CLINICAL TRIALS *ARTICLE* *Clinical Trials* 2010; **7:** 303–311

Assessing contamination and compliance in the prostate component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial

Paul F Pinsky[a], *Amanda Blacka*[a], *Barnett S Kramer*[b], *Anthony Miller*[c], *Philip C Prorok*[a] *and Christine Berg*[a]

Background Recently, the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial published 7-year complete prostate cancer mortality results, which showed no benefit of screening with prostate specific antigen (PSA) and digital rectal examination (DRE). An issue of concern was the substantial level of 'contamination', or use of PSA and DRE in control arm men.
Purpose To provide a detailed description of contamination in PLCO.
Methods Surveys inquiring about the most recent PSA and DRE use were given to a sample of control arm men throughout the screening phase of PLCO (years 0–5). A probability model was utilized to translate survey results into actual frequency counts of tests. To assess the impact of contamination, Surveillance, Epidemiology, and End Results (SEER) incidence rates from the pre-screening era (1985–1987) as well as contemporaneous rates, were applied to PLCO person-years of observation.
Results Of 38,350 control arm men, 2427 were surveyed. Pre-trial screening and college education were statistically significantly associated with increased contamination rates. The estimated mean number of screening PSAs (DREs) in the control arm was 2.7 (1.1); this compares to 5.0 (3.5) in the screened arm. 1984 and 2538 prostate cancers were observed in the control and screened arms, respectively, during the screening phase. In the absence of screening, 960 and 949 would have been expected; with contemporaneous incidence rates, 1630 and 1611 were expected.
Limitations Due to the limitations of the surveys, in terms of both reach and scope, the exact level of PSA and DRE use in control arm men cannot be known.
Conclusions Use of prostate screening by control arm men was substantial, but also substantially less than in screened arm men. Detailed quantitative analyses of screening use across arms are critical for understanding current and future findings from the prostate component of PLCO. *Clinical Trials* 2010; **7**: 303–311. http://ctj.sagepub.com

Introduction

Recently, two high profile randomized trials of prostate cancer screening utilizing prostate specific antigen (PSA), the US-based Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, and the European Randomized Study of Screening for Prostate Cancer (ERSPC), published

[a]Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, [b]Office of Disease Prevention, National Institutes of Health, Bethesda, MD, USA, [c]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
Author for correspondence: Paul F Pinsky, Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. E-mail: pinskyp@mail.nih.gov

10.1177/1740774510374091

their early mortality results [1,2]. ERSPC reported a statistically significant mortality benefit of prostate cancer screening, with a relative risk in the screened arm of 0.80; in contrast, PLCO did not show any benefit, with a relative risk of around 1.1.

Differences in protocol and execution between the two trials included the timing of screens, the PSA cutoff utilized and whether digital rectal examination (DRE) was employed. There were also differences across countries and over time within ERSPC. Further, the reporting period cutoffs in terms of study years were different; 10 years for PLCO and up to 14 years for ERSPC. Nonetheless, one specific difference that has been highlighted by some commentators as contributing importantly to the disparity in results was the difference between the trials in contamination levels, that is, screening tests among control arm men. Contamination, based on surveys of a sample of control arm men, was estimated to be in the neighborhood of 50% in PLCO; the contamination rate in ERSPC, although not reported in the mortality manuscript, was projected to be 20% in the design stage of the trial [1,2].

Accurate prediction of contamination in the control arm, and of 'compliance' within the intervention arm (the rate of receiving the scheduled interventions), is critical in powering a trial correctly. More relevant to the question at hand, after a trial has reported findings showing no difference in outcomes between arms, accurate assessments are critical in assessing the true power of a trial to find a significant difference. In addition, they are critical with respect to interpreting the magnitude of any difference that was found. A 20% mortality reduction with minimal contamination would be interpreted differently and have different public health implications than the same magnitude of reduction in the presence of 50% contamination, since, for example, the cost–benefit tradeoffs would differ substantially between the two scenarios.

With a one-time, all or none, single modality intervention, measuring contamination (and compliance) is straightforward – either the subject received the intervention or not. However, in many trials, including PLCO and ERSPC, quantitative assessment of contamination and compliance is more problematic, since the intervention is not all or none. For example, in the prostate component of PLCO, intervention arm subjects were scheduled to receive six PSA tests and four DREs. How then should one measure compliance and contamination? By those receiving any test, at least one from each test, or all the protocol tests? It may not be possible to summarize either quantity adequately with a single number. Additionally, the contamination surveys only measured screening test use, not diagnostic follow-up; if screening tests were not appropriately followed up then observed contamination rates may have had little impact on control arm incidence rates and, potentially, on mortality rates.

