Tidsskriftet

Tidsskrift for Den norske legeforening

Statistical analysis of early cancer diagnosis requires new methods

PER-HENRIK ZAHL

per-henrik.zahl@fhi.no
Per-Henrik Zahl, cand.med, cand.scient. in mathematical statistics and MD PhD in biostatistics. The author has completed the ICMJE form and declares no conflicts of interest.

The methods for diagnosing cancer have traditionally been based on the concept that everything grows. However, immunotherapy and screening trials show that some tumours resolve spontaneously.

The traditional understanding of cancer, which forms the basis for diagnosis, treatment and statistical analyses, is based on the notion that tumours grow continuously and spontaneous regression is a rare phenomenon. Modern immunotherapy and experiences from mass screenings show that this is not always the case. This challenges the concept of lead time and the notion that early diagnosis always increases the chances of survival.

Lead time and sojourn time

Estimation of lead time, sensitivity and sojourn time are key concepts in the evaluation of early diagnosis of cancer. There are various definitions of lead time, meaning that estimates are not directly comparable. Long lead times are not statistically or biologically plausible, and there are limits for the detection of small tumours. Long lead times result in considerable overdiagnosis, and assumptions about long lead times are central to the evaluation of cancer screening [1].

Lead time is defined as the length of time by which the diagnosis is brought forward when a new diagnostic technique is introduced, or when routine screening for a disease is performed [2].

The *sensitivity* of a test is the probability that the test will successfully diagnose the disease. Thus (1-sensitivity) is the probability of not being able to diagnose the disease, even though it is present, i.e. the probability of a false negative finding (2). Low sensitivity can entail both a low probability of diagnosing a disease that is actually present (many false negatives), and a high probability that many small lesions that are detected later resolve spontaneously (false positives). In the former case, the screening test can give a false sense of security; in the latter case, much unnecessary anxiety is created in addition to overtreatment. It is generally assumed that sensitivity decreases as tumours shrink. The smaller the tumours, the more difficult they are to detect. Figure 1 illustrates the association between length of lead time and the sensitivity function.
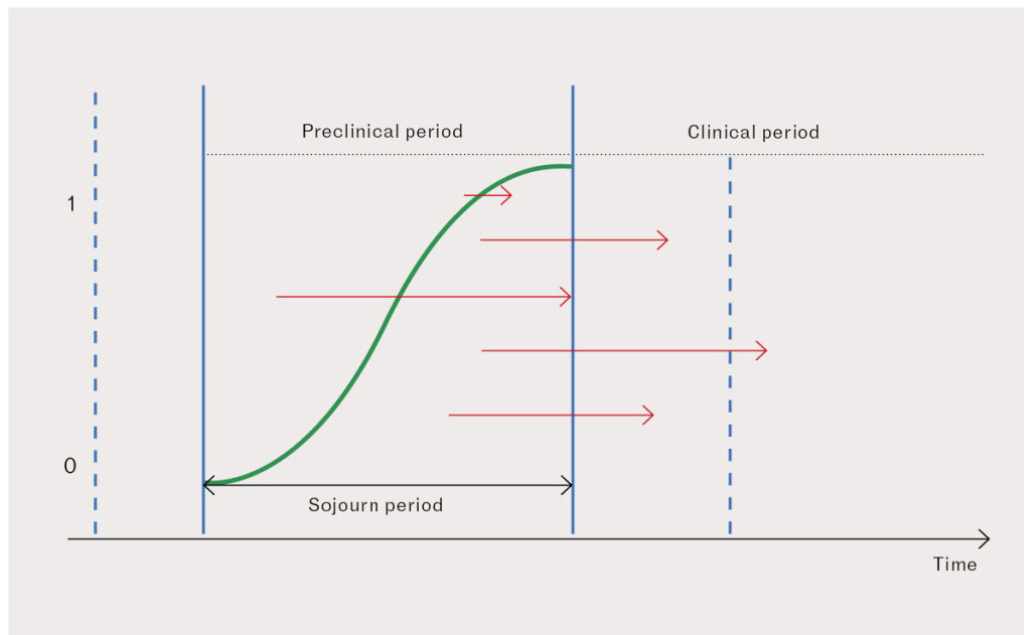

Figure 1 Association between length of lead time (red arrows) and sensitivity function (green curve). The sojourn period is between the two solid blue lines. The blue dotted line on the left is a new diagnostic method with lower detection threshold, the blue dotted line on the right is the old clinical detection threshold.

Sojourn time is defined as the interval in which it is theoretically possible to make a diagnosis (e.g. when a new diagnostic technique is introduced) before it is possible to make the diagnosis clinically. It can also be defined as the maximum lead time or an interval that encompasses all possible lead times (from the longest to the shortest possible lead time) (3). Note that these are not two identical definitions – the second definition extends into the clinical period. Sensitivity will typically be minimal or equal to zero at the start of the sojourn time, only to increase to 1 at the end of the period. When diagnostic methods become more sensitive, the detection threshold falls (illustrated with a dotted blue line on the left of Figure 1).

