Author's Accepted Manuscript

Risk Prediction of Prostate Cancer with Single Nucleotide Polymorphisms (SNPs) and Prostate-Specific Antigen (PSA)

Sam Li-Sheng Chen , Jean Ching-Yuan Fann , Csilla Sipeky , Teng-Kai Yang , Sherry Yueh-Hsia Chiu , Amy Ming-Fang Yen , Virpi Laitinen , Teuvo LJ. Tammela , Ulf-Håkan Stenman , Anssi Auvinen , Johanna Schleutker , Hsiu-Hsi Chen

PII: S0022-5347(18)44024-4
DOI: https://doi.org/10.1016/j.juro.2018.10.015
Reference: JURO 15863

To appear in: *The Journal of Urology*
Accepted Date: 4 October 2018

Please cite this article as: Chen SLS, Fann JCY, Sipeky C, Yang TK, Chiu SYH, Yen AMF, Laitinen V, Tammela TL, Stenman UH, Auvinen A, Schleutker J, Chen HH, Risk Prediction of Prostate Cancer with Single Nucleotide Polymorphisms (SNPs) and Prostate-Specific Antigen (PSA), *The Journal of Urology*® (2018), doi: https://doi.org/10.1016/j.juro.2018.10.015.

Risk Prediction of Prostate Cancer with Single Nucleotide Polymorphisms (SNPs) and Prostate-Specific Antigen (PSA)

Sam Li-Sheng Chen[1], Jean Ching-Yuan Fann[2], Csilla Sipeky[3], Teng-Kai Yang[4,5], Sherry Yueh-Hsia Chiu[6], Amy Ming-Fang Yen[1], Virpi Laitinen[7], Teuvo LJ Tammela[8], Ulf-Håkan Stenman[9,10], Anssi Auvinen[11], Johanna Schleutker[12], Hsiu-Hsi Chen[5]

[1] School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan

[2] Department of Health Industry Management, School of Healthcare Management, Kainan University, Tao-Yuan, Taiwan

[3] Department of Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, Finland

[4] Department of Urology, National Taiwan University Hospital

[5] Graduate Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan

[6] Department and Graduate Institute of Health Care Management, College of Management, Chang Gung University, Tao-Yuan, Taiwan

[7] Institute of Biomedical Technology/BioMediTech, University of Tampere, Tampere, Finland

[8] Department of Urology, Tampere University Hospital, and Faculty of Medicine and Biosciences, University of Tampere, Tampere, Finland

[9] Department of Clinical Chemistry, Helsinki University Central Hospital, Helsinki, Finland

[10] Faculty of Medicine, University of Helsinki, Helsinki, Finland

[11] Faculty of Social Sciences/Section of Health Sciences, University of Tampere, Tampere, Finland

[12] Department of Medical Genetics, Turku University Hospital, Turku, Finland

Running head: Prediction for Prostate Cancer by PSA and Genetic Polymorphisms

Correspondence & reprint requests to: Hsiu-Hsi Chen, Ph.D., Division of Biostatistics, Graduate Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Room 521, 5F, No. 17 Suchow Road, Taipei 100, Taiwan
Tel: +886-2-33228033, Fax: +886-2-23587707
E-mail: chenlin@ntu.edu.tw

Abstract

Purpose: Combined information on single nucleotide polymorphisms (SNPs) and prostate-specific antigen (PSA) offers opportunities for improving the performance of screening by risk stratification. We aimed to predict the risk of prostate cancer (PrCa) based on PSA together with SNPs information.

Materials and Methods: Prospective study of 20,575 men with PSA test and 4,967 men with polygenic risk score for PrCa based on 66 SNPs from the Finnish population-based screening trial for PrCa and 5,269 samples on seven SNPs from the Finnish PrCa DNA study. Bayesian predictive model was built for estimating the risk of PrCa by sequentially combining genetic information with PSA in comparison with PSA alone among study subjects limited with 4 ng/mL or above.

Results: The posterior odds for PrCa based on seven SNPs together with the PSA level ranged from 3.7 at 4 ng/mL, 14.2 at 6 ng/mL, 40.7 at 8 ng/mL, to 98.2 at 10 ng/mL. The area under receiver operating characteristic curve was elevated to 88.8% (95% CI: 88.6%-89.1%) with PSA in combination with the risk score based on seven SNPs in comparison with 70.1% (95% CI: 69.6%-70.7%) with PSA alone. It was further escalated to 96.7% (95% CI: 96.5%-96.9%) when all prostate cancer susceptibility polygenes were combined.

Conclusions: Expedient use of multiple genetic variants together with information on PSA levels better predicts the risk of PrCa than PSA alone and allows higher PSA cut-offs. Combined information also provides a basis for risk stratification that can be used for optimizing the performance of PrCa screening.

Introduction

Several international collaborative genome-wide association studies have been conducted to identify genetic factors in association with hereditary predisposition to prostate cancer (PrCa). A constellation of >120 single-nucleotide polymorphisms (SNPs) have been revealed with several located in five chromosomal regions-three at 8q24 and one each at 17q12 and 17q24.3.[1-5] Although the effect of each of the SNPs on the risk for PrCa is small to moderate, a strong cumulative association has been demonstrated by using several SNPs in combination. [6] Multiple prostate cancer-specific multigene panels have been evaluated for detection of PrCa.[7] Use of the major SNPs offers an opportunity to identify sub-groups of men with PrCa risk substantially below and above the population average.

In parallel with these genome-wide studies, the effectiveness of population-based screening for PrCA with prostate specific-antigen (PSA) has been intensively researched. However, the effectiveness of screening in reducing mortality is still debatable due to conflicting results of the two major randomized trials and the balance between benefits and harms remains uncertain.[8,9] To enhance the efficiency and reduce the harm, i.e. overdiagnosis caused by screening, combining genetic information together with PSA given age and genetic variant holds promise for more accurate identification of high-risk men with potential for large screening benefits.

The purpose of this study was to develop a Bayesian algorithm to predict the risk of PrCA based on PSA data together with the SNPs identified from the Finnish PrCa DNA study and the Finnish population-based screening trial for PrCA in order to compare the

performance of the risk prediction for PrCa between PSA alone and PSA with the incorporation of information on SNPs.

Methods

Study Subjects and Design

To estimate the risk of PrCa based on PSA and selected SNPs, we combined two Finnish datasets, one from the population-based randomized screening trial during 1996-2007 (20,575 men enrolled) and an unselected patient series from Tampere University hospital during 1994-2013. The details of study design and preliminary results for the former have been published previously[10-12] and the mortality results have been published also as a part of the ERSPC trial.[8] The dataset included DNA samples collected from 2,959 individuals who participated in the Finnish screening trial (518 prostate cancers and 2,441 prostate cancer-free subjects) plus 2310 prostate cancer patients from the Tampere University Hospital. It should be noted that information derived from the two datasets are complementary with each other as the genetic dataset from the unselected patients included wide-scale genetic information but with incomplete PSA data whereas the opposite was for the screening trial. Figure 1 gives a summary of the estimates of interest, the use of model and distribution, and data sources.

In order to do the risk stratification of PrCa, we adopted Bayesian sequential design by first classifying PSA into 13 categories with an increment of 0.5 ng/mL in study subjects limited with 4 ng/mL following the usual PSA threshold for referral to biopsy. Given the risk of PrCa by PSA level, we then added information on SNPs in a sequential manner from seven selected SNPs based on unselected patients to 66 SNPs based on the screening trial. Receiver operating characteristic (ROC) was used to assess the performance of the combined PSA and information on SNPs in comparison with PSA alone following the risk predicted by Bayesian algorithm.

