

PERSPECTIVE

Lead-Time Models Should Not Be Used to Estimate Overdiagnosis in Cancer Screening

Per-Henrik Zahl, DrMedSci¹, Karsten Juhl Jørgensen, DrMedSci², and Peter C. Gøtzsche, DrMedSci²

¹Norwegian Institute of Public Health, Oslo, Norway; ²The Nordic Cochrane Centre, Rigshospitalet, Dept. 7811 Copenhagen, Denmark.

Lead-time can mean two different things: Clinical lead-time is the lead-time for clinically relevant tumors; that is, those that are not overdiagnosed. Model-based lead-time is a theoretical construct where the time when the tumor would have caused symptoms is not limited by the person's death. It is the average time at which the diagnosis is brought forward for both clinically relevant and overdiagnosed cancers. When screening for breast cancer, clinical lead-time is about 1 year, while model-based lead-time varies from 2 to 7 years. There are two different methods to calculate overdiagnosis in cancer screening—the excess-incidence approach and the lead-time approach—that rely on two different lead-time definitions. Overdiagnosis when screening with mammography has varied from 0 to 75 %. We have explained that these differences are mainly caused by using different definitions and methods and not by variations in data. High levels of overdiagnosis of cancer have usually been explained by detection of many slow-growing tumors with long lead-times. This theory can be tested by studying if slow-growing tumors accumulate in the absence of screening, which they don't. Thus, it is likely that the natural history of many subclinical cancers is spontaneous regression.

KEY WORDS: cancer; screening; lead-time; overdiagnosis.

J Gen Intern Med 29(9):1283–6

DOI: 10.1007/s11606-014-2812-2

© Society of General Internal Medicine 2014

Etzioni et al.¹ have recently compared the two different methods to calculate overdiagnosis in breast and prostate cancer screening: the *excess-incidence approach* and the *lead-time approach*. They write that the main limitations of the excess-incidence approach are that the prevalence screening rounds are included and that follow-up after screening is insufficient, whereas the lead-time approach relies on model choices and assumptions that are not transparent. They conclude that they “wonder whether it is possible to compare and integrate results across published

studies of overdiagnosis”. Etzioni et al.¹ do not state which approach they consider most reliable. We have recently published a comparison of the two approaches.² We concluded that the excess-incidence approach is preferable. The lead-time approach seriously underestimates overdiagnosis because of flawed assumptions about how much screening advances the time of diagnosis.²

It is generally accepted that prostate specific antigen (PSA) screening for prostate cancer leads to about 50 % overdiagnosis, which means that one in three prostate cancers in a population offered screening are overdiagnosed. This estimate is based on the excess-incidence approach applied to the randomized trials.¹ Using the same approach, we calculated 31 % overdiagnosis in randomized trials of mammography screening;³ in observational studies, we found 30–50 % overdiagnosis of invasive breast cancer in countries with publicly organized screening programs.^{4,5}

WHY THE EXCESS-INCIDENCE APPROACH IS PREFERABLE

With the excess-incidence approach, the reduction in the number of cancers after screening has stopped is subtracted from the incidence increase observed during screening. With long follow-up after screening has stopped, this approach will correctly adjust for advancement of the time of diagnosis (lead-time),² provided that there is little opportunistic screening outside the targeted age range. The conditions of absence of opportunistic screening and long follow-up in age groups no longer offered screening cannot always be met, and Etzioni et al. are correct that this can bias the excess-incidence based approach. However, for a claim of no overdiagnosis to be valid, it is important that *all* extra cases detected during the screening period are compensated for, including those detected in the prevalence screening round. If, for example, a large reservoir of dormant or slow-growing lesions is a major cause of overdiagnosis, it will be most prominent in the first screening round. Simply disregarding the prevalence peak, as Etzioni et al.¹ suggest, and as we also did to reach a conservative estimate,^{4,5} underestimates overdiagnosis.