Not all cancer grows monotonically

A mathematical assumption for estimating lead time is that the disease is growing monotonically (either exponentially, log-normally or with incremental growth) (2). Tumours cannot resolve spontaneously in these models. Some tumours grow rapidly and others slowly (4). Some tumours grow so slowly that individuals die from other causes before the disease can be clinically diagnosed or cause symptoms. This is known as overdiagnosis (5). The fact that tumours do not behave uniformly means that traditional (exponential) growth models cannot describe the behaviour of all tumours. The entire modelling framework must therefore be revised.

> *«Since we cannot know in advance which tumours will continue to grow and which will disappear, we need to develop a statistical growth model that allows for various possible scenarios»*

Estimated lead time is an average. By definition, indolent subclinical tumours and tumours that resolve spontaneously at a subclinical level have no lead time (they do not cross the threshold for clinical detection), but theoretical lead times can nevertheless be defined for these types of tumours. This will result in a dramatic increase in the average lead time. Such theoretical lead times make little sense because they cannot be interpreted as the length of time by which the diagnosis is brought forward. Theoretical lead times are frequently estimated for screening and diagnosis of prostate cancer and breast cancer. The only sensible medical approach is to estimate lead times for tumours that are progressing and will develop into clinical cancer. These estimates are called clinical lead time (5).

It is easy to check whether there are many indolent tumours or a reservoir of slow-growing tumours in autopsy studies (6) or prevalence studies when commencing large-scale screening programmes (7).

Test 1 – estimate the proportion of tumours with a long lead time

If there are many tumours with long lead times (longer than the screening intervals), a high peak in prevalence should be observed in the first screening round (8). If screening takes place every two years and the detection rate is the same in the first and second screening round, the maximum lead time is two years. All new tumours detected in the second screening round were not diagnosable in the first screening round. If the rate in the second round is divided by the rate in the first round, the result is the proportion with a lead time of less than two years. When mammography screening was introduced in

Norway and Sweden, the detection rate in the first and second screening round was found to be approximately the same for women below the age of 60 years (9, 10). The maximum lead time for these women was around two years.

Lead time and volume doubling time

Lead time can easily be converted to volume doubling time. This gives a relative measurement for the length of time the diagnosis is brought forward in relation to total growth rate (5). After 19 volume doublings, the diameter is approximately 1 mm, but tumours start to spread long before this (11). Around 8–9 volume doublings later, tumours can be diagnosed using the most sensitive imaging modalities, such as MRI (Table 1) (12, 13).

Table 1

Number of cell doublings, number of cells, diameter, volume and medical relevance for breast cancer (11–13).

Number of cell doublings	Number of cells	Diameter (mm)	Volume (mm^3)	Medical relevance
0	1	0,012	0.000001	Primary cancer cell
19	524 288	1,0	0.52	Onset of metastasis (11)
28	268 435 456	8,0	268	Mammography detection threshold (13)
29	536 870 912	10,0	536	Clinical detection threshold (13)
31	2 147 483 648	16,0	2 148	Average diameter in randomised controlled trials (13)
40	2.2 × 1012	135,1	1 099 776	1 kg and death

If the preclinical period at screening is three volume doublings, the average lead time will be slightly more than 1.5 volume doublings. This sets a ceiling for what constitutes realistic lead times. If the volume doubling time is one year, it will take around 30 years from the start until the tumour can be clinically detected. This is unreasonable. Lead times of more than 1–2 years and preclinical periods of 3–7 years are not biologically plausible. Moreover, a decelerating growth rate can typically be observed after 29 volume doublings (when the tumour can be diagnosed clinically) (4). Such long lead times and decelerating growth rates also mean that, on average, people will live 10–20 years with the tumours after they are clinically diagnosable. This is also biologically unreasonable.

Test 2 – low sensitivity

Sensitivity is reported as an average estimate for sensitivity function. In fact, sensitivity has two interpretations. As mentioned in the introduction, this is the probability of successfully diagnosing the disease. Or alternatively, (1-sensitivity) is an estimate of conditions that resolve spontaneously. If there is a reservoir of tumours with long lead times that are detected with low sensitivity, this reservoir should gradually empty over the course of several screening rounds, resulting in a decreasing sequence of interval cancer rates (rate of cancer between two screening tests). If the interval cancer rate decreases over time, this indicates a large number of tumours with long lead time being detected with low sensitivity. This has not been observed for breast cancer.

Test 3 – compensatory fall when screening stops

Some tumours grow so slowly that some die from other causes before a disease detected through screening develops into a clinical disease. The number is determined by calculating the cumulative incidence for a birth cohort from the start of screening until approximately ten years after screening minus the rate in the absence of screening. Almost all increases in the incidence of disease during a screening period will be compensated for by a reduction in incidence when screening stops. In the case of mammography screening, there will normally only be around a 2 % increase in breast cancer when the lead time is 4.8 years (8). Following a large increase in incidence at screening, an equally large compensatory fall in incidence should subsequently be observed (8). After screening ends, the observed incidence of the disease over a ten-year period is compared to the expected incidence without screening.