We provide herein a more detailed analysis of contamination than reported previously [1]. First, we present a thorough analysis of the contamination survey results themselves, examining the entire range of responses. Additionally, we analyze the effect of demographic and screening history variables on contamination. Second, using modeling, we extrapolated the survey results, which asked only about the most recent test use, to estimate the actual counts and frequency of PSA (and DRE) tests performed during the screening phase of the trial in control arm men; in this way, the control and screened arms can be directly compared in terms of numbers of tests performed. Finally, in an attempt to examine the actual impact of control arm screening, we estimated the excess of prostate cancer cases attributable to screening in each arm; estimation was done by applying pre-PSA era population incidence rates to the person years (PY) at risk for PLCO men. This estimate can serve as a check on whether what the surveys were measuring had the expected impact when translated into cancer incidence.

Methods

The design of PLCO has been described previously [3]. Enrollment of men and women aged 55–74 was initiated in 1993 and completed in 2001 at 10 US screening centers. Each institution obtained IRB approval and all participants provided written informed consent before being randomized. The primary exclusion criteria were the history of a PLCO cancer, current cancer treatment and starting in April 1995, having had more than one PSA blood test in the prior 3 years. At study entry, participants completed a baseline questionnaire that inquired about demographics, medical history, and screening history, including the utilization of PSA and DRE testing in the 3-year period prior to PLCO enrollment. Subjects were to be followed for at least 13 years from randomization.

Men randomized to the intervention arm were offered annual PSA testing for 6 years (study years 0–5) and annual DRE testing for 4 years (study years 0–3). Men randomized prior to June, 1994 (about 4% of total) received PSA tests only at study years 0–3 and those randomized from June 1994 through May 1995 (about 17% of total) received PSAs at years 0–3 and at year 5 but not at year 4. A serum PSA level of >4.0 ng/mL was considered positive.

PLCO trial staff obtained medical records related to diagnostic follow-up of all positive screens, and medical record abstractors recorded information on diagnostic procedures and any cancer diagnoses. Information on cancers diagnosed in the absence of a positive screen was obtained from annual study update forms sent to each participant.

The utilization of screening tests outside of the trial protocol in the control arm, which is denoted as 'contamination', was assessed through periodic surveys of a sample of the participants. Surveys were performed annually from 1997 to 2007 except for the years 2000 and 2006. The survey consisted of a self-administered health status questionnaire (HSQ) that inquired about the use of 11 preventive health procedures, including the screening tests under evaluation in the trial. With respect to prostate screening, the lead questions were, 'Have you ever had a PSA blood test for prostate cancer' and 'Have you ever had a DRE of the prostate'. For each question, respondents who answered affirmatively were asked when they had the most recent test, with response categories of 'within the past year', '1–2 years ago', '2–3 years ago', and 'more than 3 years ago', and the main reason for having this test, with response categories 'because of a specific health problem', 'follow-up of a previous health problem', and 'as part of a routine health check-up'.

In the PLCO publication comparing prostate cancer mortality across arms, the definition of PSA (DRE) 'contamination' in the control arm was defined as having had a PSA (DRE) in the past year as part of a routine health check-up, to coincide with the screening protocol in the intervention arm [1]. Men who reported having received more than one PSA test in 3 years prior to trial entry ('baseline contaminated' men) were not included in the HSQ sampling frame during the screening years because they were *a priori* assumed to be contaminated (they were included in the post-screening years). The overall contamination rate was thus computed as a weighted average of the rate in surveyed men and the assumed 100% rate in the baseline contaminated men, who made up about 10% of the cohort.

