Author's response to reviews

Title: A population-based cohort study of chest x-ray screening in smokers: Lung cancer detection findings and follow-up

Authors:

Lorenzo Dominioni (lorenzo.dominioni@uninsubria.it)
Nicola Rotola (nicola.rotolo@uninsubria.it)
William Mantovani (william.mantovani@univr.it)
Albino Poli (albino.poli@univr.it)
Salvatore Pisani (pisanis@asl.varese.it)
Valentina Conti (v.conti@hotmail.it)
Massimo Paolucci (massimo.paoluccicolico@tin.it)
Fausto Sessa (fausto.sessa@uninsubria.it)
Antonio Paddeu (anpaddeu@tin.it)
Vincenzo D’Ambrosio (vincenzo.dambrosio@aogallarate.it)
Andrea Imperatori (andrea.imperatori@uninsubria.it)

Version: 2 Date: 21 November 2011

Author's response to reviews: see over
Dear Editor,

Varese, November 21, 2011

Re: Cover Letter for submission of revised MS 6205820665816792 - Title: A population-based cohort study of chest x-ray screening in smokers: Lung cancer detection findings and follow-up.

This is to submit the revised manuscript, addressing the Editorial Requirements and the Reviewers’ concerns and comments.

**Editorial Requirements**

**Kindly expand the background section**

The background section has been expanded and revised as follows (page 5 of revised manuscript).

**Background (reference numbers are those of the revised manuscript)**

The historical observational studies of chest x-ray (CXR) screening for lung cancer (LC) showed improved LC-specific survival but no reduction of LC-specific mortality [1-6]. Four case-control studies performed in the 1990s in Japan on a population level however suggested CXR screening effectiveness, as indicated by LC mortality reductions of about 40% [7, 8]. The results of these case-control studies were obtained in scarcely defined target populations, were likely biased by self-selection and are yet to be confirmed in western countries. Randomized trials of LC screening by chest radiography performed in the 1970s [9, 10] failed to answer the question of efficacy of radiographic screening, likely due to methodological flaws [11]; however, the possibility was recognized that a small but important benefit from annual CXR could have been missed [12-14]. The lung component of the Prostate, Lung, Colon and Ovarian (PLCO) cancer screening randomized trial was carried out to definitively assess the efficacy of CXR screening [15]. The recently published PLCO results showed that four annual chest radiographs did not reduce LC-specific mortality in volunteers [16], but it remains uncertain whether this finding may be generalized on a population level. In 1997 we started to study the PREDICA cohort (hereafter concisely referred to as “cohort”), a clearly defined population-based cohort of 5,815 heavy or long-term smokers of the Province of Varese, Italy, to ascertain whether CXR screening at community level decreases LC mortality [17]. The cohort was invited to an annual CXR screening program and was followed-up for 13.5 years. Our aim is to evaluate here the LC detection results and the long-term LC survival in screening participants as well as in nonparticipants of this population-based cohort. We also examine LC survival by intent-to-screen, in the entire cohort invited to screening. This analysis is preliminary to the evaluation of CXR screening effectiveness on a population level by the LC mortality indicator, to be reported at a later date.

**We would like to ask clarification on ethical approval**

Attached below please find scanned certificate letter of Ethics Committee approval.
Varese, November 7, 2011

Prot. n.

Rif: AM/cg

Object: PREDICA Project for early diagnosis of lung cancer by screening radiography in the Province of Varese

This is to certify that this Comitato Etico - Ethics Committee - of Azienda Ospedaliera Ospedale di Circolo di Varese - Varese Hospital - on May 20, 1997, approved the "PREDICA Project for early diagnosis of lung cancer by screening radiography in the Province of Varese" study.

Its purpose is early diagnosis of lung cancer, in the asymptomatic phase, by annual chest radiography screening of the population at risk in the Province of Varese. Duration of screenings is 4 years after enrolment, starting from 1997. Participation of at risk subjects (smokers and former smokers) to the screening is on volunteer basis, after informed consent of the participants, and it is free of charge.

The principal investigator, responsible for the study, is Prof. Lorenzo Dominioni, Thoracic Surgery Centre, University of Insubria, Varese.

Best regards,

The Ethics Committee President
Giovanni Chelazzi
**Tables and Figures**

Tables and figures are provided following the instructions received.

We amended two mistakes in Tables:

In Table 2: there were 49 (not 50) LC patients in the cohort with stage I disease, and staging was not available in 17 cases (not 16, as previously reported wrongly in footnote). The corresponding percentage values were amended.