Genetic polymorphisms

To incorporate information on SNPs in association with PrCa, we assessed the combined effects of seven SNPs, rs4242382, rs6983267, rs1601979, and rs1447295 at 8q24, rs104865677 at 7p15.2, rs138213197 and rs1859962 at 17q21. The risk allele A of rs424238 at 8q24 has been previously reported to be associated with PrCa and aggressive PrCa. The risk allele G of rs10486567 at 7p, the intron 2 of the JAZF zinc finger1 gene (*JAZF1*) is commonly seen in Europeans. [13] The association between rs138213197 in HOXB13 and the risk of hereditary PrCA has also been addressed, [14] and the effect has been shown to be especially strong in the Finnish population.[15] With the advent of more SNPs in association with PrCa susceptibility, the analysis using polygenetic risk score was based on 66 SNPs for a sample of the trial participants, 1093 men with PrCa and 3874 men without PrCa.[16]

Statistical Analysis

To fit the normal distribution, the PSA concentrations were transformed into logarithms. The distributions of PSA in men with and without PrCa as well as aggressive PrCa (Gleason score ≥7) are given in the Appendix Tables 1 and 2. To incorporate information on SNPs, we first assessed the effects of each of seven SNPs on PrCa and aggressive PrCa by logistic regression analysis. The effects of the combined seven SNP on PrCa and aggressive PrCa was evaluated by two ways, treating each SNP as a dichotomous variable, and treating seven SNPs as a polygenic risk score. Such a risk-score-based approach was further applied to 66 SNPs.

The optimal cutoff of PSA based on receiver operating characteristics (ROC) curve was calculated by the largest value of the formula, Sensitivity + Specificity – 1, from each PSA cut-off. The bootstrap method was adopted by sampling individuals with replacement from the original sample to validate the prediction model. The sample size varied according to the number of events per variable (EPV) from 10 to 80. As the genetic variants associated with PCa are heterogeneous, it is necessary to make a comparison across different ethnic groups or populations by using the information on the proportion of each SNP in population and the effect of each SNP to PrCa risk. We used results from the previous Zheng's study[6] for external validation of developed model. The details of the algorithm developed by using Bayesian underpinning are given in the Appendix. Data analysis was performed with SAS 9.4 and Winbugs software.

Results

Estimates of the risk for PrCa (Posterior Odds) by different levels of PSA

Table 1 shows the likelihood ratios for log(PSA) and the SNPs, as well as the posterior odds by PSA levels given the prior odds (1: 2.78) for the risk of PrCa for men aged 60 years or younger at baseline. Our model was used to discern the PrCa cases from 4 ng/mL upward given the posterior odds by combining PSA and 7 SNPs, increasing from 3.7 (95% CI:1.6-10) at 4 ng/mL of PSA to 98.2 (95% CI:27.3-437.5) at 10 ng/mL of PSA. The likelihood ratio based on the presence of the risk alleles of 7 SNPs was 2.8 considering the weighted distribution (the proportion of each SNP in population) contributed from each SNP (see the footnote of Table 1). The frequencies of these 7 SNPs in patients and controls are listed in Appendix Tables 2-1 to 2-7. The corresponding posterior odds for PrCa based on risk-score model with 7 SNPs as well as all susceptibility polygenes are presented in Appendix Tables 3-1 and 3-2. The posterior probability of PrCA by age and PSA level taking 7 SNPs into account was simulated and the results are shown in Figure 2.

ROC curves limited to men with PSA ≥ 4 ng/mL

Adding SNP information to this risk group substantially enhanced the performance of risk prediction for PrCA as the area under curve (AUC) from 70.1% based on PSA only to 95.8% based on PSA combined with 7 SNPs when each of SNP was treated as a binary variable (Figure 3).

Figure 4 (A) shows the corresponding figure was 88.8% with 7 SNPs and was elevated to 96.7% when 66 SNPs were considered on the basis of the risk-score-based approach. It is very interesting to see PSA combined with 66 SNPs was not able to enhance the performance of risk prediction for aggressive PrCA as good as PSA combined with 7 SNPs (Figure 4(B)). The AUC increased from 77.0% based on PSA alone to 83.8% based on PSA combined with 7 SNPs whereas 80.6% of the AUC was noted when PSA was combined with 66 SNPs.

External Validation

The proposed predictive model was further extended to incorporate 5 SNPs from the Zheng's study.[6] Considering the 5 SNPs, the odds of PrCa was 2.4 at 4 ng/mL compared with 2 at 0.6 ng/mL (Table 2). The optimal cutoff was 9.9 ng/mL when using PSA plus 5 SNPs. The corresponding AUC was 86.8% (95% CI: 86.6%-87.0%).

The external validation based on four common SNPs (rs1859962, rs16901979, rs6983267, and rs1447295) from the Finnish and Zheng's studies was also conducted. The predicted ROC curve was built by applying the regression coefficients of 4 SNPs obtained from the Zheng's study to the empirical Finnish PSA data. The comparison between the externally predicted ROC and the observed ROC of the Finnish PSA data is shown in Figure 5. We found that AUC of 81.7% (95% CI: 81.5%-82.0%) in the Zheng's study was slightly lower than the 85.3% (95% CI: 85.1%-85.5%) in Finnish data ($P<0.0001$). The statistical significant difference suggests the results are not compatible even if the difference in the ROC values was not substantial.

Discussion

Using a novel clinical prediction algorithm with Bayesian underpinning that provides a feasible approach for the risk stratification of PrCa by combing information on PSA multiple genetic variants identified from genome-wide studies, we demonstrate here that adding available SNPs information to subjects with PSA > 4 ng/mL increased predictive ability of AUC substantially (by 25 percentage points) in our analysis. The enhanced predictive ability resulting from additional information on SNPs noted in the current study was supported by the recent finding that of 7.7-fold difference between the top and bottom 10 percent of polygenic risk score using 147 prostate cancer-susceptibility variants.[17] This finding leads to the following three merits for PSA screening. Firstly, the combined use of information on PSA and the SNPs may reduce false negative cases missed at PSA screen (such as interval cancer), as some men with low PSA levels may nevertheless have an increased risk of PrCA if they carry one or more high-risk alleles. The posterior odds was 4-fold higher than the prior based on PSA alone at 4 ng/ml if all 7 risk SNPs were present. Second, so doing may also reduce false positive results. The optimal cut-off was raised from 9.1 to 10.7 when information on the 7 SNPs is added, which is likely to reduce the frequency of screen-positive findings (Among men with ≥4 ng/mL, 17.5 % of men had PSA>9.1 ng/mL in our screening data). Third, the large contrast in PrCa risk between high and low-risk groups provides opportunities for individually tailored screening strategy including the adoption of screening tool, inter-screening interval, and age to begin with screen. The higher the risk predicted by the proposed model, the more advanced detection method, the shorter inter-screening interval, and the earlier age of commencing screening should be considered[12].

There are four concerns that should be addressed here from both methodological and application viewpoints. A key methodological concern is the assumption that the SNPs are independent of PSA level, which remains imperfectly verified. It could be debated whether such an assumption is reasonable. A previous study demonstrated that the five PrCA associated SNPs were independent of PSA levels[6]. There is no significant association between SNPs and PSA concentration in patient samples[18,19]. Accordingly, the joint effect of PSA and these seven SNPs can be easily decomposed into the product of their independent effects. Although this assumption is supported by Zheng et al., it should be empirically verified before applying our PrCa risk stratification algorithm for screening. It should be noted that AUC decreased from 95.8% in independent effect model treating the effect of each SNP as a binary variable to 88.8% in risk-score-based model for selected 7 SNPs. The risk-score-based model can capture the correlations within selected SNPs. Unfortunately, the risk-based approach cannot be validated by Zheng's study as risk score was not available from their study.

As far as the consistency of results across studies is concerned, the performance of our results were compatible but slightly higher than those previous findings that predicting the risk for PrCa with PSA limited to PSA≥4 ng/mL[20-24] with the ranges of AUCs from 61 to 71%. The higher AUC in ROC analysis might be arguable with whether the predictive model is reliable in terms of sample size. To relieve this concern, the internal-validation by bootstrap method was therefore performed. With 500 bootstrap replications, the mean AUCs were ranged between 69.10% and 68.01% for EPV from 10 to 80, respectively. The estimated optimism was 0.06% (= 69.10%-68.04% (full samples))

for EPV=10, which shows good discrimination. With large sample size (EPV=40 or 80), a reduction in optimism but not a substantial difference was found, suggesting the reliability of the prediction model.

