Reliability of self-report versus chart-based prostate cancer, PSA, DRE and urinary symptoms

Eric C. Sayre, MSc,[1,2] Peter S. Bunting, PhD,[3,4] Jacek A. Kopec, PhD[2,5]

[1]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada

[2]Arthritis Research Centre of Canada, Vancouver, BC, Canada

[3]Department of Pathology and Laboratory Medicine, The Ottawa Hospital, Ottawa, Ontario, Canada

[4]Ottawa Health Research Institute, Ottawa, Ontario, Canada

[5]Department of Health Care and Epidemiology, University of British Columbia, Vancouver, BC, Canada

SAYRE EC, BUNTING PS, KOPEC JA. Reliability of self-report versus chart-based prostate cancer, PSA, DRE and urinary symptoms. The Canadian Journal of Urology. 2009;16(1):4463-4471.

Introduction: *Medical chart-review and self-reported questionnaire are two common methods of determining cancer screening and symptoms. We investigate the validity of these methods and therefore of a class of clinical/epidemiological studies. We compare variables on prostate cancer, any prostate-specific antigen (PSA) test, asymptomatic screening PSA, any digital rectal exam (DRE), and urinary symptoms. We used data from a 2005 case control study of PSA and metastatic prostate cancer (MPC) (253 cases and 496 controls). Data were collected from 1999 to 2002.*

Methods: *We calculated kappa, percent agreement (PPA) and prevalence adjusted bias adjusted kappa (PABAK). We compared percentage positive response (PPR) and sensitivities/specificities of questionnaire against chart and vice versa. We measured the degree of differential agreement between cases and controls using odds ratios.*

Results: *We found almost perfect agreement on prostate cancer, moderate agreement on any PSA and DRE, and slight agreement on asymptomatic screening PSA and urinary symptoms. PABAK ranged from 0.134 (urinary symptoms) to 0.879 (prostate cancer). Differences between cases/controls in PPR are similar according to chart or questionnaire, though PPR itself is usually higher on the questionnaire. Only for any PSA (including diagnostic), cases had better recall than controls. We found no evidence of differential agreement that might lead to bias in a case control study.*

Conclusions: *Some variables are more reliable than others comparing medical chart review and self-report. Diagnosis of prostate cancer has near perfect agreement, but for less catastrophic events such as PSA (especially asymptomatic screening tests), DRE or urinary symptoms, agreement ranges from slight to moderate.*

Key Words: prostate cancer, screening, prostate-specific antigen, digital rectal exam, urinary symptoms, medical chart review, self-report, validity, reliability

Introduction

There have been numerous studies of prostate cancer screening efficacy in reducing the risk of metastatic prostate cancer (MPC) and/or prostate cancer mortality.[1-13] Much publicity has gone to the debate about screening for prostate cancer with prostate-specific antigen (PSA) and digital rectal exam (DRE), in the scientific literature and public news media. Yet many men and prostate cancer patients have poor knowledge of screening.[14,15] Randomized controlled trials[8,9] are the best way to evaluate a screening test such as PSA. However, the results of the ongoing trials are still some years away. Other study designs that can evaluate PSA screening include ecological studies,[2,3,5-,7] analyzing administrative databases with lab tests,[4] medical chart review[1,11-13] and self-report.[1,16] Neither of the latter two methods is a clear gold standard. Self-reported screening data is error prone due to imperfect recall, misunderstanding of medical terms, insufficient information provided by the physician and other reasons. Chart-review is limited by how far back in time information is available, information loss upon changing doctors, that not all tests may be recorded, or that recorded tests may be missed by chart reviewers.[17] Misclassification of exposure or outcome

Accepted for publication October 2008

Address correspondence to Dr. Eric C. Sayre, Research Statistical Analyst, Arthritis Research Centre of Canada, 895 West 10th Avenue, Vancouver, BC V5Z 1L7 Canada

variables can lead to bias in analytical results from epidemiological or clinical studies and it is therefore important to understand the validity and reliability of these measures.[18,19] Reliable PSA screening data are also essential for monitoring screening rates and evaluating campaigns that attempt to increase rates of PSA screening, and accurate self-reported screening histories are important to physicians caring for new patients whose past medical charts detailing their screening behaviors are not always complete.[18,20,21]

Previous studies have investigated the reliability of medical chart review versus self-reported prostate cancer screening tests (DRE and/or PSA).[17-19,21-23] Other studies have compared chart review versus self-reported screening for other forms of cancer including mammograms and pap smears for breast and cervical cancer,[17,20,21,24,-33] and fecal occult blood testing and colonoscopies for colorectal cancer.[17,21,23,34,35] Sensitivities of self-reported prostate cancer screening tests in predicting these events in the medical chart ranged from 63%[23] to 92%[22] for DRE, and 69%[22] to 81%[19] for PSA. Specificities ranged from 41%[22] to 72%[23] for DRE, and 64%[22] to 87%[19] for PSA.