Received November 25, 2013

Revised December 20, 2013

Accepted January 27, 2014

Published online March 4, 2014

The excess-incidence approach has been criticized because it does not follow closed cohorts for many years with screening, and subsequently, for many years after screening, it has stopped to fully identify the compensatory reduction in cancers. However, this has in fact been done, with results that are comparable to our previous estimates.⁶

WHY THE LEAD-TIME APPROACH IS MISLEADING

We have shown that lead-time models wrongly adjust away most true overdiagnosis, because they rest on the false assumption that all the incidence increase during screening is caused by early detection of slow-growing tumors that would have surfaced later.² There are two different lead-time approaches. The classical lead-time adjustment method includes all cancers diagnosed in a long time period after screening has stopped (10–15 years) in both the screening and the control group when comparing cumulative hazard rates.⁷ This ratio is initially equal to the excess-incidence estimate, but rapidly approaches zero difference between the two groups with increasing follow-up.² Another lead-time approach is using a multistage statistical model where all tumors are assumed to grow (but with different speed), which estimates overdiagnosis as the fraction of tumors detected at screening that would never become clinical during the lifetime of the patient. This approach is highly model dependent; for example, three different lead-time models gave overdiagnosis estimates ranging from 23 % to 42 % using the same Dutch data set.² The MISCAN model represents such an approach, and the assumptions in this model are multiple and not transparent.^{1,2} The lead time

models always give lower overdiagnosis estimates than the excess-incidence approach if screening detects dormant tumors or tumors that regress spontaneously.

The theoretical incidence rate in a screened cohort with many slow-growing tumors and little overdiagnosis is presented in Figure 1a. An initial prevalence peak is followed by a period where the rate stabilizes slightly above the background incidence level, after which a compensatory decline appears when screening stops. The validity of this model can be studied by comparing Figure 1a to observed incidence rates in a screened population (Fig. 1b). The incidence increased markedly when women aged 50–51 years were invited to their prevalence screen in Norway, and also for women above 60 years when screening started in Norway (Fig. 1b)⁸ and Sweden.⁹ Second, the incidence increase persists during all following screening rounds and is consistently 50 % above the expected incidence without screening, which is an increase that is two to three times larger than that predicted with the lead-time approach (Fig. 1b). Third, the compensatory incidence decline after screening had stopped at age 69 is very small, and disappears much earlier than the assumed long lead-times would suggest, and compensates for only a small fraction of the observed incidence increase.^{2,4,5,8,10}

LIFETIME RISK OF OVERDIAGNOSIS

No studies have followed all women until they have all died. Kalager et al.⁶ included all cancers diagnosed in a 10-year period after screening had stopped to adjust for long