Another concern is the variation in genetic risk prediction across populations, i.e. population stratification. The genetic determinants of PrCa risk from different populations are not highly consistent, suggesting that the genetic factors underlying hereditary susceptibility may vary between populations. The validation was not well fitted in our analysis of external validation. This suggests that different SNPs will need to be incorporated in different populations. It is still unclear to what extent the proposed model can be applied to populations other than where it has been developed (possible overfitting). The contribution of additional SNPs depends on their frequency, effect size and independence of the already incorporated SNPs. However, we found that PSA combined with risk-score-based approach based on 7 SNPs out-performed PSA combined with polygenetic risk score based on 66 SNPs for risk prediction of aggressive tumor. Such a finding for aggressive prediction was not identical to that for the risk of PrCA. This suggests that the majority of 59 additional SNPs may not be predictive of aggressive PrCa in comparison with seven SNPs. The explanation is that seven SNPs may predispose people with family history of PrCa to aggressive PrCa but other 59 SNPs may not have such a predisposition. This postulate was supported by the recent finding from Chen et al study [25] that the incorporation of GRS to family history can improve the detection of aggressive PrCa. However, this deserves a further research to verify.

Finally, although our risk stratification by combing PSA and SNP information can provide an efficient personalized preventive strategy by reducing false negative results

and also false positive findings, the incorporation of genetic information may involve substantial costs. It is of great concern over whether improved performance in early detection can outweigh the cost incurred by the genetic testing, particularly when the unit cost of such genetic testing at population level would be reduced due to an economic scale. However, this requires a formal cost-effectiveness analysis for the evaluation of the net balance between costs from genetic testing and benefits from early detection.

In conclusion, the expedient use of multiple genetic variants in seven chromosomal regions associated with PrCa risk together with information on PSA through a Bayesian reasoning algorithm improves risk stratification, which could provide the basis for risk-adapted PrCa screening to maximize its benefits and minimize the harms.

Reference

1. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet 2006; 38:652-658
2. Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet 2007;39:638-644
3. Yeager M, Orr N, Hayes RB, et al: Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 2007; 39:645-649
4. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 2007; 39:631-637
5. Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet 2007; 39:977-983.
6. Zheng SL, Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer. N Engl J Med 2008; 358:910-919.
7. Little J, Wilson B, Carter R, et al. Multigene panels in prostate cancer risk assessment: a systematic review. Genet Med. 2016;18:535-544.
8. Schröder FH, Hugosson J, Roobol MJ, et al: Prostate-cancer mortality at 11 years of follow-up. N Engl J Med 2012; 366:981-990.
9. Andriole GL, Crawford ED, Grubb RL 3rd, et al. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med. 2009;360:1310-1319.
10. Kilpeläinen T, Tammela T, Malila N, et al: Prostate cancer mortality in the Finnish randomized screening trial. J Natl Cancer Inst 2013;105:719-725.
11. Finne P, Stenman UH., Määttänen L et al. The Finnish trial of prostate cancer screening: where are we now? BJU Int 2003; 92:22-26.
12. Wu GH, Auvinen A, Yen AMF, et al. A Stochastic Model for Survival of Early Prostate Cancer with Adjustments for Leadtime, Length Bias, and Over-detection. Biom J 2012; 54:20-44.
13. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genomewide association study of prostate cancer. Nat Genet 2008;40: 310-315.

14. Ewing CM, Ray AM, Lange EM, et al. Germline mutations in HOXB13 and prostate-cancer risk. N Engl J Med 2012;12;366:141-149.
15. Laitinen VH, Wahlfors T, Saaristo L, et al. HOXB13 G84E mutation in Finland: population-based analysis of prostate, breast, and colorectal cancer risk. Cancer Epidemiol Biomarkers Prev 2013; 22:452-460.
16. Pashayan N, Pharoah PD, Schleutker J, et al. Reducing overdiagnosis by polygenic risk-stratified screening: findings from the Finnish section of the ERSPC. Br J Cancer. 2015;113:1086-1093.
17. Schumacher FR, Al Olama AA, Berndt SI et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018;50:928-936.
18. Bao BY, Pao JB, Lin VC, et al. Individual and cumulative association of prostate cancer susceptibility variants with clinicopathologic characteristics of the disease. Clin Chim Acta. 2010 6;411(17-18):1232-1237
19. Kote-Jarai Z, Mikropoulos C, Leongamornlert DA, et al. Prevalence of the HOXB13 G84E germline mutation in British men and correlation with prostate cancer risk, tumour characteristics and clinical outcomes. Ann Oncol. 2015;26:756-61
20. Djavan B, Zlotta A, Remzi M, et al: Optimal predictors of prostate cancer on repeat prostate biopsy: a prospective study of 1,051 men. J Urol 2000; 163:1144-1148.
21. Hara I, Miyake H, Hara S, et al: Significance of prostate-specific antigen--alpha(1)-antichymotrypsin complex for diagnosis and staging of prostate cancer. Jpn J Clin Oncol 2001; 31:506-509.
22. Punglia RS, D'Amico AV, Catalona WJ, et al: Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. N Engl J Med 2003; 349:335-342.
23. Jacobsen SJ, Bergstralh EJ, Guess HA, et al. Predictive properties of serum-prostate-specific antigen testing in a community-based setting. Arch Intern Med 1996; 156:2462-2468.
24. Morgan TO, Jacobsen SJ, McCarthy WF, et al. Age-specific reference ranges for prostate-specific antigen in black men. N Engl J Med 1996; 335:304-310.

25. Chen H, Liu X, Brendler CB, et al. Adding genetic risk score to family history identifies twice as many high-risk men for prostate cancer: Results from the prostate cancer prevention trial. Prostate. 2016;76:1120-1129.

FIGURE LEGENDS

Figure 1. Summary of the estimates of interest, the use of model and distribution, and data sources.

Figure 2. Posterior odds of prostate cancer by age with or without considering seven SNPs Finnish Study.

Figure 3. Receiver operating characteristic curves for prostate cancer based on PSA alone and PSA plus genetic data (seven SNPs).

Figure 4. Receiver operating characteristic curves for prostate cancer (A) and aggressive prostate cancer (B) based on PSA alone, PSA plus risk socre with seven SNPs, and PSA plus polygenetic risk score.

Figure 5. Receiver operating characteristic curves for external validation based on four common SNPs(rs1859962, rs16901979, rs6983267, and rs144729

Table 1. Posterior odds of prostate cancer by PSA level based on seven SNPs, the Finnish prostate cancer screening trial

PSA Level	Men younger 60 years				Men aged 63-71 years			
	$P(PSA\|D)/P(PSA\|\overline{D})$ Likelihood Ratio given PSA (A)	$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|\overline{D})}$ SNP-specific risk (B)	Posterior Odds by combing PSA and 7 SNPs (C)		$P(PSA\|D)/P(PSA\|\overline{D})$ Likelihood Ratio given PSA (A)	$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|\overline{D})}$ SNP-specific risk (B)	Posterior Odds by combing PSA and 7 SNPs (C)	
	Estimate	Estimate	Estimate	95%CI	Estimate	Estimate	Estimate	95%CI
4.0	3.8	2.8	3.7	(1.6-10)	1.39	2.8	1.3	(0.6-3)
4.5	5.5	2.8	5.4	(2.2-15.3)	1.86	2.8	1.8	(0.8-4.2)
5.0	7.7	2.8	7.6	(3-23.3)	2.42	2.8	2.3	(1.1-5.6)
5.5	10.7	2.8	10.5	(3.9-33.6)	3.10	2.8	2.9	(1.3-7.5)
6.0	14.5	2.8	14.2	(5.2-47.5)	3.89	2.8	3.7	(1.6-10)
6.5	19.3	2.8	18.8	(6.5-65.7)	4.82	2.8	4.6	(2-12.7)
7.0	25.3	2.8	24.7	(8.4-89.1)	5.90	2.8	5.6	(2.4-16)
7.5	32.7	2.8	31.9	(10.3-122.6)	7.16	2.8	6.8	(2.8-20.1)
8.0	41.8	2.8	40.7	(12.9-157.4)	8.60	2.8	8.2	(3.2-24.7)
8.5	52.9	2.8	51.5	(15.6-207)	10.25	2.8	9.7	(3.8-30.2)
9.0	66.1	2.8	64.9	(18.9-267)	12.14	2.8	11.5	(4.4-36.7)
9.5	81.9	2.8	79.8	(22.6-348.4)	14.27	2.8	13.6	(5-44.5)
10.0	100.8	2.8	98.2	(27.28-437.5)	16.64	2.8	15.7	(5.7-54)