Statistical methods

As mentioned above, the definition of contamination used in the PLCO mortality paper was receiving a PSA test as part of a routine examination within the past year, which approximates the average PSA compliance rate across all PLCO scheduled examinations. When the study intervention, as in PLCO, is not all or none, the measure used to quantitatively assess the degree of contamination in the control arm should be as much similar as possible to that used to assess compliance in the intervention arm, so that direct comparisons between arms are most meaningful. The closest to a direct comparison is to perform a 'virtual' survey in the screened arm analogous to that performed in control arm men. For each screened arm man we randomly chose a 'survey date' during the screening phase of the trial and assessed what proportion of screened arm men had a PSA screen within 1 year before this date. We assumed, as for the baseline contaminated men in the control arm, that the baseline contaminated men in the screened arm had PSA screens in the past year. Note that, due to the random survey times, the estimate has some variability; however, due to the large sample size, this variability is negligible (on the order of 0.1%).

Another way to directly compare test use between arms is to translate the control arm survey (HSQ) responses into actual counts of test use during the screening phase of PLCO (years 0–5). To this end, we utilized a relatively simple probability model that translates the survey response data into counts as above (see Appendix for details).

Note that a response of receiving a PSA test in the past year, even if representative of long-term behavior, does not necessarily translate into annual use (i.e., a count of six tests during the screening period); for example, if a man regularly received PSA screens every 2 years, there would be a 50% chance of his reporting on the survey that he had a PSA test in the past year. Models were run separately for (1) routine test use and (2) any use, and also separately for PSA and DRE. For DRE, since screened arm men were only examined for study years 0–3, the model also only considers that period. Baseline contaminated men (for PSA) were assumed, as before, to have had routine PSA tests each study year. For the screened arm, PLCO screens were counted as routine use; for calculating any use, diagnostic PSAs or DREs also were included.

We performed multiple logistic regression to assess the influence of various factors on PSA and DRE contamination in the control arm, as defined by routine use in the past year. The factors assessed were as follows: PSA test use prior to PLCO (0 or 1 time), DRE use prior to PLCO (0, 1, 2, or more times), family history of prostate cancer, age, education, calendar year of survey, and study year of survey.

We estimated the expected number of cancers that would have occurred in each arm in the absence of screening by applying incidence rates for an unscreened population. A recent modeling study performed a similar estimate in the overall US population during the time interval of the

PLCO Trial [4]. Specifically, to estimate what US prostate cancer incidence rates would have been in the absence of PSA screening, the study utilized Surveillance, Epidemiology, and End Results (SEER) rates for the period of 1985–1987, the last period prior to the start of the PSA era. We similarly used SEER rates for 1985–1987, applying age (5-year age group) and race (white, black, and others) specific SEER rates for the above period to the PY at risk for control and screened arm men during the screening period of the trial (the first 6 years) [5]. The observed minus the expected number of cases in each arm is thus a measure of the effective level of screening (and accompanying diagnostic follow-up) in each arm. To assess cumulative incidence by arm further, we performed an analysis similar to the above, but applied contemporaneous SEER rates, instead of the rates in 1985–1987, to the PLCO person years; this allows comparison of the observed numbers of cases to that expected, if PLCO subjects were representative of the general population.

Results

Of a total of 38,350 control arm men, 3815 (10.0%) without prior prostate cancer diagnosis were surveyed overall, and 2427 (6.3%) were surveyed during the screening phase of the PLCO Trial, study years 0–5.

Table 1 shows reported test usage rates in the control arm by study year and interval since last use, both for testing as part of a routine health visit (routine testing) and for any testing. Rates of routine PSA testing within the last year increased from 33% at study year 0 to 46% at study year 5, while rates of any PSA testing within the last year increased from 40% to 55% during this period. Correspondingly, the percent of men reporting no history of PSA use (for any reason) decreased from 38% at year 0 to 15% at year 5. For routine PSA testing more than 1 year earlier, 14–18% and 3–8% reported such testing 1–2 years and 2–3 years earlier, respectively. Overall, 40% of men reported routine PSA testing within the last year; when this

