



OPEN ACCESS

Evidence categories in systematic assessment of cancer overdiagnosis

Anton Barchuk ,^{1,2} Niko K Nordlund ,¹ Alex L E Halme ,¹
Kari A O Tikkinen ^{3,4}

10.1136/bmjebm-2024-113529

¹Faculty of Medicine,
University of Helsinki,
Helsinki, Finland

²Department of Medical
Informatics, Erasmus MC,
Rotterdam, The Netherlands

³Department of Urology,
Helsinki University Hospital,
Helsinki, Finland

⁴Department of Surgery,
Päijät-Häme Central Hospital,
Lahti, Finland

Correspondence to:
Dr Anton Barchuk; anton.
barchuk@helsinki.fi



© Author(s) (or their
employer(s)) 2025. Re-use
permitted under CC
BY-NC. No commercial
re-use. See rights and
permissions. Published by
BMJ Group.

To cite: Barchuk A,
Nordlund NK, Halme ALE,
et al. *BMJ Evidence-Based
Medicine* Epub ahead of
print: [please include Day
Month Year]. doi:10.1136/
bmjebm-2024-113529

Abstract

The phenomenon of cancer overdiagnosis, the diagnosis of a malignant tumour that, without detection, would never lead to adverse health effects, has been reported for several cancer types in different populations. There has been an increase in studies focused on overdiagnosis, creating an opportunity to synthesise evidence on specific cancer types. However, studies that systematically assess evidence across different research domains remain scarce, with most of them relying on data from studies that already mentioned overdiagnosis as a potential concern. In this review, we consider several evidence categories that are used to systematically assess the presence and magnitude of overdiagnosis, including (1) data from cancer surveillance, (2) studies exploring the 'true' prevalence of cancer in the population, (3) studies that explore the use of diagnostics and its effect on incidence and mortality and (4) studies that explore changes and progress in cancer management and its effect on cancer mortality. This article highlights the strengths and weaknesses of different evidence categories, provides examples of studies on different cancer types and discusses how these categories can help synthesise evidence on cancer overdiagnosis.

Introduction

Cancer overdiagnosis occurs when a malignant tumour—that would never lead to any harm if left undetected—is found. This has been reported for several cancer types, such as thyroid, prostate and breast cancers, and is more common in populations with better access to diagnostics.¹ In recent years, there has been an increase in studies focused on overdiagnosis, creating opportunities to synthesise evidence on specific cancer types. For instance, a recent scoping review on overdiagnosis in malignant melanoma assessed available evidence from studies in different populations and settings to conclude the presence and magnitude of this phenomenon.² The growing awareness of the potential harms associated with cancer overdiagnosis is also reshaping how future healthcare providers are trained.³ This shift in teaching is particularly significant for those who once believed that earlier detection was unequivocally beneficial. Table 1 provides explanations of key concepts relevant to overdiagnosis studies.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Cancer overdiagnosis, the detection of tumours that would not cause harm if left untreated, is a recognised issue in oncology, particularly with the increased use of advanced diagnostic technologies. However, systematic reviews on this topic are uncommon, and the categorisation of studies and standard approaches to data extraction is still limited.

WHAT THIS STUDY ADDS

⇒ We propose several evidence categories that can be used to synthesise evidence on cancer overdiagnosis and aid in detecting it across different populations. The article provides examples of studies, highlighting their strengths and limitations, and offers a structured approach to analysing the data, allowing for a more comprehensive understanding of overdiagnosis patterns.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Systematic evidence synthesis using the proposed evidence categories can improve the identification of cancer overdiagnosis and clarify its public health impact. Integrating data from various research domains can lead to a more balanced approach to overdiagnosis assessment, prioritising patient outcomes and optimising resource allocation within healthcare systems.