In Table 4: there were 25 cases (not 24) of Nonscreen-detected LCs stage IIIB-IV and extensive SCLC. The corresponding percentages were amended.

These amendments had no influence in the interpretation of the results. After amending these mistakes, the text of the manuscript required no modifications.

Point-by-point responses to the Reviewers’ concerns are attached.

Best regards,

Lorenzo Dominioni, MD
on behalf of all the authors
1. quality of data collected
   a. the authors selected a cohort of potential research subjects, but only 21% participated. this draws in question whether the data are population-based, as the authors claim.

   **REPLY** As specified in the Methods section of manuscript, the data that we present are of a population-based cohort of smokers, which includes all residents in 50 communities of the Varese Province possessing the screening enrolment criteria: 45-75 years of age, both genders, current or ex-smokers of >10 pack-years. We analyzed the lung cancer detection findings and follow-up in the whole population-based cohort invited to CXR screening (n=5,815), and separately in screening participants (n=1,244) and nonparticipants (n=4,571).

   b. the smoking exposure are questionable - there is no explanation as to how the researchers knew that their subjects had at least 10 pack-years.

   **REPLY** Manuscript revised as follows: (page 6, line 24)  "The 50 GPs served nearly 100% of residents in their respective communities, in the NHS. They recruited all cigarette smokers in their public medical practices who were screening-eligible, based on self-reported smoking history abstracted from medical records."

   The smoking history was self-reported by subjects enrolled in the whole cohort invited to CXR screening. Inaccurate self-reported smoking status occurs relatively infrequently in adults of the general population (Partrick DL et al., Am J Public Health 1994;84:1086-93; Wells AJ et al., Am J Public Health 1998;88:1503-9) as well as in participants in lung cancer screening (Studts JL et al., Cancer Epidemiol Biomarkers Prev 2006;15:1825-8).

   c. how were the 50 GPs selected? is this a random sample of GPs?

   **REPLY** The selected GPs were not a random sample of GPs. From previous experience we knew that GPs working in the immediate surroundings of the city of Varese are highly cooperative in population studies, while those working at the periphery of the Varese Province are less cooperative. Therefore, randomization of GPs, even by areas, would have caused unbalanced representation of screening candidates resident in the various geographical areas of the Province. Consequently, we identified a sample of 50 GPs well distributed geographically over 44 different towns, willing to collaborate in the screening programme and serving communities representative of urban and rural areas of the Varese Province (see Appendix B in reference 17, list of towns where the screening was conducted). All smokers with screening criteria resident in these 50 communities (n=5,815) formed the cohort of our study. This cohort was found to be representative of the Varese smokers population (reference 17, page 434).

   d. were investigators blinded to method of detection (screen vs. symptom) when assigning cause of death.

   **REPLY** The manuscript was revised as follows: (In the Methods section, page 9, line 13)  “Death certificates were reviewed by the mortality review committee members (LD, AI, NR, FS, SP, APo and WM), who were not blinded to the mode of LC detection. Those deaths definitely attributed to LC we recorded as LC-specific deaths.”

   (In the Methods section, page 10, line 1 ) “The cause of death of the control LC group subjects was assigned by the mortality review committee, as described above for deceased subjects of the cohort.”

   (In the Discussion section, page 15 line 19) “Another limitation is that the investigators assigning the cause of death in the cohort and in the control LC group were not blinded to mode of LC detection. Sensitivity of LC death certificates however was shown to be high and similar in screening participants and nonparticipants [17]; therefore, selective misclassification of the cause of death unlikely occurred.”
e. screening participation rates at later years were extremely low.

REPLY The discussion was revised as follows (page 15, line 14):

“Compliance with annual screening progressively decreased, more markedly than in other CXR screening trials. By year 3 we recorded 59% adherence, while 79% was observed in the PLCO radiography screening [16] and about 80% in the Mayo Lung Project [36].”

2. analyses and data interpretation
a. comparisons should be restricted to participants vs. non-participants.omit data for the entire cohort (what's the benefit of showing this data?) and the control group.

REPLY The reason to show the data for the entire cohort is to examine LC survival by intent-to-screen, in the screening invited cohort. This is specified at the end of the revised Background Section (page 5, line 19):

“Our aim is to evaluate here the LC detection results and the long-term LC survival in screening participants as well as in nonparticipants of the cohort. We also examine LC survival by intent-to-screen, in the entire cohort invited to screening. This analysis is preliminary to the evaluation of CXR screening effectiveness on a population level by the LC mortality indicator, to be reported at a later date.”

b. treatment of impact of screening biases (leadtime, length sampling, and overdiagnoses) are confusing, not rigorous, and in places in error.