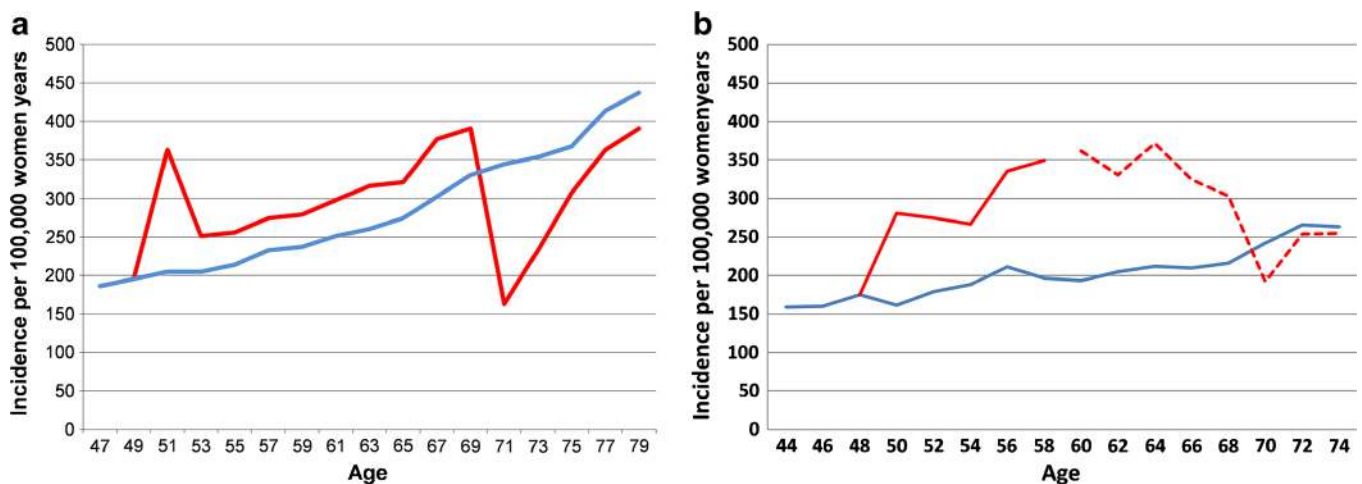


Figure 1. a Breast cancer incidence with screening (red line) as predicted with the MISCAN model, which is a lead-time model.⁷ The blue line is the no screening group. b Observed breast cancer incidence rates for Norwegian women in the period 1991–1995 before screening started (blue line), and for a cohort of 70,000 Norwegian women aged 50 years invited to biennial screening (first screening in 1996–2001) with 10 years follow-up (solid red line). Another cohort of women aged 60 years (43,000) in 1996–2001 were also followed up for 10 years with biennial screening (the prevalence screening was done before age 60) and then in 1–5 years without screening (dotted red line). To compare, the breast cancer incidence rate for women under age 50 years was constant throughout this period.

lead-times. They found that mammography screening in Norway was associated with 15–25 % overdiagnosis of invasive breast cancers. If the follow-up is life-long, this method estimates lifetime risk of overdiagnosis, and the percentage is therefore smaller than the estimated level of overdiagnosis in the screening period of 30–50 %. In practice, different studies have included different periods of follow-up after screening stopped, and the individual percentage estimates are therefore not comparable.

Both the excess incidence approach and the lead-time approach need to adjust for changes in the underlying incidence rate. A strong underlying incidence increase has been commonly assumed, mainly due to hormone replacement therapy (HRT). However, in the last decade, HRT use has dropped 80 % in Norway in post-menopausal women without any significant corresponding drop in breast cancer incidence, which we would expect if HRT is causing breast cancer.¹⁰ Thus, opportunistic screening and increased screening sensitivity seem more likely explanations for the increases in background incidence in the non-screened areas and non-screened age groups observed in recent years. Assuming a strong underlying incidence increase most likely underestimates the true level of overdiagnosis with screening, but not very much, as the results of Kalager et al.⁶ correspond well with ours.^{4,5,10}

A FORMAL TEST OF THE LEAD-TIME MODEL

If all the incidence increase is caused by advanced diagnosis of primarily slow-growing tumors, as is assumed in Figure 1a, these slow-growing tumors should all be present in the breasts of non-screened women and become detected once they enter a screening program at a later time in their lives. Whether this happens has been studied in Norway⁸ and Sweden.⁹ We compared cumulative incidence rates in matched cohorts of women where the study group included all women who were invited for regular screening over 6 years. All women in the control group were invited to undergo a prevalence screen at the end of the study period when they were 6 years older. After these 6 years of follow-up, the cumulative incidence rate of invasive breast cancer was 16–22 % higher in the group screened several times during the observation period, and this difference did not disappear after one more screening round of both groups,⁸ or after several rounds.⁹ In fact, almost all overdiagnosis can be explained by the detection of tumors that would have regressed spontaneously.⁹ Based on US data, Fryback et al.¹¹ have suggested the same; however, they used a different approach. They modeled tumor growth and concluded that the “best model uses 42 % LMP tumors (tumors with low malignant potential). LMP tumors progress to a maximum of approximately 1-cm diameter, dwell at this size for 2 years, and then regress if undetected.

Examination of the best scoring model without regression of LMP tumors reveals an apparent depletion of the occult pool of localized invasive cancers.”