Studies have typically assessed cancer overdiagnosis by analysing cancer surveillance data to identify specific epidemiological patterns ('signatures' or 'fingerprints') that may explain trends.^{4–6} Researchers have sometimes interpreted the same trends as supporting or opposing cancer screening.^{7,8} However, studies that systematically assess evidence from different research domains are still rare, with most relying on data from studies that already flag overdiagnosis as a concern. Overdiagnosis research, therefore, needs

Research methods and reporting

Table 1 Terms and concepts used in cancer overdiagnosis research (in alphabetical order)

Term/concept	Explanation
Active surveillance (cancer management option)	The strategy of following patients with cancer without providing any treatment until disease progression is detected.
Awareness campaigns	Media and public campaigns are needed to improve the general population's knowledge of cancer prevention, symptoms and detection methods. Campaigns are often focused on symptoms of specific cancer types.
Cancer progression/regression	An increase or decrease in the size of a tumour or in the extent of cancer that can be also reflected in up/down staging of the disease.
Cancer surveillance	Methods to monitor cancer burden. Cancer incidence data are typically collected at population-based cancer registries, and mortality data through the civil registration systems. Comparability, accuracy and completeness of the cancer incidence and mortality data vary globally.
Diagnostic utilisation	Measures to assess the use of specific diagnostic services expressed in absolute numbers or rates.
Earlier detection	A phenomenon where specific diagnostics improve the healthcare system's ability to provide an earlier diagnosis for patients with cancer. Terms like 'early detection' may overestimate the ability to detect cancer early enough to be cured.
False positives/negatives	Diagnostic test result that incorrectly indicates the presence or an absence of cancer.
Harms of screening	Harms of screening include physical effects and psychological effects (caused by false positive, false negative test results, work-up after initial test and overdiagnosis), financial strain and opportunity costs. Harms cannot be completely avoided but can be reduced within organised cancer screening programmes.
Incidence	A group of epidemiological measures that reflect the number and rates of newly diagnosed cancer cases in a specified period.
Indolent tumours	Slower-growing tumours that may not be causing symptoms and can be detected during an autopsy and screening test.
Lead-time bias	Bias in studies that use cancer survival to assess the effectiveness of earlier detection and screening. Earlier diagnosis always leads to increased survival, but not necessarily results in improved outcomes (reduction in the number of cancer deaths).
Length-time bias	A bias in studies comparing symptomatic tumours to those found via screening, as screening tends to detect slower-growing tumours with better prognoses, potentially overestimating screening benefits.
Mortality	A group of epidemiological measures that reflect the number and rates of deaths attributed to cancer in a specified period.
Natural history	The process of cancer development in an individual uninterrupted by any intervention or treatment, from cancerogenesis to death, attributed to cancer or other causes.
Opportunistic screening	Cancer screening that is outside of an organised screening programme with tests and procedures that may not adhere to screening guidelines and recommendations.
Overdiagnosis (of a cancer)	The diagnosis of a malignant tumour that, without detection, would never lead to any adverse health effects, such as symptoms or death.
Positive/negative predictive value	The probability of presence or absence of disease, given a positive or negative test result. Predictive values depend on the prevalence of disease in the population.
Positivity (in cancer screening)	The proportion of positive test results among all participants tested in a screening trial/programme.
Screening	Earlier diagnostics in an attempt to identify certain early-stage cancer types and pre-cancer lesions in the asymptomatic general population. It may vary from highly organised screening programmes to opportunistic activities with little quality control.
Screening sensitivity/specificity	Sensitivity indicates how well screening detects the condition of interest, and specificity describes how accurately subjects free of the target disease are identified.
Stage migration	Improved cancer diagnostics shift patients to higher disease stages by detecting cases earlier or at a more sensitive level, increasing survival in each stage without necessarily improving overall patient outcomes.
Survival	A group of epidemiological measures that reflect the risk of death in patients with cancer.

a systematic approach to evaluate the role of emerging cancer diagnostics, which increasingly affects the epidemiology of many cancer types.

The individuals harmed by unnecessary tests are often not the same as those who benefit from earlier detection. The balance of benefits and harms of cancer diagnostics should be assessed at the population level through randomised trials. However, cancer screening trials are often limited to high-resource settings and may lack the required follow-up to establish trustworthy evidence on overdiagnosis. In addition, due to the extended follow-up required, some trials may become outdated, making their overdiagnosis estimates applicable only to selected settings, time periods and populations.^{9–11}

In this paper, we examine a range of evidence categories that can be relevant in the context of cancer overdiagnosis studies: (1)

cancer surveillance data, (2) studies assessing the true prevalence of cancer in the population, (3) research on diagnostic utilisation and its impact on cancer incidence and mortality and (4) studies investigating changes and progress in cancer management and its effect on cancer mortality. We explore why and how these specific categories should be considered in overdiagnosis research, discuss practical applications for systematically synthesising evidence on overdiagnosis and highlight their potential limitations.