Table 1 Contamination survey results in control arm men

Test/study year[a]		Time period of latest test				
	Number of men surveyed[b]	< 1 year	1–2 years	2–3 years	> 3 years	
PSA		% Routine use (% any use)				% Never received test for any reason
0	181	33 (40)	15 (16)	3 (4)	2 (3)	38
1	422	31 (38)	14 (17)	6 (6)	5 (5)	34
2	385	41 (47)	17 (18)	5 (6)	4 (4)	24
3	410	39 (46)	16 (18)	8 (9)	5 (6)	21
4	435	46 (55)	15 (16)	7 (8)	3 (4)	17
5	392	46 (54)	18 (20)	5 (6)	3 (4)	15
0–5	2225	40 (47)	16 (18)	6 (7)	4 (4)	23
0–5 (adjusted)[c]		46 (52)	14 (16)	5 (6)	4 (4)	21
0–5 (screened arm)[d]		78	8	3	2	9
DRE						
0–5	2336	28 (34)	17 (19)	8 (9)	9 (11)	28
PSA or DRE*						
0	196	39 (46)	16 (16)	6 (7)	10 (10)	20
1	454	37 (44)	20 (22)	8 (8)	10 (11)	15
2	415	49 (55)	17 (18)	7 (7)	6 (7)	13
3	450	43 (50)	20 (21)	10 (10)	7 (7)	12
4	466	49 (58)	17 (17)	7 (7)	6 (6)	12
5	418	52 (58)	22 (23)	5 (5)	5 (6)	8
0–5	2399	46 (52)	19 (20)	7 (7)	7 (8)	13
0–5 (adjusted)[c]		51 (57)	17 (18)	6 (6)	6 (7)	12

[a]Study year of survey defined as participants' study year at the date 6 months prior to the completion of the survey. [b]Number includes only those answering the question about the given test. Thus, for PSA, 2225 out of 2427 (91.6%) answered the relevant questions, and for DRE, 2336 out of 2427 (96.3%) answered the questions. [c]Weighted average of survey results and presumed results for 'baseline contaminated' controls. The latter, comprising 10% of control arm men, were not included in the survey and were assumed to have had a routine PSA test within the last year (see text). [d]'Virtual survey' results in screened arm (see text). *Year category reflects most recent of the two tests.

figure was adjusted for the baseline contaminated controls, it increased to 46%. The results of the 'virtual' survey in the screened arm showed a comparable (adjusted) figure of 78% receiving a routine PSA in the past year.

A total of 28% of participants reported having had a routine DRE test in the past year and an additional 17% in the past 1–2 years. DRE rates did not vary by study year (data not shown). For the combination of PSA or DRE, routine usage in the last year increased from 39% to 52% and any use in the last year from 46% to 58% from study years 0 to 5; around 20% reported use in the past 1–2 years over these study years. A total of 13% reported never having had a PSA or DRE test. Overall, 46% reported routine PSA or DRE testing in the prior year; adjusting for the baseline contaminated controls increases this figure to 51%.

Table 2 shows the results of the logistic model of factors affecting rates of PSA and DRE routine testing within the last year. For PSA, having had a PSA test in the 3-year period prior to PLCO enrollment (OR = 1.5) and having a college degree (OR = 1.3) were significantly associated with greater testing, in addition to the study year of survey. For DRE testing, having a college degree (OR = 1.5) and having had a DRE test(s) prior to PLCO enrollment (OR = 1.8 for 1 prior test; OR = 2.8 for 2 or more prior tests) were significantly associated with increased test usage.

Table 3 shows the modeled frequency of test use in the control arm during the screening phase of

Table 2 Factors associated with reporting routine PSA and DRE testing in the past year

Factor	Number in survey	PSA		DRE	
		% Routine use in past year	OR (95% CI)[a]	% Routine use in past year	OR (95% CI)[a]
No prior PSA	1220	33.5	1.0 (ref.)	26.2	1.0 (ref.)
Prior PSA	894	46.6	1.5 (1.2–1.8)	31.0	0.9 (0.7–1.1)
No prior DRE	1139	37.0	1.0 (ref.)	21.2	1.0 (ref.)
One prior DRE	756	39.9	0.9 (0.7–1.2)	32.6	1.8 (1.4–2.3)
Two or more prior DREs	352	48.3	1.2 (0.9–1.6)	43.5	2.8 (2.1–3.8)
No college degree	1392	36.5	1.0 (ref.)	24.7	1.0 (ref.)
College degree	920	44.3	1.3 (1.1–1.6)	34.0	1.5 (1.2–1.8)

Calendar year of survey, age, history of enlarged prostate, and family history of prostate cancer were not statistically significant for either PSA or DRE test use. Study year was significant only for PSA use. Prior PSA/DRE refers to the 3-year period preceding PLCO enrollment.
[a]OR (odds ratios) based on multiple logistic regression model.