The purpose of our study is to compare self-reported versus chart review-based prostate cancer diagnosis, any PSA (including diagnostic), asymptomatic screening PSA, any DRE, and urinary symptoms including retention, frequent daytime urination, intermittent stream, urgency, weak stream, hesitancy and nocturia. We use data from a 2005 case control study of PSA and MPC by Kopec et al.[1] Our study differs from previous studies of reliability of self-reported data and chart review in several important ways. First, previous studies calculated sensitivity from the perspective of self-reported data predicting chart review (chart review was considered the gold standard). However, we have indicated some of the caveats of both sources of data, and as such we choose to study this question from both directions. Secondly, while previous studies have investigated the reliability between chart-review and self-reported data on PSA and DRE, to our knowledge this is the first to consider the reliability of data on asymptomatic screening PSA specifically (excluding diagnostic PSA tests), an important variable in studies of screening. Hiatt et al did report differences of asymptomatic screening rates between chart review and self-report in asymptomatic patients, however they reported sensitivities and other reliability measures on the global sample of symptomatic as well as asymptomatic patients.[23] Gordon et al investigated the reliability of reported reasons for testing (screening versus diagnostic) but reliability measures on the tests themselves were again on the global sample.[17] Finally, we compare differences in the reliability of data seen in cases of MPC versus that seen in controls. Significant differences between cases and controls in the reliability of their data could pose potential biases to case-control studies utilizing self-reported data, for example, if cases of MPC remembered their screening PSAs more readily than controls. Furthermore, in the Kopec et al study, charts of cases usually had far more investigations than those of controls. This can lead to loss of information in very large charts and/or lack of comparable investigations in control charts. The two sets of data in cases and controls may not be comparable, and it is therefore important to check reliability in both directions (chart and self-reported) and to compare reliability between cases and controls, something which previous studies have not considered.

Methods

Data collection

The study population consisted of all residents of Metropolitan Toronto and the five surrounding counties of Durham, Halton, Peel, York and Simcoe. This describes a population of 5.2 million (2001 Census of Canada) served by two Regional Cancer Centres (RCCs). Cases were men who developed metastatic prostate cancer between August 1, 1999 and May 31, 2002. Patients with metastases to the lymph nodes and those with local spread of the original tumor into adjacent organs were excluded because of the variable extent to which data on them was available, and they may or may not have had distant metastatic disease (which was defined as having a positive bone scan). Cases had to be 40 to 84 years old when diagnosed with MPC, diagnosed with prostate cancer on or after January 1, 1990, living in the study area at the time of diagnosis and able to answer a questionnaire in English. The date of diagnosis of prostate cancer was the date of biopsy. New cases of MPC were found by searching computerized lists of prostate cancer patients treated at the two RCCs. To ascertain cases that were not referred to RCCs, monthly contacts were maintained with the participating urologists (88 out of 90) and all the oncologists in the study area who would normally treat men with MPC. We estimate that > 80% of metastatic patients were seen in RCCs. All physicians were contacted and all available charts were reviewed, whether the patient was treated in an RCC or not. Note that information about PSA screening and symptoms prior to the diagnosis of prostate cancer came primarily from the charts of family physicians (not RCC charts) for both cases and controls since screening by definition does not apply to patients already diagnosed

with prostate cancer. The chart review was done by a visit to the physician's office to review the chart by trained chart abstractors. From the initial diagnosis of prostate cancer, the mean (median) time to questionnaire administration in cases was 3.7 years (3.1 years). From the subsequent diagnosis of metastatic prostate cancer, the mean (median) time to questionnaire administration was 152 days (114 days). There were 253 cases.

Controls were men without MPC selected randomly from the municipal tax records database for the study area and able to answer a questionnaire in English. There were 496 controls. Men known to have non-metastatic prostate cancer were eligible to be controls as long as their diagnosis was after January 1, 1990. This was allowed because in the original study MPC (being incurable) was used as a surrogate for death from prostate cancer, and many men live for many years with prostate cancer without developing MPC. Eleven controls developed prostate cancer during the study. Controls were sampled throughout the period of case accrual and matched to cases for age (5 year intervals) and region (8 regions). Controls were further matched on observation time to individual cases; their charts were searched for relevant events between January 1, 1990 and the date of diagnosis of prostate cancer for the matching case (the "reference date" for the case and his matched controls).

The response rate was 69% for cases, 51% for controls, which is fairly typical for this kind of study in North America, and medical records were obtained for 90% of respondents (93% for cases, 88% for controls). Less than 2% of cases and controls were excluded on the basis of language proficiency. No significant differences among cases and controls were found based on self-reported race, which was about 80% Caucasian.