Considering seven SNPs (rs4242382 & rs10486567 & rs16901979 & rs6983267& rs138213197 & rs1447295 & rs1859962) from the Finnish DNA study
The likelihood ratios: 1.88 (95% CI:1.42-2.49) for rs4242382, 1.68 (95% CI:1.35-2.09) for rs10486567, 1.45 (95% CI:1.18-1.77) for rs1601979, 1.54 (95% CI:1.36-1.74) for rs6983267, 8.98 (95% CI:5.51-14.65) for rs138213197, 1.93 (95% CI:1.46-2.56) for rs1447295, and 1.42 (95% CI:1.25-1.61) for rs1859962,
$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+|\overline{D})}$:exp (0.0438*log(1.88) +0.933*log(1.68) +0.0818*log(1.45) +0.285*log(1.54) +0.0367*log(8.98) +0.0441*log(1.93) +0.7558*log(1.42)) =2.8

$(C) = \frac{P(D)}{P(\overline{D})} \times (A) \times (B)$

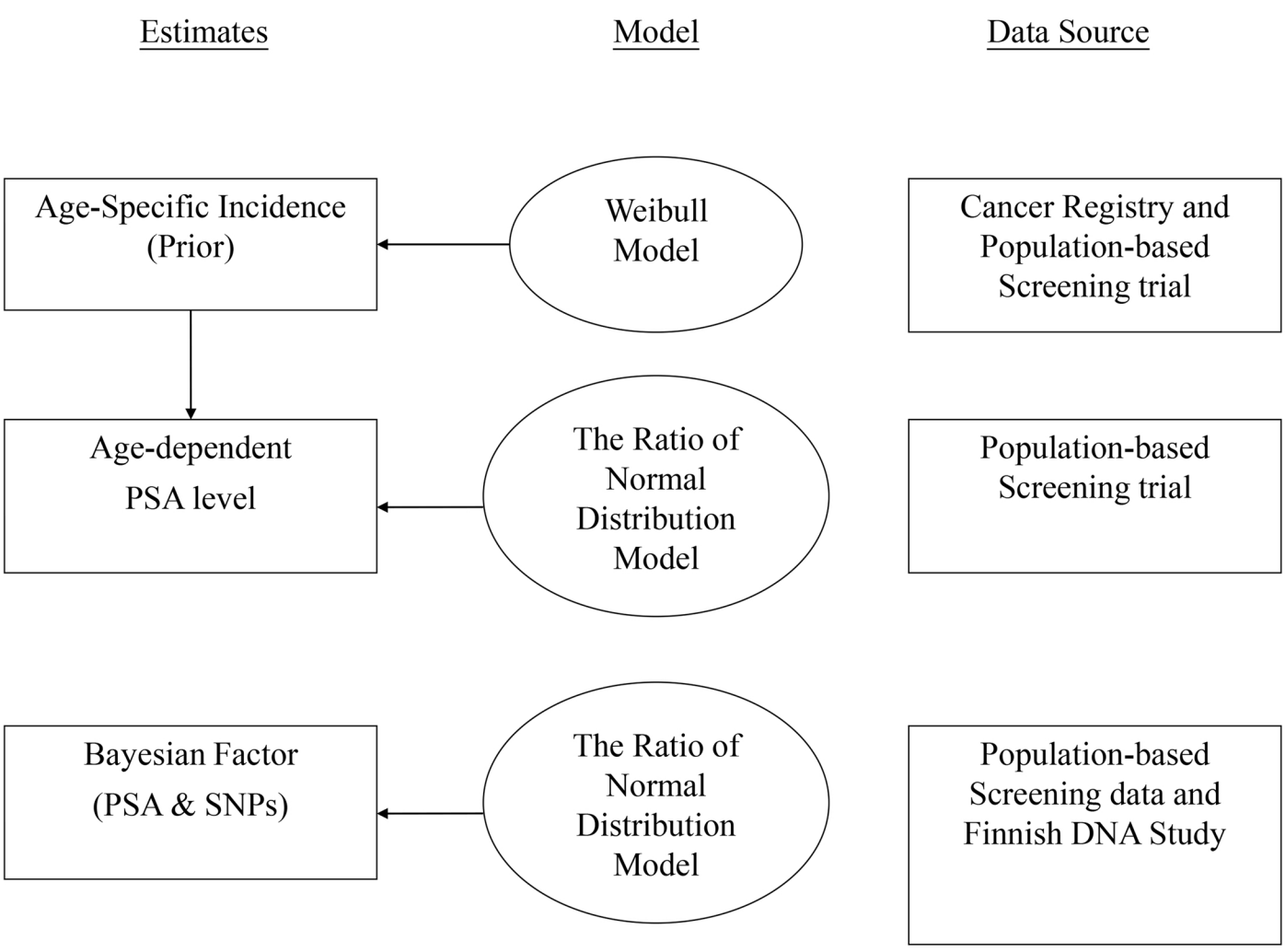
Table 2. Posterior odds of prostate cancer by PSA level based on five SNPs data from Zheng's study

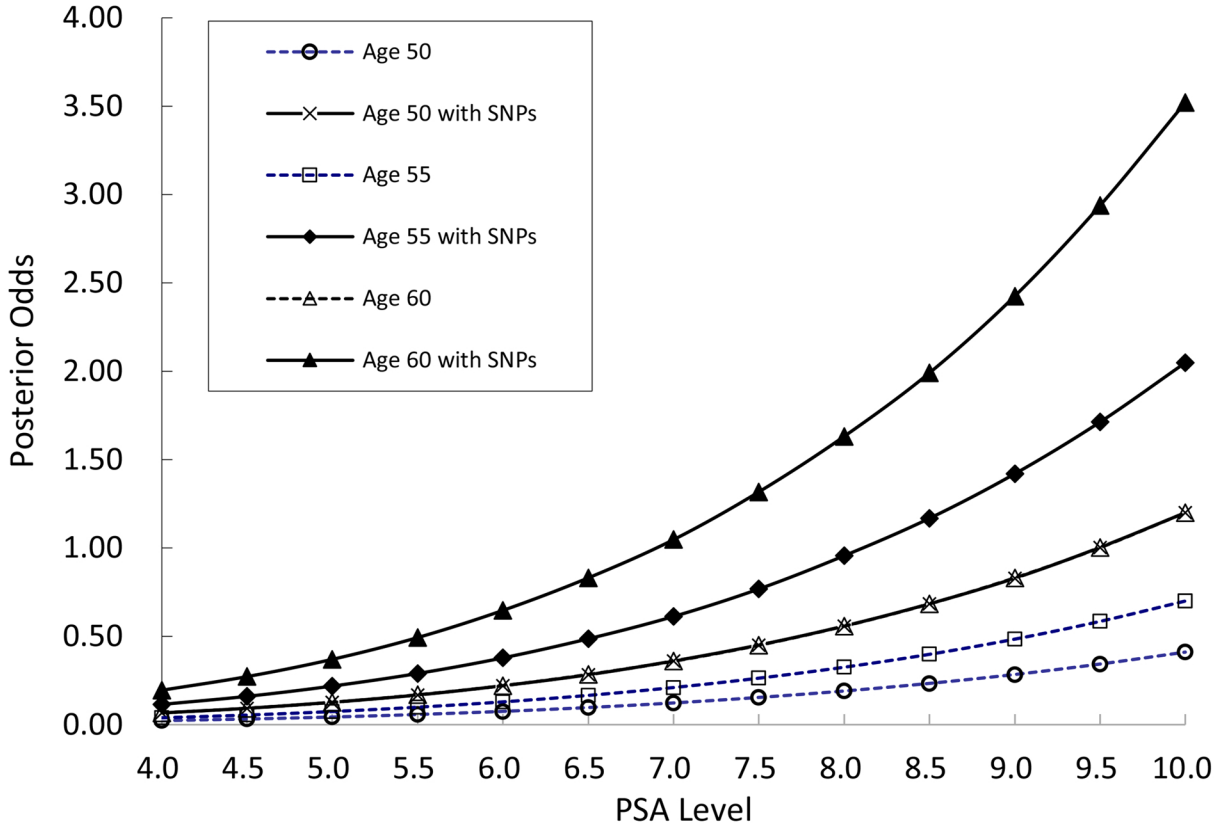
PSA Level	Men younger 60 years				Men aged 63-71 years			
	P(PSA|D)/P(PSA|$\overline{D}$) Likelihood Ratio given PSA (A)	$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|\overline{D})}$ SNP-specific risk (B)	Posterior Odds by combing PSA and 5 SNPs# (C)		P(PSA|D)/P(PSA|$\overline{D}$) Likelihood Ratio given PSA (A)	$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|\overline{D})}$ SNP-specific risk (B)	Posterior Odds by combing PSA and 5 SNPs# (C)	
	Estimate	Estimate	Estimate	95%CI	Estimate	Estimate	Estimate	95%CI
4.0	3.8	1.7	2.3	(1-6.1)	1.39	1.7	0.8	(0.4-1.8)
4.5	5.5	1.7	3.4	(1.4-9.5)	1.86	1.7	1.1	(0.6-2.6)
5.0	7.7	1.7	4.8	(1.9-14.4)	2.42	1.7	1.5	(0.7-3.5)
5.5	10.7	1.7	6.7	(2.6-21)	3.10	1.7	1.9	(0.9-4.6)
6.0	14.5	1.7	9.0	(3.4-29)	3.89	1.7	2.3	(1.1-6.1)
6.5	19.3	1.7	11.9	(4.3-41.4)	4.82	1.7	2.9	(1.3-7.8)
7.0	25.3	1.7	15.7	(5.4-55.8)	5.90	1.7	3.5	(1.5-9.9)
7.5	32.7	1.7	20.3	(6.7-75.5)	7.16	1.7	4.3	(1.8-12.5)
8.0	41.8	1.7	26.0	(8.3-100.2)	8.60	1.7	5.2	(2.1-15.4)
8.5	52.9	1.7	32.7	(10.1-129.9)	10.25	1.7	6.2	(2.5-18.7)
9.0	66.1	1.7	40.8	(12.3-168.2)	12.14	1.7	7.3	(2.8-23.1)
9.5	81.9	1.7	50.7	(14.7-217.1)	14.27	1.7	8.6	(3.3-27.5)
10.0	100.8	1.7	62.4	(17.74-271)	16.64	1.7	10.0	(3.7-33.1)