REPLY The impact of screening biases in this study have been addressed in the Discussion section, that has been amended in the revised manuscript, as follows:

(From page 14, line 20, until page 15, line 12.)

“In the context of our study the impact of lead-time bias on long-term LC specific survival of participants seems negligible, because the LC specific survival curve of screening participants plateaued after five years, until ten years from diagnosis (Figure 1B). Enrolment of asymptomatic individuals in the cohort was a source of healthy selection. The latter however did not influence the long-term LC survival, as shown by comparison of the Kaplan-Meier LC survival curve of cohort’s nonparticipants and of control group (Figure 1D); the LC survival was initially greater in nonparticipants, likely due to healthy selection, but after five years from diagnosis it was virtually identical to that of control group (Figure 1D and Table 6). The latter finding is also consistent with the similar LC resection rate (27% vs. 26%) and similar proportion of histological diagnoses and of LC stage distribution in the nonparticipants and in the control group (Table 2 and Table 3). To evaluate length-time bias and overdiagnosis bias in our study, we focused on the VDT of screen-detected LCs, that is also an indicator of tumor aggressiveness [33-35]. Incidence screen-detected LCs had short median VDT (80 days), and only one of these cancers had VDT >300 days, indicating that most of them grew rapidly and unlikely were overdianosed, in agreement with the observations of other authors about CXR screen-detected LCs [22]. Moreover, among screen-detected cancers we found no cases of bronchioloalveolar carcinoma, a slow-growing subtype that may be overdianosed. Furthermore, we previously showed that the LC incidence standardized rate ratio in the whole cohort was 1.07 [17], suggesting that the number of possibly overdianosed LCs was minimal.”

c. too many tables and figures

REPLY The number of tables and figures presented is reasonable for this type of study, which reports the lung cancer detection results and survival curves in several subgroups of patients: screening participants, nonparticipants and the entire cohort.

3. other
a. limitatons of RCTs are discussed, but limitations of observational research are not.

REPLY Limitations are presented in the revised Background section (page 5).

b. selective discussion of previous research in this area and state of the science.
Previous research and state of the science are presented in the revised Background section (page 5).

c. research is somewhat outdated given the results of the NLST

REPLY We do not think that our research is outdated. The NLST greatly contributed to clarify the effectiveness of CT screening for lung cancer. Yet, much has to be learned about the cost effectiveness and accessibility of CT screening in the context of limited financial resources of health systems worldwide. Moreover, our study differs from previous screening studies because it explores the results that may be obtained in the whole target population invited to CXR screening, not only in those individuals who volunteer to participate.

d. is this study called PREDICA? if not, what is predica?

REPLY This is explained in the revised Background section (page 5, line 16):

“In 1997 we started to study the PREDICA cohort (hereafter concisely referred to as “cohort”), a clearly defined population-based cohort of 5,815 heavy or long-term smokers of the Province of Varese, Italy, to ascertain whether CXR screening at community level decreases LC mortality [17].”

e. criticism of NLST in discussion section is unfounded - the idea is not to necessarily screen all smokers - just those at the highest risk.

REPLY We are not criticizing the NLST. We are just reporting the limitations to the generalizability of the NLST results, as outlined in the NLST report by the research team (Reference 28, page 408).
Reply to Reviewer: Heidi C. Roberts

(reference numbers are those of the revised manuscript)

Reviewer’s report:
The authors present a very interesting, thorough and well written study on the utility of CXR for lung cancer screening.

My only comment is on the potential bias between participants and non-participants: is there a demographical difference between those groups, i.e. do the non-participants have a lower education level etc. that would explain their refusal to get screened?

REPLY We revised the Discussion as follows:

(page 14, line 1) “Only 21% of this cohort participated in screening, a low attendance rate similar to that observed in mass screening for LC in Japan [29, 30]. Analyzing the demographic features and risk factors of the whole screening-invited cohort, we previously reported that participation was possibly prompted by increased awareness of LC risk, while it was not related to educational level [17].”
Reply to Reviewer: Paul Pinsky

(reference numbers are those of the revised manuscript)

Reviewer's report:
This manuscript presents interesting data relating to characteristics of lung cancers by mode of detection. The study design is a bit complex but adequately described. The data and statistical analysis is generally sound.

Major Changes

1. With respect to VDT, where a median of 80 days was reported, please specify how many cases this was calculated for. Was it calculated for all non-baseline screen detected cases (meaning that the lesion was always evident at the prior screen)? Please give more details of the distribution of VDT. Even if median was 80 days, this does not imply no overdiagnosis since a significant fraction could still have had very large VDTs.