Prostate cancer diagnosis, PSA, DRE and urinary symptoms were ascertained from the beginning of the patient history, and collected by both a review of the subject's medical chart up to the reference date, and via self-report on a questionnaire administered to the subject. All subjects signed a consent form for participation in the study. The study was approved by the University of Toronto Ethics Review Board and by the ethics review boards of the participating hospitals.

Self-reported prostate cancer was determined with the question, "Have you ever been diagnosed with prostate cancer?" If the answer was yes, a follow up question asked the month and year of diagnosis.

Self-reported PSA was determined with the question "Have you ever had a PSA test?" If yes, the reason for the first PSA test was asked, the choices being, "I had no symptoms, but I requested it to test for prostate cancer"; "I had no symptoms, but the doctor suggested it to test for prostate cancer"; "It was part of a routine check-up"; "It was ordered by the doctor to investigate symptoms I had at that time"; "It was ordered by the doctor to investigate the results of an X-ray or bone scan"; "It was ordered by the doctor to investigate the results of a rectal exam or ultrasound"; "Other reason for PSA test" and "Don't know". A self-reported PSA was considered an asymptomatic screening test if the reason was one of the first three. The date of the first PSA was also asked.

In the chart, all PSAs were recorded between January 1, 1990 and the reference date. Whether a PSA in the chart was screening or diagnostic was determined by analyzing the physician's notes in the chart.[36] We compared any PSAs between self-report and chart-based variables, as well as screening PSAs up to the reference date (if the self-reported date of the first PSA was after the reference date, then "self-reported PSA up to the reference date" was considered negative). (Asymptomatic screening PSA was the primary risk factor investigated in the case-control study.)

Self-reported history of DRE was determined by the question "How many times did you have a rectal exam, that is, an exam where the doctor checks your rectum with his finger?" Categorized frequencies were elicited for the periods 1980 to 1984, 1985 to 1989, and 1990 to 1994. We compared any DRE between 1980 and 1994 in the subgroup with reference dates on or after January 1, 1995 so that all subjects had an equal opportunity to have a DRE in this time period. This subgroup contains 363/496 controls and 189/253 cases.

We also compared urinary symptoms between the chart and responses to questions from a validated American Urological Association questionnaire.[37] Symptoms were indicated from the chart if any symptom in question was recorded up to December 31, 1994. Symptoms were indicated on the questionnaire if the selected response to the question "About 10 years ago, how often were you bothered by the following symptoms" was more than "not at all" (other possible answers ranged from "less than 1 time in 5" up to "almost always"). Symptoms were, with questionnaire wording in quotes and the symptom in the chart in parentheses: "a sensation of not emptying your bladder completely after urinating" (retention), "having to urinate again less than 2 hours after you finished" (frequent daytime urination), "stopping and starting again several times while urinating" (intermittent stream), "difficulty postponing urination" (urgency), "a weak urinary stream" (weak stream), "pushing or straining to begin urination" (hesitancy), "getting up to urinate during the night" (nocturia). This comparison

was made in the subgroup with reference dates on or after January 1, 1995.

Data analysis

Statistical analyses were done using SAS 9.1.3. All variables were compared on dichotomous scales. Measures of agreement between chart- and questionnaire-based dichotomous measures included simple kappa with 95% asymptotic confidence intervals (CIs), percent perfect agreement (PPA), prevalence adjusted bias adjusted kappa (PABAK), and Spearman's rank correlation coefficient with asymptotic p value. For PABAK, we use standard nomenclature for kappa: almost perfect agreement is > 80%, substantial is 61% to 80%, moderate is 41% to 60%, fair is 21% to 40%, slight is 0% to 20%, and poor is <0%.[38]

We compared the percentage positive response (PPR) to each variable between the chart and questionnaire. We also calculated the sensitivities/specificities of chart-based (or self-reported) outcomes given the outcome in self-reported (or chart-based) data. For positive outcomes this is "sensitivity", for negative outcomes it is "specificity".

Equally relevant to whether there are discrepancies between chart and questionnaire is whether the degree of discrepancy is the same in cases and controls. "Differential agreement" between cases and controls—a different degree of discrepancy between self-report and chart-review data in cases compared with the discrepancy in controls—might indicate a source of bias in a case-control analysis of that variable. For example, if cases are more likely to falsely report being screened by PSA, an analysis of screening efficacy would be biased against screening. To investigate differential agreement, odds ratios of case/control versus each chart-review variable were calculated for each fixed level of the questionnaire-based variable. ORs of case/control versus each questionnaire variable were calculated for each fixed level of the chart-review variable. These ORs measured differential agreement (OR = 1 indicates non-differential agreement, which is desirable). Asymptotic 95% CIs for the ORs were obtained.