Table 3 Modeled (control arm) and observed (screened arm) testing frequency during screening phase of PLCO

	Routine testing		Any testing	
	% Screened arm	% Control arm (modeled)	% Screened arm	% Control arm (modeled)
Number of PSAs[a]				
0	5	26	5	17
1—2	4	16	4	17
3—4	15	38	13	42
5—6	76	20	78	24
Mean	5.0	2.7	5.3	3.1
Mean (first 3 years)	2.7	1.2	3.0	1.4
Number of DREs[a]				
0	5	45	5	34
1–2	7	43	6	49
3–4	88	12	89	16
Mean	3.5	1.1	4.0	1.3

For screened arm frequencies, men dying or developing prostate cancer before the projected date of the last screening exam were excluded. For any testing, some screened arm men have > 6 PSA tests and > 4 DREs. [a]Period covered is first 6 study years for PSA and first 4 years for DRE, corresponding to the protocol period of screening for each test.

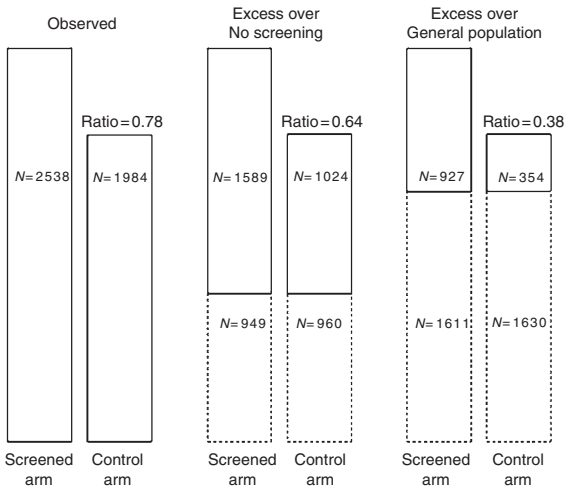

Figure 1 Observed prostate cancers by arm during screening period and excess over that expected based on pre-PSA screening era rates and general population rates. Dotted rectangles give expected numbers of cases, with expected plus excess adding up to total observed cases. See text for details

the trial as well as the actual frequency in the screened arm. For routine testing, the model shows that 26% of control arm men had no PSAs, 16% had 1–2, 38% had 3–4 and 20% had 5–6, with a mean of 2.7 tests. This is comparable with the screened arm figures of 5% having no screening PSAs, 76% having 5–6 and a mean of 5.0 tests. For any PSA use, 17% of control arm men had none, 17% had 1–2, 42% had 3–4, and 24% had 5–6. The mean number of screening DREs was 1.1 for control arm men, compared to 3.5 for screened arm men.

Figure 1 shows the observed number of prostate cancers by arm during the screening phase of the trial (study years 0–5), as well as the estimated excess cases over that expected based on pre-PSA screening era incidence rates and contemporaneous general population incidence rates. A total of 1984 (control arm) and 2538 (screened arm) cancers were observed, for a control to screened arm ratio of 0.78. The excess over pre-screening era expected cases was 1024 (control) and 1589 (screened), giving a control to screened arm ratio of 0.64 for cases presumably detected only due to screening. The excess in cases over that expected based on general population rates was 354 and 927 in the control and screened arms, respectively; the 354 reflects a 22% excess over the expected number of 1630.

Discussion

We have characterized the level of control arm contamination in PLCO in a number of ways that allow for direct comparisons of the use, and effect of that use, of PSA and DRE testing between study arms. Qualitatively, all point to a similar conclusion, namely, that there was widespread use of prostate cancer screening in the control arm but that the magnitude of that use was still considerably less than that observed in the screened arm of the trial.

By our original measure of control arm contamination, PSA use in the control arm was 46%; the comparable measure in the screened arm gave a use rate of 78%. These values give a rough control to screened arm 'ratio' of PSA use of 46/78 or 59%. This ratio is quite close to two other ratios derived here that also can be used to assess relative test usage across arms. The first of these is that of the mean number of PSA tests in the control versus screened arm; as estimated by the model this ratio is 54% (2.7/5.0). The other ratio is the excess diagnosed cancers in the control versus screened arm, which was estimated at 64% (1024/1589). This level of quantitative agreement suggests that all of these statistics are measuring a similar underlying quantity, which adds to the credibility of the reported findings. Utilizing a number of methods to assess contamination, which approach the problem from varying perspectives and utilize different data sources, is important because any given method, including those employed here, has limitations.