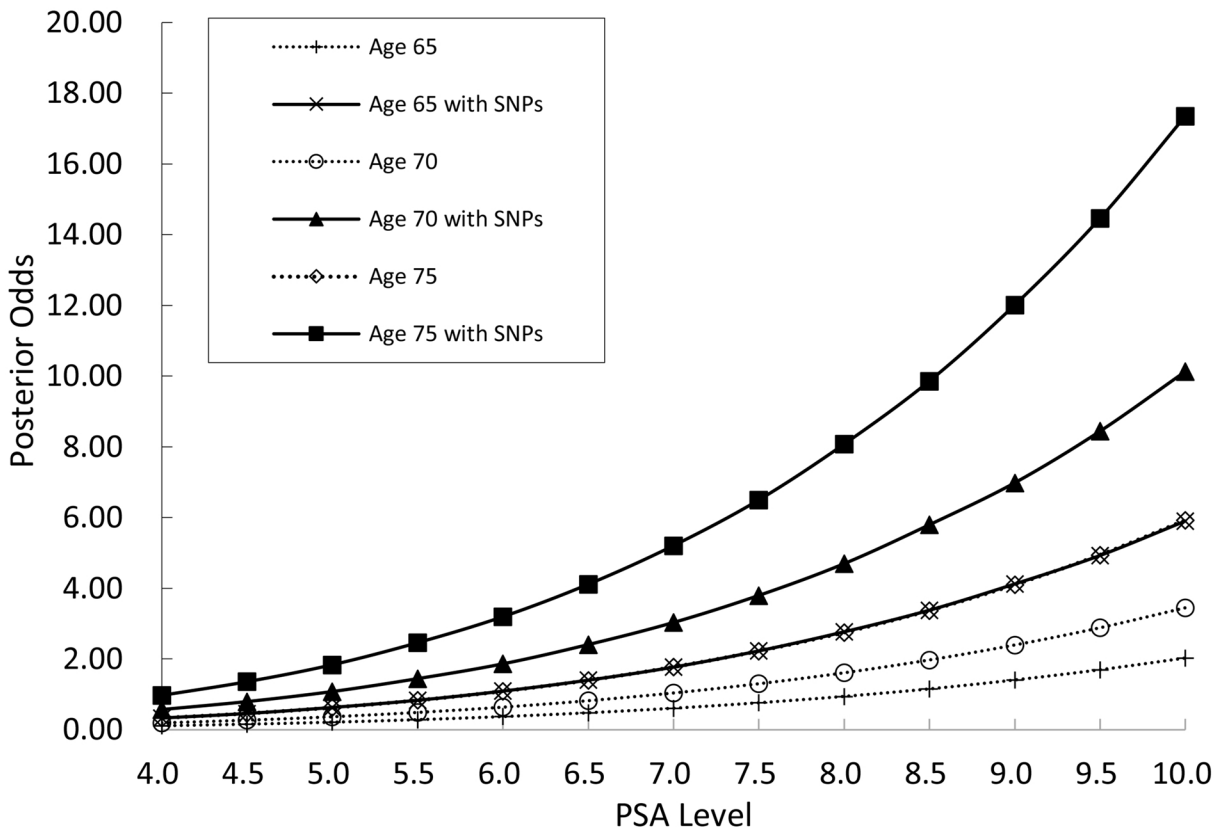
Considering five SNPs (rs4430796, rs1859962, rs16901979, rs6983267, and rs1447295) from Zheng's study

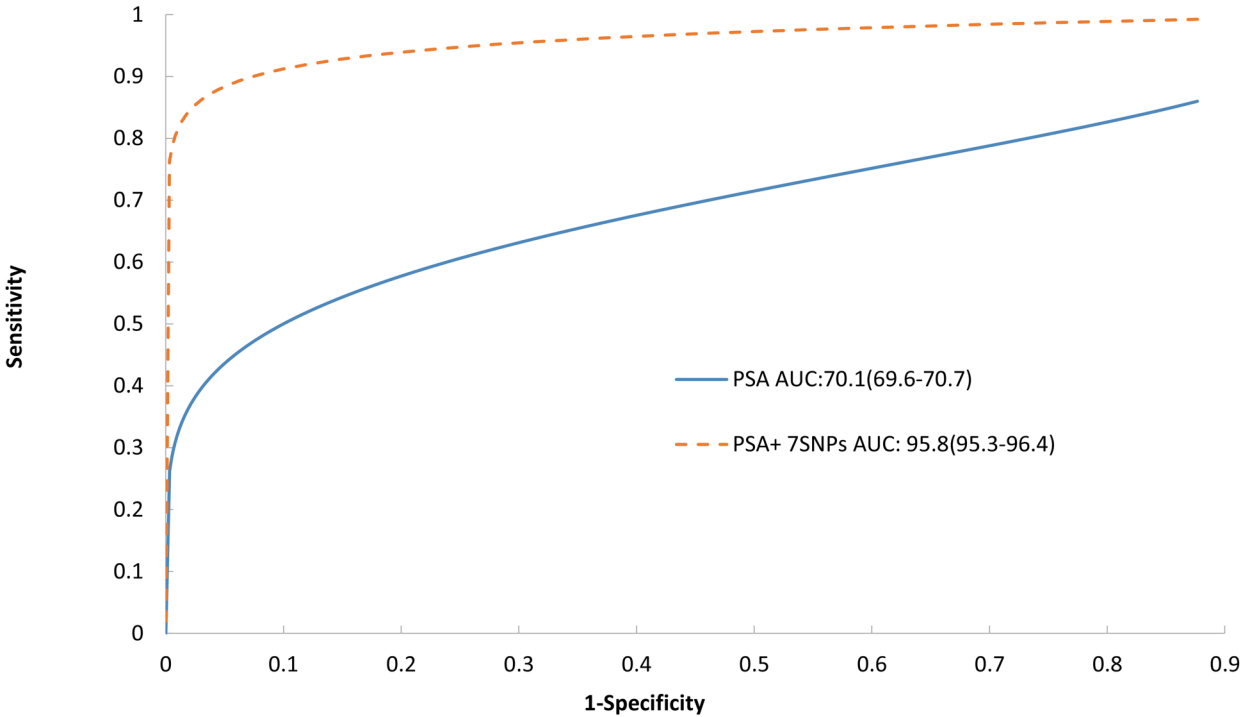
$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_5^+|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_5^+|\overline{D})}$:exp(0.56*log(1.38)+0.5*log(1.28)+0.03*log(1.53)+0.51*log(1.37)+0.14*log(1.22)) =1.7