TABLE 1. **Measures of agreement**

Variable subset (n)	**Percent perfect agreement (PABAK)**	**Simple Kappa (95% CI)**	**Spearman's rank correlation (p value)**
Prostate cancer			
Controls (n = 493)	91.9 (0.838)	0.330 (0.183, 0.478)	0.445 (< .001)
Cases (n = 251)	98.0 (0.960)	n/a	n/a
Combined (n = 744)	94.0 (0.879)	0.871 (0.835, 0.908)	0.876 (< .001)
Ever had a PSA test			
Controls (n = 446)	73.1 (0.462)	0.455 (0.372, 0.538)	0.455 (< .001)
Cases (n = 243)	79.0 (0.580)	0.260 (0.109, 0.410)	0.260 (< .001)
Combined (n = 689)	75.2 (0.504)	0.452 (0.382, 0.522)	0.452 (< .001)
Ever had a PSA and first was screening			
Controls (n = 496)	53.4 (0.069)	0.099 (0.029, 0.169)	0.123 (0.006)
Cases (n = 253)	69.6 (0.391)	0.236 (0.117, 0.355)	0.261 (< .001)
Combined (n = 749)	58.9 (0.178)	0.145 (0.084, 0.205)	0.172 (< .001)
[1]Had DRE between 1980 and 1994			
Controls (n = 359)	73.8 (0.476)	0.273 (0.170, 0.376)	0.309 (< .001)
Cases (n = 182)	67.6 (0.352)	0.281 (0.149, 0.413)	0.310 (< .001)
Combined (n = 541)	71.7 (0.434)	0.283 (0.202, 0.364)	0.316 (< .001)
[1]Prostate symptoms up to 1994 or about 10 years ago			
Controls (n = 363)	55.1 (0.102)	0.166 (0.081, 0.250)	0.193 (< .001)
Cases (n = 189)	59.8 (0.196)	0.194 (0.065, 0.324)	0.210 (0.004)
Combined (n = 552)	56.7 (0.134)	0.183 (0.112, 0.254)	0.207 (< .001)

[1]Subset with reference date on or after January 1, 1995

Results

Table 1 shows measures of agreement between chart and questionnaire. All variables are positively correlated between chart and questionnaire. Correlations (combined sample) range from 0.172 (screening PSA) to 0.876 (prostate cancer diagnosis), with any PSA test measuring 0.452 correlation. The correlation for DRE (0.316) is at the lower end of the previously published range, 0.306[17] to 0.710[21] (males), while for any PSA the correlation is in line with a previous study finding of 0.382.[18]

Kappa coefficients are significantly greater than 0. However, because of the skewed distributions, PABAK is more appropriate than simple kappa. The most extreme example is that amongst cases kappa is 0 for prostate cancer because all cases had chart-based prostate cancer by definition (shown as "n/a" in the table). Perhaps not surprisingly, there was almost perfect agreement for prostate cancer; patients are unlikely to mistake a diagnosis of prostate cancer for something else. PPA (PABAK) is 91.9% (0.838) in controls (11 developed prostate cancer) and 98.0% (0.960) in cases. In the combined sample it is 94.0% (0.879). There was moderate agreement for any PSA and any DRE, with PPA (PABAK) overall 75.2% (0.504) and 71.7% (0.434) respectively. These results fall near the middle of the range of previous study results for various samples (PPA of 57%[18] to 84.2%[19] for PSA and 59%[22] to 82%[22] for DRE). Asymptomatic screening PSA and urinary symptoms each had slight agreement, respectively PPA (PABAK) were 58.9% (0.178) and 56.7% (0.134).

Table 2 shows overall PPR to chart and questionnaire variables, as well as sensitivity/specificity of the questionnaire in predicting the chart, and the converse of that, sensitivity/specificity of the chart in predicting the questionnaire. By definition, all cases had prostate cancer in their charts. 2.2% of controls developed prostate cancer as indicated in the chart. Ninety-eight percent of