A limitation of the survey (HSQ) findings relates to the accuracy of self-report. In the typical epidemiologic study, in which individuals must be classified as either having received a test or not, accuracy at the subject level is critical. In assessing contamination, however, the bar is lower; the only concern is getting the overall proportion (using the tests) correct. Two recent studies, which verified self-reports with chart review, showed that the proportions of men with a PSA test, in the prior 2 years in one study and ever in the other, were essentially equivalent whether based on self-report or medical records [6,7]. In contrast, these studies found that DRE use was 10–15% higher when based on self-report as compared to medical records. One of these studies also examined reasons for testing and found that the percentages of PSA tests that were classified as for screening (as opposed to for symptoms or follow-up of an abnormal test) were similar according to self-report and chart review; for DRE, however, screening indications were 15% higher on self-report than on chart review.

The PLCO contamination survey was kept simple and manageable to help to insure high levels of accuracy and compliance; accordingly it did not attempt to ascertain all the test usages comprehensively during the PLCO screening period. To correct for that limitation, we employed a statistical model. As with any model, the findings

from it reflect the model's underlying assumptions. However, we believe that the assumptions were minimal, and that the results are robust to modest deviations from the assumptions.

A limitation of the approach of utilizing pre-PSA era incidence rates to estimate the expected numbers of cancers in the trial in the absence of screening is that the method assumes that there have been no secular trends in the underlying (age and race specific) background incidence rates in the interim. If such trends existed, due to changes in risk factors or other reasons, then the estimate of expected cases would be off by some amount, as would be the resulting numbers of cases above expected in each arm. Nonetheless, the ratio between arms of excess cases over expected (in the absence of screening), given at 64%, is robust to modest changes in background incidence. For example, changing background incidence rates by ±20% would result in values of the above ratio changing only by ±4%.

Compared to the situation with no contamination and perfect compliance, high levels of contamination and/or low levels of compliance tend to push the relative risk for outcomes towards unity, that is, no difference between arms. Thus, the finding of a significant mortality difference between arms is made more difficult with increased contamination, and the power of the trial is thus reduced. The PLCO protocol revision assumed a 38% contamination rate and a 90% compliance rate, which still would have left the trial, over its full length of 13 years, with approximately 90% power, although the terms contamination and compliance were not rigorously defined. Whatever measure is used for these quantities, the gap between them appears to be less than that between 38% and 90%; therefore, the power of the study is diminished. Given that the partial outcome of the trial (7–10 years) is already known, it makes little sense to re-calculate the power of the full-length trial as of the start of the study. A calculation of conditional power, the probability of the study finding a significant difference given the updated contamination and the outcome data to date, would be informative. However, such a calculation would be complex, and would require several assumptions and is beyond the scope of this analysis.

Several of the ERSPC centers have reported contamination assessments [8,9]. The Italian center, similarly to PLCO, performed a survey on a sample of controls, with a resulting rate of PSA testing in the last year of 30–35%. Other ERSPC centers took advantage of available database linkages to connect control arm men with laboratory records of PSA testing. For the Rotterdam center, data were collected from a regional general practitioner (GP) laboratory that served about 75% of control arm subjects' GPs. With results from these subjects extrapolated to the entire control arm; a total of 31% were estimated to have received PSA testing. A sample of men identified as receiving PSA testing were subsequently surveyed to differentiate (asymptomatic) screening from other forms of PSA testing; approximately half were determined to have had screening [10]. The center in Spain utilized a similar linkage approach on a subset of controls, and reported a low rate of ever PSA use of 7%. Overall, these data show a lower level of contamination in ERSPC than in PLCO.

When the level of contamination is substantial, it is tempting to somehow 'adjust' trial results for this contamination (and noncompliance) [11]. However, quantitative adjustments of the estimate of the intervention effect for contamination and noncompliance are problematic. Ignoring randomization and comparing those who received the assigned intervention with those who did not is invalid, as those choosing to receive an intervention are likely to be different from those not so choosing, with different underlying risks for the outcome of interest. More sophisticated approaches, which avoid this obvious type of bias, rely on identifying contaminated and uncontaminated subsets of participants in the control arm and compliant and noncompliant subsets in the intervention arm [11,12]. Since these approaches rely on all-or-none (binary) definitions of compliance and contamination, utilizing them in PLCO would require specific definitions to be agreed upon; different definitions, all defensible, could lead to different results. Recently, the ERSPC published risk estimates adjusted for noncompliance and contamination using the above methodology, where screened arm compliance was defined as participating in the initial round of PSA screening and control arm contamination was defined as receiving a PSA test for asymptomatic screening at any time after randomization [10]. However, as contamination rates were not available for the total ERSPC control population, the authors extrapolated the contamination rates (and outcomes) in the Rotterdam component of ERSPC to the other countries in the trial.