Adjustment for lead time

Incidence rates and overdiagnosis should only be adjusted for clinical lead time (5). It is important to distinguish between the growing incidence of clinical cancer and the growing incidence of overdiagnosed tumours. There is no reason to believe that the risk factors are the same for overdiagnosed tumours as for clinical cancer. Overdiagnosis may therefore be a major confounding factor and explain why many risk factors for a cancer diagnosis are not risk factors for mortality from a cancer diagnosis (14).

> *«It is important to distinguish between the growing incidence of clinical cancer and the growing incidence of overdiagnosed tumours»*

If individuals have many indolent tumours in the prostate (15) and kidneys (16) or tumours in the kidneys (16) and neuroblastoma (17) that resolve spontaneously, then screening obviously results in considerable overdiagnosis. Spontaneous regression is much more common that most doctors think – this conclusion can also be drawn based on the outcome of modern immunotherapy, which is based on this principle (18).

Adjustment for lead time is a dubious statistical practice because estimated lead time varies depending on how lead time is defined, the diagnostic method used and the extent to which doctors and individuals look for small clinical tumours. Moreover, lead time is speculative because it is a non-observable quantity. There is no reason to believe that adjusting for lead time reduces bias; it may just as easily increase bias.

REFERENCES

1. The Research Council of Norway. Research-based evaluation of the Norwegian Breast Cancer Screening Program. https://www.forskningsradet.no/siteassets/publikasjoner/1254012138940.pdf Accessed 30.4.2024.

2. Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. Biometrics 1984; 40: 1–14. [PubMed][CrossRef]

3. Oxford reference. Sojourn time. https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100516669 Accessed 12.3.2024.

4. Spratt JA, von Fournier D, Spratt JS et al. Decelerating growth and human breast cancer. Cancer 1993; 71: 2013–9. [PubMed][CrossRef]

5. Zahl P-H, Jørgensen KJ, Gøtzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. Br J Cancer 2013; 109: 2014–9. [PubMed][CrossRef]

6. Nielsen M, Jensen J, Andersen J. Precancerous and cancerous breast lesions during lifetime and at autopsy. A study of 83 women. Cancer 1984; 54: 612–5. [PubMed][CrossRef]

7. Smith-Bindman R, Chu PW, Miglioretti DL et al. Comparison of screening mammography in the United States and the United kingdom. JAMA 2003; 290: 2129–37. [PubMed][CrossRef]

8. http://dx.doi.org/10.1016%2FS0140-6736(94)90105-8 doi: 10.1016/S0140-6736(94)90105-8. Boer R, Warmerdam P, de Koning H et al. Extra incidence caused by mammographic screening. Lancet 1994; 343: 979 [Letter]. 10.1016/S0140-6736(94)90105-8.10.1016/S0140-6736(94)90105-8[CrossRef]

9. Zahl P-H, Maehlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. Arch Intern Med 2008; 168:

2311–6. [PubMed][CrossRef]

10. Zahl P-H, Gøtzsche PC, Mæhlen J. Natural history of breast cancers detected in the Swedish mammography screening programme: a cohort study. Lancet Oncol 2011; 12: 1118–24. [PubMed][CrossRef]

11. Folkman J, Watson K, Ingber D et al. Induction of angiogenesis during the transition from hyperplasia to neoplasia. Nature 1989; 339: 58–61. [PubMed][CrossRef]

12. DENSE Trial Study Group. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. N Engl J Med 2019; 381: 2091–102. [PubMed][CrossRef]

13. Gøtzsche PC, Jørgensen KJ, Zahl P-H et al. Why mammography screening has not lived up to expectations from the randomised trials. Cancer Causes Control 2012; 23: 15–21. [PubMed][CrossRef]

14. Strand BH, Tverdal A, Claussen B et al. Is birth history the key to highly educated women's higher breast cancer mortality? A follow-up study of 500,000 women aged 35-54. Int J Cancer 2005; 117: 1002–6. [PubMed][CrossRef]

15. Draisma G, Boer R, Otto SJ et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst 2003; 95: 868–78. [PubMed][CrossRef]

16. Jewett MA, Mattar K, Basiuk J et al. Active surveillance of small renal masses: progression patterns of early stage kidney cancer. Eur Urol 2011; 60: 39–44. [PubMed][CrossRef]

17. Schilling FH, Spix C, Berthold F et al. Neuroblastoma screening at one year of age. N Engl J Med 2002; 346: 1047–53. [PubMed][CrossRef]

18. The nobel prize. The Nobel Prize in Physiology or Medicine 2018. https://www.nobelprize.org/prizes/medicine/2018/press-release/ Accessed 12.3.2024.