TABLE 2. **Overall percentages, sensitivities and specificities**

Variable subset (n)	Overall (%)		Sensitivity and specificity of chart predicting questionnaire (%)		Sensitivity and specificity of questionnaire predicting chart (%)	
	Chart	Quest.	Sens.	Spec.	Sens.	Spec.
Prostate cancer						
Controls (n = 493)	2.2	10.3	21.6	100.0	100.0	91.7
Cases (n = 251)	100.0	98.0	100.0	0.0	98.0	n/a
Combined (n = 744)	35.2	39.9	86.5	98.9	98.1	91.7
Ever had a PSA test						
Controls (n = 446)	56.7	54.9	77.1	68.2	74.7	71.0
Cases (n = 243)	81.9	84.0	86.3	41.0	88.4	36.4
Combined (n = 689)	65.6	65.2	81.3	63.7	80.8	64.6
Ever had a PSA and first was screening						
Controls (n = 496)	22.2	53.0	27.0	83.3	64.5	50.3
Cases (n = 253)	17.4	34.4	31.0	89.8	61.4	71.3
Combined (n = 749)	20.6	46.7	28.0	86.0	63.6	57.6
[1]Had DRE between 1980 and 1994						
Controls (n = 359)	69.4	86.1	75.1	66.0	93.2	30.0
Cases (n = 182)	58.8	78.0	66.9	70.0	88.8	37.3
Combined (n = 541)	65.8	83.4	72.5	67.8	91.9	33.0
[1]Prostate symptoms up to 1994 or about 10 years ago						
Controls (n = 363)	37.7	65.6	44.5	75.2	77.4	41.6
Cases (n = 189)	30.7	49.7	40.4	78.9	65.5	57.3
Combined (n = 552)	35.3	60.1	43.4	76.8	73.8	47.3

[1]Subset with reference date on or after January 1, 1995

cases reported prostate cancer, and all 11 controls with prostate cancer reported it on the questionnaire; 8.3% of controls without prostate cancer in the chart (erroneously) reported being diagnosed with prostate cancer; 81.9% of cases had any PSA in the chart while only 56.7% of controls had one. However, counting only first PSAs that are screening, the numbers are reversed: 22.2% of controls and 17.4% of cases have charts indicating this. On the questionnaire the numbers are similar: 84.0% of cases reported any PSA test while only 54.9% of controls reported one, and counting only first PSA tests that are screening, 53.0% of controls and 34.4% of cases reported a PSA. In the combined sample of cases/controls, PPR was higher for self-reported variables than for chart-based, considering PSA when first is screening, DRE between 1980 and 1994, and urinary symptoms up to 1994. This same over reporting (or under recording) of screening tests has been seen in previous studies of self-reported mammograms,[17,20,23,24,29,30] pap smears,[17,23,27-31] colonoscopy or fecal occult blood test,[17,22,23,35] sigmoidoscopy,[17,22,23] clinical breast examination[17,23] and male DRE,[17,18,22,23] with few exceptions.[21,32,33] Previous comparisons for PSA are mixed[18,19,22] due most likely to the definition of PSA that was not restricted to asymptomatic screening tests.

Table 2 helps to illustrate the tradeoff between sensitivity/specificity. Treating chart review as gold standard (last two columns), sensitivities of self-report in the combined sample (omitting prostate cancer) range from 63.6% (PSA test and first was for screening) to 91.9% (DRE between 1980 and 1994), and specificities in the combined sample range from 64.6% (any PSA) to 33.0% (DRE). Specifically, 91.9% of those with a DRE in the chart between 1980 and 1994 report a DRE in that time period (91.9% sensitivity). This falls at the high end of the previously published range (63%[23] to 92%[22]). Sixty-seven percent of those with no DRE in the chart reported having one (33.0% specificity), somewhat lower specificity than previous findings (41%[22]). Sensitivity of self-reported PSA (any test) is 80.8% and specificity 64.6%, which are also at the high end of sensitivity (69%[22] to 81%[19]) and the low end of specificity (64%[22] to 87%[19]).

Treating the questionnaire as the gold standard (middle two columns) produces almost exactly the same result for any PSA, but for screening PSA, DRE and urinary symptoms, sensitivities of the chart review for predicting self-report are lower and specificities higher than when chart review was the gold standard.

The odds ratios in Table 3 measure differential agreement (OR = 1 is non-differential agreement). ORs for self reporting a variable given presence of the variable in the chart (third column), range from 0.56 to 2.59. Cases are more likely than controls to recall any PSA test (OR = 2.59, 95% CI = 1.54, 4.35) (higher sensitivity), but there is no difference in recall of PSA

TABLE 3. **Odds ratios for case/control versus free variable (questionnaire or chart) for fixed levels of other variable (T = true, F = false)**

Variable (n)	OR (case versus chart = T | SR = T) (95% CI)	OR (case versus chart = T | SR = F) (95% CI)	OR (case versus SR = T | chart = T) (95% CI)	OR (case versus SR = T | chart = F) (95% CI)
[1]Prostate cancer (n = 744)	-	-	-	-
Ever had a PSA test (n=689)	1.862 (1.132, 3.064)	3.077 (1.522, 6.219)	2.591 (1.542, 4.353)	4.281 (2.151, 8.522)
Ever had a PSA and first was screening (n=749)	1.217 (0.717, 2.066)	0.568 (0.309, 1.043)	0.872 (0.424, 1.795)	0.407 (0.284, 0.583)
[2]Had DRE between 1980 and 1994 (n = 541)	0.671 (0.435, 1.036)	0.832 (0.34, 2.035)	0.580 (0.267, 1.261)	0.719 (0.387, 1.338)
[2]Prostate symptoms up to 1994 or about 10 years ago (n = 552)	0.845 (0.520, 1.372)	0.809 (0.427, 1.532)	0.556 (0.283, 1.089)	0.532 (0.344,0.822)