In the US, with its open, decentralized health care system, it is hard to avoid contamination in screening or prevention trials as long as the employed modality is available to the public for some indication, particularly when it has been widely promoted by advocacy groups. A variety of factors, including the invasiveness and level of discomfort or side effects associated with the modality, the modality's expense and the extent of its coverage by insurance plans, the intervention (if any) received in the control arm, and the level of

media exposure of the modality, will influence the contamination rate in controls. In the case of PSA, virtually all of the above factors were working in the same direction, to favor contamination.

In contrast, another US screening study shows the reverse situation, where the majority of these factors worked against contamination, with a resultant low level of use in the control arm. In the Lung Screening Study (LSS), a pilot feasibility study for lung cancer screening with low dose spiral CT (LD-CT) versus chest radiograph (CXR), the intervention arm modality of LD-CT was expensive and not widely covered (for a screening indication) by insurance plans and the control arm did receive some intervention; further, while there was some favorable media coverage, it was not as extensive as for PSA. In the CXR arm of the LSS, only 1% of subjects reported receiving routine LD-CT in the first year of the study [13].

It is interesting to compare the level of PSA utilization in PLCO control arm men with that in the general population. A recent study examined trends in PSA testing as reported from the National Health Interview Survey (NHIS). Among white men in the birth cohorts of 1925–1934 and 1935–1944, 71% and 57%, respectively, were estimated ever to have had a PSA test by 2000 [14]. Based on combining the 1999 and 2001 HSQ surveys (there was no 2000 survey), PLCO control arm white men reported a PSA test ever at rates of 75% ($n = 132$) and 81% ($n = 165$) in the 1925–34 and 1934–44 birth cohorts, respectively. Note that adjusting for the 'baseline contaminated' men not included in the survey would further increase these proportions by 2–3%. The same study also reported the frequency distribution of the number of PSA tests administered for the 5-year period 1996–2000 among whites 60–64 years old in 2000 (i.e, born between 1936 and 1940) based on the NHIS data. A total of 38% had received no PSAs, 22% received 1–2 PSAs, 10% received 3, and 30% received 4 or more PSAs during this period. Based on the model for the PLCO survey data, for the 5-year period from study year 0–4, which generally coincided with the period 1996–2000, the frequency distribution among all control arm men (median age 64 in 2000) of number of PSA tests for any reason was 19% with none, 25% with 1–2, 27% with 3 and 29% with 4 or more. Assuming that those with 4 or more PSAs in NHIS averaged 4.5 tests, the mean number of tests in the NHIS men was 1.95, as compared to 2.5 for the PLCO control arm men. These findings are consistent with the approximately 20% excess of observed cancers in the PLCO control arm relative to that expected based on the general population incidence rates.

Given the high level of contamination in PLCO, one way of interpreting the trial is that it is a study of an active (organized) program of screening versus 'passive (opportunistic) screening', that is, the current level of screening ongoing in the community. Viewed in this way, 'contamination' could be conceived as the level of screening in the usual care arm over and above that in the general community, and by this standard the level of contamination would be relatively low. From this perspective, a finding of no benefit implies that an organized program does not confer significantly greater mortality benefit for prostate cancer than opportunistic screening as currently carried out in the US However, the study interpreted in this manner would not be able to answer the question of whether that level of opportunistic screening is conveying a mortality benefit over no screening.

Acknowledgments

The PLCO Trial was funded by contracts from the National Cancer Institute. The ClinicalTrials.gov identifier for PLCO is NCT00002540. The specific research described here received no additional grant or contract from any funding agency in the public, commercial, or not-for-profit sectors. The authors thank the PLCO PIs and the PLCO participants.

