation, the updated PCPTRC calculators were put online, soon thereafter they were validated, often multiple times [17–19]. Nevertheless, simulation studies are currently underway to assess the potential extent of bias by this approach, and methodological studies are being performed to find methods to ease the bias, such as by weighting observations in the case–control study according to their similarity to the patients used to make the base PCPTRC calculator.

Genetic markers, particularly single-nucleotide polymorphisms (SNPs), measured through increasingly less expensive technological platforms, have the potential to increase the practical utility of clinical risk prediction tools. Multiple genome-wide association studies (GWASs) have identified the most common SNPs associated with major diseases, and efforts have begun to assess their added value if added to existing clinical risk factors. In the common cancers, the improvement by adding genetic factors has been modest at best [20–29], but interest remains in external validation of risk predictions incorporating the genetic factors and the potential for new markers, including gene–environment interactions to improve disease prediction based on the established risk factors.

Of note is that the largest GWASs have been performed on exclusively Caucasian populations from around the world, including Iceland, Australia, Sweden, and the United States among others, see the Supplementary Appendix at www.jclinepi.com. Members of other races have been excluded in attempt to preserve homogeneity of allele frequencies, but there are still many population differences that make combining into a single meta-analysis challenging. The populations have undergone different screening and referral patterns; controls may comprise younger men (the Prostate Cancer Prevention Trial [PCPT] population was 55 years or older) or even children or women (many of the GWASs only crudely describe the populations in their Supplementary Appendix). Commonly, later published GWASs include the same participants as in a prior publication. This is often but not always explicitly stated but can be inferred from careful study of the participant descriptions. It is important to exclude prior studies so as not to double count patients; in the Supplementary Appendix at www.jclinepi.com, such excluded studies are indicated. Examining