[1]Odds ratios are missing for case/control versus prostate cancer since all cases have prostate cancer in their charts
[2]Subset with reference date on or after January 1, 1995
SR = self-reported questionnaire

test where the first was for screening (OR = 0.87). The additional recall sensitivity in cases for any PSA test is therefore for diagnostic tests, which might be expected considering cases have MPC. Cases exhibited lower recall than controls of DRE (OR = 0.58, 95% CI = 0.27, 1.26), and urinary symptoms (OR = 0.56, 95% CI = 0.28, 1.09), though differences were not significant.

Since there is no gold standard between chart review and self-report, we also considered the OR between being a case and the chart-based variables, conditioning on fixed levels of the self-reported variable. For example, amongst those who self report having a PSA, cases are more likely than controls to have a PSA in their chart (OR = 1.86, 95% CI = 1.13, 3.06), but once again there is no difference in PSA where the first was for screening (OR = 1.22). The story is the same as before for DRE and urinary symptoms: no significant difference between cases and controls in chart-based indications among those who report these variables.

Discussion

In this paper we looked at reliability/validity of self-reported and chart-based indicators of prostate cancer diagnosis, any PSA, asymptomatic screening PSA, any DRE and urinary symptoms. Using data from a case-control study of PSA and MPC we measured overall agreement between chart review and self-report and compared agreement in cases versus controls. Spearman's rank correlation, simple kappa coefficients and PABAK showed that agreement was significantly better than random chance.

Among the variables studied, we found the highest agreement for prostate cancer diagnosis; in the combined sample, PPA (PABAK) was 94.0% (0.879), almost perfect agreement according to standard classification.[38] For any PSA, and DRE between 1980 and 1994 we found moderate agreement by the same classification, with PABAK values near the middle of the range of previously published results. PSA tests counted only when the first was for screening and urinary symptoms showed slight agreement. Self-reported data and chart review agree well on presence/absence of any (including diagnostic) tests or diagnosis of prostate cancer, but not as well on less tangible variables such as urinary symptoms or PSA tests counted only when the first was for screening. A similar phenomenon of recall related to the (perceived) importance of the procedure has been observed in a study of six cancer screening tests; they found that tests generating a report (e.g., mammogram, Pap smear, fecal occult blood test, and sigmoidoscopy) were recalled more easily than those generating only a physician's note (e.g., clinical breast examination and DRE).[34] However, this may also point to the uncertainty of medical chart review and the fact that it is not necessarily a gold standard compared to self-reported data, which would reinforce the need to compare each data source to the other on an equal basis.

Considering percent positive response, differences between cases and controls on various self-reported variables is similar to differences seen in the chart review variables. Though the actual PPR differs between chart and questionnaire with self reported rates generally higher than chart review counterparts, the direction of differences between cases and controls is preserved, so general effects found in a case-control study of these variables would be unlikely to depend on which method of data collection was used.

Without a clear gold standard, we calculated sensitivity/specificity of each method treating the other as the gold standard. To be expected, variables with higher sensitivity exhibit lower specificity. For example, the questionnaire had 91.9% sensitivity in detecting DRE when one was present in the chart; but it had only 33.0% specificity. For DRE and any PSA, sensitivities are at the high end and specificities at the low end of previous findings. With the questionnaire treated as gold standard instead of the chart, we find similar results for any PSA, but for screening PSA, DRE and urinary symptoms, sensitivities of the chart are lower and specificities higher than the questionnaire. This can be explained in part by the lower PPR in the chart.

We also measured differential agreement between cases and controls using ORs between being a case and self-reported variables given presence/absence of the variable in the chart (and as there is no gold standard between the chart and questionnaire, vice versa). ORs for self-reporting a variable given its presence in the chart ranged from 0.56 to 2.59. Cases were more likely than controls to recall any PSA, but there was no difference in recalling PSA where the first is for screening—the additional recall sensitivity in cases is for diagnostic tests. Cases exhibit moderately lower recall than controls of DRE between 1980 and 1994 and urinary symptoms, but the difference is not statistically significant. Considering associations in the other direction, among those who reported any PSA, cases were more likely than controls to have a PSA in their chart, but again there was no statistically significant difference in PSA tests where the first was for screening, DRE between 1980 and 1994, or urinary symptoms up to 1994. Other than diagnostic PSA tests, we find no evidence of differential agreement between cases and controls.