(C) $= \frac{P(D)}{P(\overline{D})} \times (A) \times (B)$


Estimates
Model
Data Source
Age-Specific Incidence (Prior)
Weibull Model
Cancer Registry and Population-based Screening trial
Age-dependent PSA level
The Ratio of Normal Distribution Model
Population-based Screening trial
Bayesian Factor (PSA & SNPs)
The Ratio of Normal Distribution Model
Population-based Screening data and Finnish DNA Study

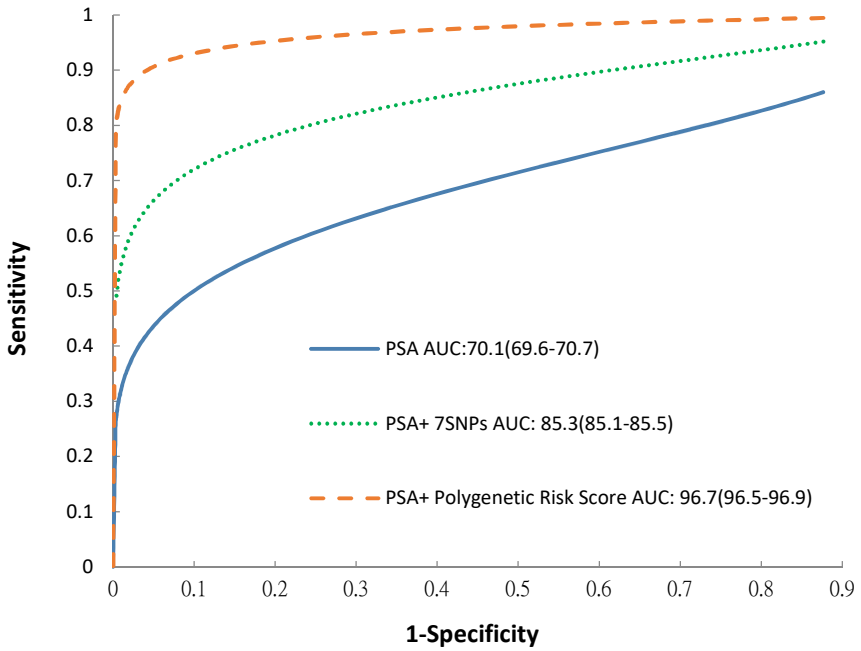

Age 50
Age 50 with SNPs
Age 55
Age 55 with SNPs
Age 60
Age 60 with SNPs
Posterior Odds
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
PSA Level


Age 65
Age 65 with SNPs
Age 70
Age 70 with SNPs
Age 75
Age 75 with SNPs
Posterior Odds
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
PSA Level

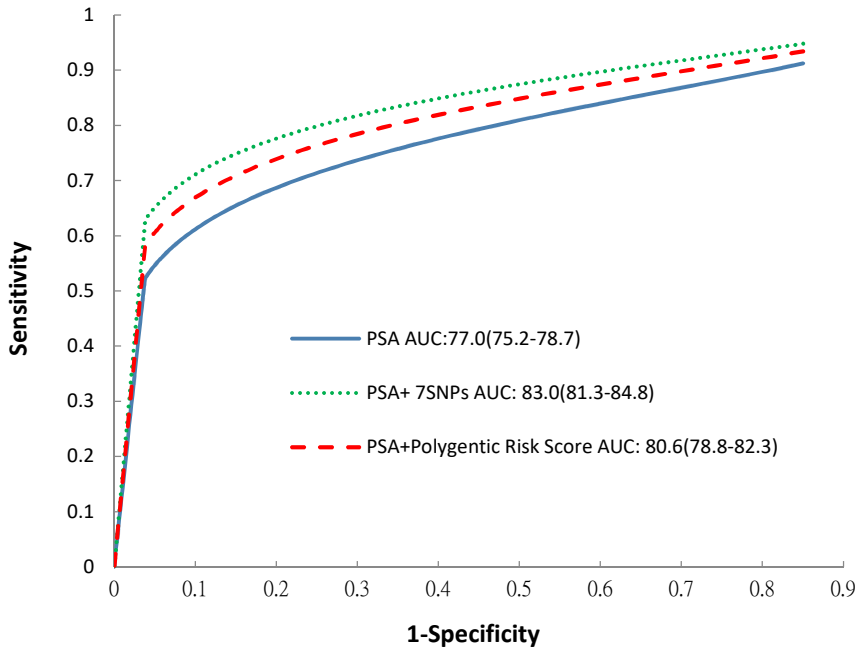


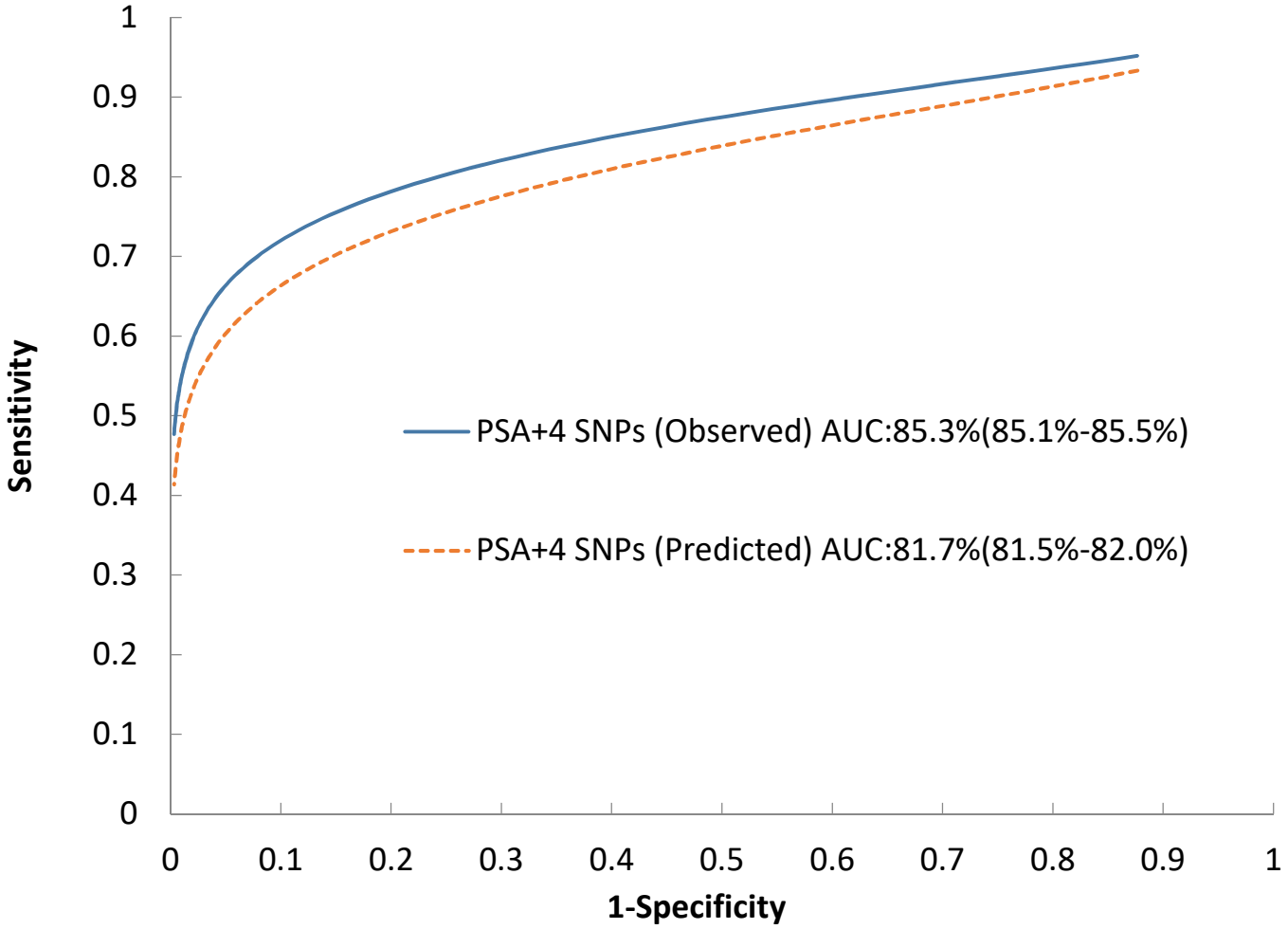
Optimal PSA Cut-off (PSA): 9.1 corresponding with 50.4% of sensitivity and 89.6% of specificity
Optimal PSA Cut-off (PSA+7 SNPs): 10.7 corresponding with 87.5 % of sensitivity and 95.8% of specificity