In the randomized PLCO trial of screening for prostate cancer with PSA, the cumulative number of prostate cancers were 2,541 for the screening group and 1,997 for the control group when the trial was closed after 6 years (absolute difference, 544 cases).¹² Two years after the trial closed the absolute difference was 422 cases, and seven years later the cumulative numbers were 4,250 and 3,815, respectively (absolute difference, 435 cases). The excess-incidence estimate of overdiagnosis is therefore: $(2,541-1,997)/1,997=1.22$ (22 %). Here, $544 - 435=109$ is the difference in the cumulative rates after 7 years of follow-up. In contrast, after 7 years of follow-up, overdiagnosis estimated with the classical lead-time adjustment is $4,250/3,815=1.12$ (12 %), illustrating how this method provides percentage estimates of overdiagnosis that are greatly dependent on study duration, are incomparable for various lengths of follow-up, and can therefore be misleading. Furthermore, prostate cancer incidence rates were similar in the previously screened and non-screened groups during the 2–7 years after the trial closed, indicating that PSA screening did not detect many tumors with long lead-times, but rather dormant tumors or tumors that spontaneously regress.

In conclusion, the excess-incidence approach for calculating overdiagnosis is preferable, because this method yields an estimate with a simple and intuitive interpretation. The lead-time approach depends on a method that is not transparent and usually has no interpretation. Further, estimates obtained with this model depend both on the number of screenings, as well as on the length of the follow-up after screening has stopped, making comparisons between studies impossible.

Acknowledgements: All authors declare no conflict of interest: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

Corresponding Author: Per-Henrik Zahl, DrMedSci; Norwegian Institute of Public Health, PO Box 4404, Nydalen 0403 Oslo, Norway (e-mail: Per-henrik.zahl@fhi.no).

REFERENCES

1. **Etzioni R, Gulati R, Mallinger L, Mandelblatt J.** Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. *Ann Intern Med.* 2013;158:831–839.
2. **Zahl P-H, Jørgensen KJ, Gøtzsche PC.** Overestimated lead-time in cancer screening has led to substantial under-estimating of overdiagnosis. *Br J Cancer.* 2013;109:2014–2019.

3. **Gøtzsche PC, Jørgensen KJ.** Screening for breast cancer with mammography. *Cochrane Database Syst Rev.* 2013;6, CD001877. doi:10.1002/14651858.CD001877.pub5.
4. **Zahl P-H, Strand BH, Mæhlen J.** Breast cancer incidence in Norway and Sweden during introduction of nation-wide screening: prospective cohort study. *BMJ.* 2004;328:921–924.
5. **Jørgensen KJ, Gøtzsche PC.** Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ.* 2009;339:b2587.
6. **Kalager M, Adami H-O, Bretthauer M, Tamimi RM.** Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian Screening program. *Ann Intern Med.* 2012;156:491–499.
7. **Boer R, Warmerdam P, de Koning H, et al.** Extra incidence caused by mammographic screening. *Lancet.* 1994;343:979.
8. **Zahl P-H, Mæhlen J, Welch HG.** The natural history of invasive breast cancers detected by screening mammography. *Arch Intern Med.* 2008;168:2311–2316.
9. **Zahl P-H, Gøtzsche P, Mæhlen J.** Natural history of breast cancers detected in the Swedish mammography screening program; a cohort study. *Lancet Oncol.* 2011;12:1118–1124.
10. **Zahl P-H, Mæhlen J.** Overdiagnosis of breast cancer after 14 years with mammography screening. *Tidsskr Nor Lægeforen.* 2012;132:414–417.
11. **Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL.** The Wisconsin breast cancer epidemiology simulation model. *J Natl Cancer Inst Monogr.* 2006;36:37–47.
12. **Andriole GL, Crawford ED, Grubb RL III, et al.** Prostate cancer screening in the randomized prostate, lung, colorectal and ovarian cancer screening trial: mortality results after 13 years follow-up. *JNCI.* 2012;104:125–132.