Cancer surveillance

Data from population-based cancer surveillance are among the most commonly used domains in studies that address cancer overdiagnosis.^{4 5 12 13} Epidemiological measures used to assess changes in cancer trends focus on cancer incidence and mortality in the general population. The risk of death in patients with cancer is also

assessed through survival analysis, referred to as cancer survival. Studies often involve linear modelling, where rates are analysed as a function of time, and information on the relative change of the rates is captured over a specified period. For example, in a recent study, long-term annual trends of mortality and incidence rates of thyroid cancer from 43 countries were compared, and the number of overdiagnosed cases was estimated.¹⁴ Trend analyses can be supplemented by breakpoint and age-period-cohort analyses. In the age-period-cohort analysis of kidney cancer incidence rates from 1978 to 2007 in 16 populations, overdiagnosis was hypothesised as one of the reasons for changes.¹⁵ Stratification by stage can help identify stage-specific changes and determine whether shifts in incidence are attributed to early-stage disease.¹⁶ Increased early-stage cancer incidence accompanied by unchanged late-stage disease incidence might be an indicator of overdiagnosis. This approach may have limitations, as the changes in the incidence of late-stage disease over time may be affected by changes in risk factor prevalence and diagnostic accuracy. Comparison across and outside the age groups targeted by screening intervention and the addition of proper control groups to comparison might be useful to overcome these difficulties;¹⁷ still, the availability of information about screening participation may be a limitation. Individual-level studies using national databases that capture individual characteristics, screening participation and relevant outcomes may help refine this approach.

Cancer surveillance helps detect meaningful changes in cancer trends. It also helps to suspect underlying causes of cancer. Surveillance can be considered as an instrument for overdiagnosis research based on the reasonable assumption that increased diagnostic activity leads to higher cancer incidence. Generally, a substantial increase in incidence without similar changes in mortality, accompanied by an increase in survival, is considered a sign of diagnostic effects and overdiagnosis.^{4,5} A well-known example is the increase in thyroid cancer incidence in South Korea, which is consensually explained by the use of ultrasound in regular check-ups.¹² Cancer surveillance data have also been used to assess the magnitude of overdiagnosis in prostate and thyroid cancer when baseline historical trends were compared with those observed in the period of the increased diagnostic activity.^{6,14}

The methods and criteria to quantify overdiagnosis using surveillance data are not, however, standardised. There is no consensus on how large changes in cancer rates are considered substantial. As periods used to identify meaningful changes are often arbitrary, studies may not be comparable. In addition, stage-specific information is often unavailable or incomplete in population-based cancer registries that consolidate and produce cancer incidence data.

Cancer surveillance studies typically have an ecological study design without data on exposures. This limits the opportunities for causal inferences. Cancer incidence rates reflect individuals diagnosed each year. In contrast, mortality and survival rates typically reflect individuals diagnosed in previous years or decades. The gap between diagnosis and death depends on many factors: the lead time of a given cancer, the characteristics of the tumour and the effectiveness of management, including diagnostics and treatments.¹⁸

Although cancer surveillance studies are useful for detecting trends, they have highlighted the need for reproducible algorithms and clear quantitative criteria to standardise efforts to quantify overdiagnosis. Pooling the results of studies that use different methods to analyse the same population-based datasets is not the most optimal approach to generating reliable evidence. Systematic synthesis in this evidence category is likely most useful when data

from different regions acquired through similar methodologies are considered.

Changes in diagnostic practices and effect of diagnostic activities

Diagnostic utilisation studies

The surge in thyroid cancer incidence, often linked to opportunistic ultrasonography screening, exemplifies how diagnostic volume changes can lead to overdiagnosis, as seen in South Korea and globally.^{12,13} Years before overdiagnosis became a recognised issue, a French study emphasised the need to account for shifts in diagnostic, medical, surgical and pathological practices when assessing thyroid cancer incidence,¹⁹ validating earlier concerns.²⁰ Indeed, the thyroid cancer 'diagnostic epidemic' had been anticipated, with the potential reservoir of tumours noted already in 1985.²¹ Despite this, many individuals and healthcare systems have faced unnecessary diagnoses.¹⁴