(A)


Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1-Specificity
PSA AUC:70.1(69.6-70.7)
PSA+ 7SNPs AUC: 85.3(85.1-85.5)
PSA+ Polygenetic Risk Score AUC: 96.7(96.5-96.9)

(B)


Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1-Specificity
PSA AUC:77.0(75.2-78.7)
PSA+ 7SNPs AUC: 83.0(81.3-84.8)
PSA+Polygentic Risk Score AUC: 80.6(78.8-82.3)



Abbreviations

AUC	Area Under Curve
PSA	Prostate-Specific Antigen
PrCa	Prostate Cancer
ROC	Receiver Operating Characteristic
SNPs	Single Nucleotide Polymorphisms

Prediction of the risk of PrCa with SNPs and PSA levels by using Bayesian clinical reasoning

We adopted a Bayesian clinical reasoning to estimate PSA- and SNP-based posterior odds for PrCa by updating the baseline risk of PrCa (prior) with the likelihood ratios between PrCa positive and negative men formed by the two corresponding distributions of PSA and the other likelihood ratio based on SNPs contribution, which is equivalent to the ratio of sensitivity to false positive, yielding the ROC curve. The posterior odds of developing prostate cancer given a specific PSA level and the SNPs of interests by the Bayesian algorithm considering different scenarios are derived as follows:

(1) With seven SNPs

$$\frac{P(D|SNP_1^+, SNP_2^+, \dots, SNP_7^+, PSA_D)}{(\bar{D}|SNP_1^+, SNP_2^+, \dots, SNP_7^+, PSA_{\bar{D}})} = \frac{P(D)}{P(\bar{D})} \times \frac{P(PSA_D|D, SNP_1^+, SNP_2^+, \dots, SNP_7^+)}{P(PSA_{\bar{D}}|\bar{D}, SNP_1^+, SNP_2^+, \dots, SNP_7^+)} \times \frac{P(SNP_1^+, SNP_2^+, \dots, SNP_7^+|D)}{P(SNP_1^+, SNP_2^+, \dots, SNP_7^+|\bar{D})}$$

, where D represents the event of prostate cancer, and $\bar{D}$ is the complement of D (non-disease). P(D) is prior probability of prostate cancer and P($\bar{D}$) is prior probability of being free of prostate cancer.

Assume PSA level is the conditionally independent of SNP once the disease status is determined. The formula can be simplified as $\frac{P(PSA_D|D)}{P(PSA_{\bar{D}}|\bar{D})}$

Let PSA_D and $PSA_{\bar{D}}$ denote PSA in men with and without prostate cancer. Both follow the two normal distributions, indicated by N(u_D , σ_D^2) and N($u_{\bar{D}}$, $\sigma_{\bar{D}}^2$); the likelihood ratio then becomes

$$\frac{P(PSA_D|D, SNP_1^+, SNP_2^+, \dots, SNP_7^+)}{P(PSA_{\bar{D}}|\bar{D}, SNP_1^+, SNP_2^+, \dots, SNP_7^+)} = \sqrt{\frac{\sigma_{\bar{D}}}{\sigma_D}} \times exp\left\{-\frac{1}{2}\left[\left(\frac{PSA_D - u_D}{\sigma_D}\right) - \left(\frac{PSA_{\bar{D}} - u_{\bar{D}}}{\sigma_{\bar{D}}}\right)\right]\right\}$$

u_D : the average estimate of PSA for prostate cancer cases

σ_D : standard deviation of PSA for prostate cancer cases

$u_{\overline{D}}$: average PSA for prostate cancer free men

$\sigma_{\overline{D}}$: standard deviation of PSA for prostate cancer free men

Appendix Table 1-1: The distribution of PSA with log transformation among men with and without prostate cancer in the Finnish prostate cancer screening trial

Age	Free of Prostate Cancer		Prostate Cancer		
	N	log(PSA), Mean(SD)	N	log(PSA), Mean(SD)	p-Value*
Age 55-59	399	1.742(0.325)	227	2.156 (0.705)	<0.0001
Age 63-71	719	1.801(0.346)	388	2.251(0.801)	<0.0001
Overall	1118	1.780(0.340)	615	2.216(0.768)	<0.0001

* Adjusting for age

Appendix Table 1-2: The distribution of PSA with log transformation among men with and without aggressive prostate cancer in the Finnish prostate cancer screening trial

Age	Free of Aggressive Prostate Cancer		Aggressive Prostate Cancer		
	N	log(PSA), Mean(SD)	N	log(PSA), Mean(SD)	p-Value*
Age 55-59	590	1.842(0.455)	36	2.718 (0.941)	<0.0001
Age 63-71	1025	1.892(0.478)	82	2.786(1.055)	<0.0001
Overall	1615	1.874(0.470)	118	2.765(1.018)	<0.0001

* Adjusting for age

Appendix Table 2-1: The risk of prostate cancer for the SNP of rs4242382 from Finnish population

	Genotype			OR	P
	GG	GA	AA	AA vs. GA/GG	
Non-prostate Cancer	1721(70.5%)	646(26.5%)	74(3%)	1.88(1.42-2.49)	<0.0001
Prostate Cancer	1729(61.1%)	942(33.3%)	157(5.6%)		

Appendix Table 2-2: The risk of prostate cancer for the SNP of rs10486567 from Finnish population

	Genotype			OR	P
	GG	GA	AA	GG/GA vs. AA	
Non-prostate Cancer	1254(51.4%)	981(40.2%)	206(8.4%)	1.68(1.35-2.09)	<0.0001
Prostate Cancer	1685(59.6%)	996(35.2%)	147(5.2%)		

Appendix Table 2-3: The risk of prostate cancer for the SNP of rs16901979 from Finnish population

	Genotype			OR	P
	AA	AC	CC	AA/AC vs. CC	
Non-prostate Cancer	4(0.2%)	160(6.6%)	2277(93.3%)	1.45(1.18-1.77)	0.0003
Prostate Cancer	6(0.2%)	261(9.2%)	2561(90.6%)		

Appendix Table 2-4: The risk of prostate cancer for the SNP of rs6983267 from Finnish population

	Genotype			OR	P
	AA	AC	CC	AA/AC vs. CC	
Non-prostate Cancer	625(25.6%)	1233(50.5%)	583(23.9%)	1.54(1.36-1.74)	<0.0001
Prostate Cancer	530(18.7%)	1377(48.7%)	921(32.6%)		

Appendix Table 2-5: The risk of prostate cancer for the SNP of rs138213197 from Finnish population

	Genotype			OR	P
	CC	CT	TT	TT/CT vs. CC	
Non-prostate Cancer	2418(99.3%)	18(0.7%)	0(0%)	8.98(5.51-14.65)	<0.0001
Prostate Cancer	2572(93.7%)	171(6.2%)	1(0%)		