REPLY
In Methods section we revised:
(page 9, line 4) “To calculate VDT of a lung tumor not evident retrospectively on the prior radiograph, the tumor was arbitrarily attributed the dimension of 6 mm on the prior CXR, corresponding to the estimated visibility threshold [22].”

In the Results section we revised:
(page 12, line 4). “The VDT was calculated for all 21 non-baseline screen-detected cases. For 16 of these cases the lung tumors were not evident retrospectively on the prior radiograph and were arbitrarily attributed the 6 mm estimated visibility threshold dimension. The median VDT of all non-baseline screen detected cases was 80 days (range: 44-318); only one of these cancers had VDT >300 days.”

For Reviewer: the frequency distribution of VDT for the 21 non-baseline screen-detected cases is shown in the figure below (not included in the revised manuscript).
2. The authors should statistically compare the survival of non-participants versus controls. This would help to address whether the control group is appropriate for the cohort, since neither group (control nor non-participants) is affected by screening. Of course, non-participants may have some selection bias (since they did not select to participate), but it would be nice to know that the groups’ survival are similar.

**REPLY** In the Figure 1, we added Figure 1D, and in the Results section the revised text was implemented as follows: (page 13, line 5) “Moreover, comparing controls versus nonparticipants (neither group was screened) we found that after 5 years from diagnosis the LC survival in these two groups was essentially identical (Figure 1D and Table 6).”

In the Discussion section of the revised version we added:
(page 16, line 7) “Moreover, for the evaluation of long-term survival of the cohort, the control group seems appropriate because the cohort’s nonparticipants and the controls (neither group was screened) had similar long-term survival (Table 6).”

3. In Table 2, there is a row for “asymptomatic LC diagnosis”. Other than the screen detected cases, how was this determined (e.g., chart review, self report)? Were these by screening or some other method?

**REPLY** Asymptomatic LC diagnosis was attributed to cases that were either screen-detected or incidentally found, as documented from reviewing patient files. In nonparticipants and in controls the asymptomatic LCs diagnosed were all incidentally found. This information has been added in the footnotes of Table 2.
“Footnote: * In nonparticipants of the cohort and in the control group the asymptomatic LCs diagnosed were all incidentally found, as documented by patient file review.”

**Discretionary Changes**

1. **It would be useful to show the total number of person years at risk for lung cancer in the participants and non-participants. That way, a lung cancer rate could be calculated in each group. This is important to assess possible over-diagnosis, which of course affects survival.**

**REPLY** We added in the revised manuscript in Discussion section (page 15, line 10) “…Furthermore, we previously showed that the LC incidence standardized rate ratio in the whole cohort was 1.07 [17], suggesting that the number of possibly overdiagnosed LCs was minimal.”

We added the above sentence to summarize the LC incidence findings in the cohort, previously reported (Table 1, of reference 17), showing that the cumulative person-years of observation were 11,315.1 in participants and 40,291.8 in nonparticipants.
The impact of LC overdiagnosis on LC incidence in the whole cohort invited to screening was minimal. This is documented on page 431 of reference 17: “…Age-standardized LC incidence observed in the whole PREDICA cohort was 5.15 (95% CI, 4.47–5.84)/1000 person-years, while the LC incidence expected in the PREDICA towns-matched population was 4.81 (95% CI, 4.06–5.68)/1000 person-years. The PREDICA cohort LC incidence SRR was 1.07 (95% CI, 0.87–1.31).”

2. **It would be of interest to see if survival within stage (say stage I) is different for the screen detected versus the others (participant non-screen detected, non-participant, control).**

**REPLY** We made the following comparisons to see if survival within stage I LC is different for the screen detected vs. the others:
   a) Comparison of 10-year survival of screen-detected stage I LCs (n=18) vs. participant non-screen detected stage I LCs (n=3) is not meaningful because of small number of the latter.
   b) The 10-year survival in screen-detected stage I LCs (n=18) was nonsignificantly higher than in non-participant stage I LCs (n=28) (p=0.054).
   c) The 10-year survival in screen-detected stage I LCs (n=18) was nonsignificantly higher than in control stage I LCs (n=23) (p=0.271).

While comparison of survival by stage and by mode of detection is of interest, we think that a comprehensive analysis of these comparisons would require a considerable extension of the manuscript and the addition of more tables. Because another Reviewer of this paper already complained of too many tables and figures, we think that comparison of survival by stage between the various group should be dealt with in separate paper.