References

1. **Andriole GA, Crawford ED, Grubb R**, *et al*. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* 2009; **360**: 1310–19.
2. **Schroder FH, Hugosson J, Roobol MJ**, *et al*. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 2009; **360**: 1320–28.
3. **Prorok P, Andriole GL, Bresalier RS**, *et al*. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* 2000; **21**: 273–309S.
4. **Draisma G, Etzioni R, Tsodikov A**, *et al*. Lead time and over diagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst* 2009; **101**: 374–83.
5. **Surveillance Research Program**. National Cancer Institute SEER*Stat software. Available at: www.seer.cancer.gov/seerstat version 6.5.1 (accessed August 2009).
6. **Hall IH, Van Den Eeden SK, Tolsma DD**, *et al*. Testing for prostate and colorectal cancer: comparison of self-report and medical record audit. *Prev Med* 2004; **39**: 27–35.
7. **Sayre EC, Bunting PS, Kopec JA.** Reliability of self-report versus chart-based prostate cancer, PSA, DRE and urinary symptoms. *Can J Urol* 2009; **16**: 4463–71.
8. **Ciatto S, Zappa M, Villers A**, *et al*. Contamination by opportunistic screening in the European Randomized Study of Prostate Cancer Screening. *Br J Urol* 2003; **92**: 97–100.
9. **Roemeling S, Roobol MJ, Otto SJ**, *et al*. Feasibility study of adjustment for contamination and non-compliance in a prostate cancer screening trial. *Prostate* 2007; **67**: 1053–60.
10. **Roobol MJ, Kerkhof M, Schroder FH**, *et al*. Prostate cancer mortality reduction by prostate-specific

antigen-based screening adjusted for nonattendance and contamination in the European Randomised Study of Screening for Prostate Cancer (ERSPC). *Eur Urol* 2009; **56**: 584–91.
11. **Cuzick J, Edwards R, Segnan N.** Adjusting for non-compliance and contamination in randomized clinical trials. *Stat Med* 1997; **16**: 1017–29.
12. **Angrist JD, Imbens GW, Rubin DR.** Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996; **92**: 444–55.
13. **Gohagan J, Marcus P, Fagerstrom R, *et al*.** Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph. *Chest* 2004; **126**: 114–21.
14. **Mariotto AB, Etzioni R, Krapcho M, Feuer EJ.** Reconstructing PSA testing patterns between black and white men in the US from Medicare claims and the National Health Interview Survey. *Cancer* 2007; **109**: 1877–86.

Appendix: Model to estimate actual number of tests performed from survey data

It was assumed that each man had a PSA (DRE) test at most once during each study year; therefore, there were $2^6 = 64$ possible temporal patterns of test use in the survey population (i.e., 0 or 1 test during each of 6 study years). With five possible survey responses in 6 different survey years, there were $6 \times (5–1) = 24$ degrees of freedom for the model, meaning that at most 24 parameters could be fit to these data (for n parameters, $n + 1$ patterns can be distinguished, since the probabilities sum to 1).

We thus proceeded as follows: we randomly selected subsets of 25 (out of 64) possible temporal patterns and for each subset fit a model with 24 parameters, which defines the 25 probabilities for the patterns. All other patterns were assigned zero probability. Each subset included the pattern for 0 tests and 6 tests. We generated 1000 subsets and averaged the results of these 1000 fit models to get our final results. Note that for DRE, since only 4 study years are of interest, there are $2^4 = 16$ possible patterns and $4 \times (5–1) = 16$ degrees of freedom, so no patterns had to be disallowed.

For a given model, as defined by a subset of allowable patterns, we fit the parameters by maximum likelihood. Given a parameter vector, that is, the frequencies of the (allowable) patterns, the likelihood of the survey response of each man was calculated by summing up the probabilities of all patterns that were consistent with that response. For example, if a man in study year 5 reported having had his most recent test 1–2 years prior, then his likelihood would sum up the frequencies of all patterns with a test in study year 4 and no test in study year 5.

As a test of the possible bias of the model, we performed simulated surveys in the screened arm similar to the virtual survey described in the methods and ran the model on the results to see if they would reproduce the actual frequency data. Specifically, we randomly selected a study year and a date within that year for each screened arm man, determined the correct survey response in terms of the time interval of the latest PSA screen, and then applied the model to the simulated survey results to get a frequency distribution of screening patterns. The mean number of PSAs and the percentage with any PSA as estimated from the model were close to the actual figures as determined directly from the PLCO screened arm data base – 5.0 (actual) versus 4.5 and 95.3% (actual) versus 91.1% for mean number and percentage with any tests, respectively.