Appendix Table 2-6: The risk of prostate cancer for the SNP of rs1447295 from Finnish population

	Genotype			OR	P
	AA	AC	CC	AA vs. AC/CC	
Non-prostate Cancer	73(3%)	710(29.2%)	1652(67.8%)	1.93(1.46-2.56)	<0.0001
Prostate Cancer	159(5.6%)	990(35.1%)	1673(59.3%)		

Appendix Table 2-7: The risk of prostate cancer for the SNP of rs1859962 from Finnish population

	Genotype			OR	P
	AA	AC	CC	AA/AC vs. CC	
Non-prostate Cancer	681(27.9%)	1193(48.9%)	566(23.2%)	1.42(1.25-1.61)	<0.0001
Prostate Cancer	604(21.4%)	1421(50.4%)	797(28.2%)		

Appendix Table 3-1. Posterior odds of prostate cancer by PSA level based on risk score with seven SNPs, the Finnish prostate cancer screening trial

PSA Level	Men younger 60 years				Men aged 63-71 years			
	P(PSA|D)/P(PSA|$\overline{D}$) Likelihood Ratio given PSA (A)	$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|\overline{D})}$ SNP-specific risk (B)	Posterior Odds by combing PSA and 7 SNPs# (C)		P(PSA|D)/P(PSA|$\overline{D}$) Likelihood Ratio given PSA (A)	$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|\overline{D})}$ SNP-specific risk (B)	Posterior Odds by combing PSA and 7 SNPs# (C)	
	Estimate	Estimate	Estimate	95%CI	Estimate	Estimate	Estimate	95%CI
4.0	3.8	1.567	2.1	(0.9-5.6)	1.4	1.567	0.8	(0.4-1.7)
4.5	5.5	1.567	3.1	(1.3-8.7)	1.9	1.567	1.0	(0.5-2.3)
5.0	7.7	1.567	4.4	(1.8-13.2)	2.4	1.567	1.3	(0.6-3.2)
5.5	10.7	1.567	6.1	(2.3-19.2)	3.1	1.567	1.7	(0.8-4.2)
6.0	14.5	1.567	8.3	(3.1-26.5)	3.9	1.567	2.1	(1-5.5)
6.5	19.3	1.567	10.9	(3.9-37.8)	4.8	1.567	2.6	(1.2-7.1)
7.0	25.3	1.567	14.4	(4.9-51)	5.9	1.567	3.2	(1.4-9)
7.5	32.7	1.567	18.5	(6.1-69)	7.2	1.567	3.9	(1.7-11.4)
8.0	41.8	1.567	23.8	(7.6-91.7)	8.6	1.567	4.7	(2-14.1)
8.5	52.9	1.567	29.9	(9.3-118.9)	10.3	1.567	5.7	(2.3-17.1)
9.0	66.1	1.567	37.3	(11.3-153.9)	12.1	1.567	6.7	(2.6-21.1)
9.5	81.9	1.567	46.4	(13.5-198.6)	14.3	1.567	7.8	(3-25.1)
10.0	100.8	1.567	57.1	(16.2-248)	16.6	1.567	9.1	(3.4-30.3)

Considering seven SNPs (rs4242382 & rs10486567 & rs16901979 & rs6983267& rs138213197 & rs1447295 & rs1859962) from the Finnish DNA study

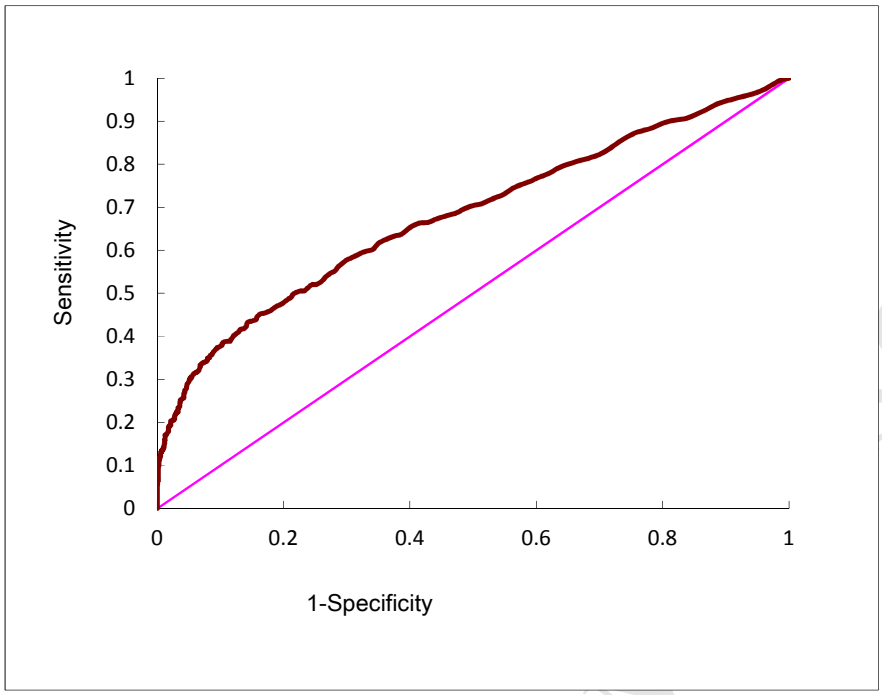
$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+|\overline{D})}$: exp(0.4492) =1.567

Appendix Table 3-2. Posterior odds of prostate cancer by PSA level based on Polygenetic Risk, the Finnish prostate cancer screening trial

PSA Level	Men younger 60 years				Men aged 63-71 years			
	$P(PSA\|D)/P(PSA\|\overline{D})$ Likelihood Ratio given PSA	$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|\overline{D})}$ Polygenetic Risk	Posterior Odds by combing PSA and Polygenetic Risk Score		$P(PSA\|D)/P(PSA\|\overline{D})$ Likelihood Ratio given PSA	$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+\|\overline{D})}$ Polygenetic Risk	Posterior Odds by combing PSA and Polygenetic Risk Score	
	Estimate	Estimate	Estimate	95%CI	Estimate	Estimate	Estimate	95%CI
4.0	3.8	3.12	4.2	(1.9-11.2)	1.39	3.12	1.5	(0.8-3.4)
4.5	5.5	3.12	6.2	(2.6-17.3)	1.86	3.12	2.0	(1-4.7)
5.0	7.7	3.12	8.8	(3.5-26.3)	2.42	3.12	2.6	(1.3-6.4)
5.5	10.7	3.12	12.1	(4.6-38.2)	3.10	3.12	3.4	(1.6-8.4)
6.0	14.5	3.12	16.4	(6.1-52.7)	3.89	3.12	4.2	(1.9-11)
6.5	19.3	3.12	21.7	(7.8-75.3)	4.82	3.12	5.3	(2.3-14.2)
7.0	25.3	3.12	28.6	(9.8-101.5)	5.90	3.12	6.5	(2.8-18)
7.5	32.7	3.12	36.9	(12.2-137.4)	7.16	3.12	7.8	(3.3-22.7)
8.0	41.8	3.12	47.4	(15.2-182.5)	8.60	3.12	9.4	(3.9-28.1)
8.5	52.9	3.12	59.5	(18.4-236.5)	10.25	3.12	11.3	(4.5-34)
9.0	66.1	3.12	74.3	(22.4-306.2)	12.14	3.12	13.3	(5.2-42)
9.5	81.9	3.12	92.3	(26.8-395.2)	14.27	3.12	15.6	(6-50)
10.0	100.8	3.12	113.6	(32.3-493.4)	16.64	3.12	18.1	(6.8-60.3)

The likelihood ratios for polygenetic risk score: 3.12(2.78-3.50)
$\frac{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+|D)}{P(SNP_1^+,SNP_2^+,\ldots,SNP_7^+|\overline{D})}$:exp (log(1.13) =3.12

Appendix Figure 1: Receive Operating Characteristic Curves for Prostate Cancer using traditional logistic regression analysis.



PSA ≥4, AUC : 0.68 (0.65-0.71)