Limitations

This study has several limitations. First, while some of the questions were from validated or previously used questionnaires (e.g., the questions posed about urinary symptoms were from a validated American Urological Association questionnaire[37]), some new questions had to be developed for the case-control study. Please note that these are relatively simple questions about specific health behaviors and not multi-item psychometric scales. These questions had been pilot-tested but their reliability had not been ascertained. The purpose of the current study is to determine how reliable such questions are and to what extent errors in responses may affect the results of clinical and epidemiological studies of prostate cancer and PSA screening.

The response rate was 69% for cases and 51% for controls, and medical records were obtained for 90% of respondents (93% for cases, 88% for controls). There is always the possibility of some selection bias, though we do not anticipate that the small differences between cases and controls will play an additional factor.

We had not collected any information specific to hormone deprivation therapy (e.g., dose and length of time of hormonal therapy), however there are likely to be more cases than controls on hormonal therapy. Due to this treatment's effect on cognition, this could possibly bias cases' recall compared to controls. □

References

1. Kopec JA, Goel V, Bunting PS, Neuman J, Sayre EC, Warde P, Levers P, Fleshner N. Screening with prostate specific antigen and metastatic prostate cancer risk: a population based case-control study. *J Urol* 2005;174(2):495-499.
2. Hankey BF, Feuer EJ, Clegg LX, Hayes RB, Legler JM, Prorok PC et al. Cancer surveillance series: interpreting trends in prostate cancer--part I: Evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates. *Natl Cancer Inst* 1999;91(12):1017-1024.
3. Potosky AL, Feuer EJ, Levin DL. Impact of screening on incidence and mortality of prostate cancer in the United States. *Epidemiol Rev* 2001;23(1):181-186.
4. Weinmann S, Richert-Boe K, Glass AG, Weiss NS. Prostate cancer screening and mortality: a case-control study (United States). *Cancer Causes Control* 2004;15(2):133-138.
5. Lu-Yao G, Albertsen PC, Stanford JL, Stukel TA, Walker-Corkery ES et al. Natural experiment examining impact of aggressive screening and treatment on prostate cancer mortality in two fixed cohorts from Seattle area and Connecticut. *BMJ* 2002;325(7367):740.
6. Coldman AJ, Phillips N, Pickles TA.Trends in prostate cancer incidence and mortality: an analysis of mortality change by screening intensity. *CMAJ* 2003;168(1):31-35.
7. Perron L, Moore L, Bairati I, Bernard PM, Meyer F. PSA screening and prostate cancer mortality. *CMAJ* 2002;166(5):586-591.
8. Labrie F, Candas B, Dupont A, Cusan L, Gomez J-L, Subury RE et al. Screening decreases prostate cancer death: first analysis of the 1988 Quebec prospective randomized controlled trial. *The Prostate* 1999;38(2):83-91.
9. de Koning HJ, Auvinen A, Berenguer Sanchez A, Calais da Silva F, Ciatto S, Denis L, Gohagan JK et al. Large-scale randomized prostate cancer screening trials: program performances in the European Randomized Screening for Prostate Cancer trial and the Prostate, Lung, Colorectal and Ovary cancer trial. *Int J Cancer* 2002; 97(2):237-244.
10. Cronin KA, Weed DL, Connor RJ, Prorok PC. Case-control studies of cancer screening: theory and practice. *J Natl Cancer Inst* 1998;90(7):498-504.
11. Friedman GD, Hiatt RA, Quesenberry CP Jr, Selby JV. Case control study of screening for prostatic cancer by digital rectal examinations. *Lancet* 1991;337(8756):1526-1529.
12. Richert-Boe KE, Humphrey LL, Glass AG, Weiss NS. Screening digital rectal examination and prostate cancer mortality: a case control study. *J Med Screen* 1998;5(2):99-103.
13. Jacobsen SJ, Bergstralh EJ, Katusic SK, Guess HA, Darby CH, Silverstein MD et al. Screening digital rectal examination and prostate cancer mortality: a population based case control study. *Urology* 1998;52(2):173-179.
14. Mercer SL, Goel V, Levy IG, Ashbury FD, Iverson DC, Iscoe NA. Prostate cancer screening in the midst of controversy: Canadian men's knowledge, beliefs, utilization, and future intentions. *Can J Public Health* 1997;88(5):327-332.
15. Lamplugh M, Gilmore P, Quinlan T, Cornford P. PSA testing: are patients aware of what lies ahead? *Ann R Coll Surg Engl* 2006;88(3):284-288.
16. Eisen SA, Waterman B, Skinner CS, Scherrer JF, Romeis JC, Bucholz K, Heath A, Goldberg J, Lyons MJ, Tsuang MT, True WR. Sociodemographic and health status characteristics with prostate cancer screening in a national cohort of middle-aged male veterans. *Urology* 1999;53(3):516-522.
17. Gordon NP, Hiatt RA, Lampert DI. Concordance of self-reported data and medical record audit for six cancer screening procedures. *J Natl Cancer Inst* 1993;85(7):566-570.
18. Jordan TR, Price JH, King KA, Masyk T, Bedell AW. The validity of male patients' self-reports regarding prostate cancer screening. *Prev Med* 1999;28(3):297-303.
19. Volk RJ, Cass AR. The accuracy of primary care patients' self-reports of prostate-specific antigen testing. *Am J Prev Med* 2002;22(1):56-58.
20. Degnan D, Harris R, Ranney J, Quade D, Earp JA, Gonzalez J. Measuring the use of mammography: two methods compared. *Am J Public Health* 1992;82(10):1386-1388.
21. Montano DE, Phillips WR. Cancer screening by primary care physicians: a comparison of rates obtained from physician self-report, patient survey, and chart audit. *Am J Public Health* 1995;85(6):795-800.
22. Hall HI, Van Den Eeden SK, Tolsma DD, Rardin K, Thompson T, Hughes Sinclair A, Madlon-Kay DJ, Nadel M. Testing for prostate and colorectal cancer: comparison of self-report and medical record audit. *Prev Med* 2004;39(1):27-35.
23. Hiatt RA, Perez-stable EJ, Quesenberry Jr C, Sabogal F, Otero-Sabogal R, McPhee SJ. Agreement between self-reported early cancer detection practices and medical audits among Hispanic and non-Hispanic white health plan members in northern California. *Prev Med* 1995;24(3):278-285.
24. Fulton-Kehoe D, Burg MA, Lane DS. Are self-reported dates of mammograms accurate? *Public Health Rev* 1992;93(130):233-240.
25. Zapka JG, Bigelow C, Hurley T, Ford LD, Egelhofer J, Cloud M et al. Mammography use among sociodemographically diverse women: the accuracy of self-report. *Am J Public Health* 1996;86(7):1016-1021.

26. Sawyer JA, Fletcher RH, Daye FF, Wynn TM. Accuracy of women's self-report of their last Pap smear. *Am J Public Health* 1989;79(8):1036-1037.
27. Bowman JA, Redman S, Dickinson JA, Gibberd R, Sanson-Fisher RW. The accuracy of Pap smear utilization self-report: a methodological consideration in cervical screening research. *Health Serv Res* 1991;26(1):97-107.
28. Michielutte R, Dignan HB, Bahnson J, Smith M, Wooten R, Hale LN. Errors in reporting cervical screening among public health clinic patients. *J Clin Epidemiol* 1991;44(4):403-408.
29. Johnson CS, Archer J, Campos-Outcult D. Accuracy of Pap smear and mammogram self-reports in a southwestern Native American tribe. *Am J Prev Med* 1995;11(6):360-363.
30. Swarez L, Goldman DA, Weiss NS. Validity of Pap smear and mammogram self-reports in a low income Hispanic population. *Amer J Prev Med* 1995;11:94-98.
31. Bownan JA, Sanson-Fisher R, Redman S. The accuracy of self-reported Pap smear utilization. *Soc Sci Med* 1997;44(7):969-976.
32. Etzi S, Lane DS, Grimson R. The use of mammography vans by low-income women: the accuracy of self-reports. *Am J Public Health* 1994;84(1):107-109.
33. King ES, Rimer BK, Trock B, Balshem A, Engstrom P. How valid are mammography self-reports? *Am J Public Health* 1990;80(11):1386-1388.
34. Baier M, Calonge N, Cutter G, McClatchey M, Schoentgen S, Hines S et al. Validity of self-reported colorectal cancer screening behavior. *Cancer Epidemiol, Biomarkers Prev* 2000;9(2):229-232.
35. Mandelson MT, LaCroix AZ, Anderson LA, Nadel MR, Lee NC. Comparison of self-reported fecal occult blood testing with automated laboratory records among older women in a health maintenance organization. *Am J Epidemiol* 1999;150:617-662.
36. Bunting PS, Neuman J, Goel V, Fleshner N, Kopec J, Levers P, Warde P. A reliable coding system to define screening prostate-specific antigen tests was developed in a case-control study. *J Clin Epidemiol* 2005;58(6):639-644.
37. Barry MJ, Fowler FJ, O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, Cockett AT. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J Urol* 1992;148(5):1549-1557;discussion 1564.
38. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-174.