Concerning diagnostic utilisation, some cancers have been included in or considered for population-based screening programmes, which are organised to detect cancer in asymptomatic populations. Examples of such cancers include breast, cervical and colorectal cancer.¹ Several countries were considering screening programmes for different cancer types, for example, thyroid cancer, melanoma and stomach cancer. Regardless of whether a country has an official national programme, the rapid advancement of technologies and lack of evidence-based decision-making have led to the widespread use of diagnostic tests that have never been tested in randomised trials as screening methods.²²

Cancer screening registries can provide individual-level data on diagnostic tests for cancers included in or considered for population-based screening programmes. Such data are, however, available in only a few regions,²³ and these registers typically do not capture information on opportunistic testing. These registries are crucial for identifying the participation rates and evaluating test characteristics, such as positivity and positive predictive value.²⁴ Even in countries with established screening programmes and similar disease incidence rates, these characteristics vary and, sometimes, point to potential overdiagnosis.²⁵

For opportunistic screening programmes and modalities such as ultrasound, MRI, prostate-specific antigen (PSA) and other tests, it is much more challenging to differentiate between screening and diagnostic use in symptomatic patients or follow-up contexts. Some studies report changes in biopsy utilisation that could be used as a proxy for primary diagnostic activities.^{26,27}

Unfortunately, individual-level studies that capture diagnostic procedures and outcomes (lesions detected) are rare outside screening trials and programmes. Studies based on individual data from multiple sources could bridge this gap, allowing large-scale characterisation that includes diagnostic tests, outcomes and reasons for referral.²⁸

Cancer screening individual-level studies

Cancer screening research serves as the foundation for many publications that address cancer overdiagnosis, with results often drawn from individual-level studies. Comparison of exposed groups (individuals who were offered or had certain screening tests) and unexposed groups can help identify the difference in incidence and mortality, forming the basis for evaluating the balance between the benefits and harms of certain diagnostic interventions or combinations of interventions.²⁹

Many screening trials are conducted in actual practice settings and are affected by group contamination and non-compliance.

Research methods and reporting

In the screening arm, some participants may not adhere to the protocol and forgo screening, while in the control group, some may undergo screening outside the trial. For example, a high frequency of screening in the control arm was one of the factors that may have contributed to the lack of observed prostate cancer mortality reduction in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial.³⁰ While contamination can underestimate the mortality reduction, it may have an even more pronounced absolute effect on excess incidence. In a Finnish randomised prostate cancer screening trial, adjustment for contamination nearly doubled both mortality reduction and excess incidence. Without correction, there were 8 fewer deaths per 10000 participants in the screening arm, increasing to 14 after correction. Incidence increased from 165 to 324 excess cases per 10000 participants when contamination was accounted for.³¹

In addition, screening trials have limited ability to identify the risk of uncontrolled use of the same diagnostics outside of research settings,³² and those who defend the beneficial effects of screening on cancer mortality based on trial results are often faced simultaneously with the need to combat opportunistic screening activities.

In addition, individual-level screening studies are often conducted in higher-income countries, which have access to quality control and a more balanced approach to healthcare utilisation. At the same time, possible overdiagnosis can escalate when the same diagnostic technologies are introduced without proper quality control in lower-resource settings.²²

The prevalence of subclinical cancer

While overdiagnosis is related to the properties and dissemination of diagnostic tests, it is also influenced by tumour characteristics. The discussion about the natural history of cancer, or the course of a disease uninterrupted by treatment, when applied to cancer screening and earlier detection studies, led to several important practical conclusions. First, it was suggested that not all asymptomatic cancers and precancers are destined to become invasive and symptomatic malignancies,³³ and there may be a substantial reservoir of potentially detectable subclinical disease.²⁰ Second, it was suggested that including these cancer cases in statistics could bias cancer survival as a measure of cancer control progress.^{34 35} While the second conclusion was more related to the interpretation of surveillance studies, the first requires empirical evidence and data to suggest the extent of such a reservoir.

Two important study types can be used to quantify the detectable subclinical cancer in the population: autopsy studies and cross-sectional diagnostic surveys. Autopsy studies are a unique source of information on the prevalence of asymptomatic tumours in people whose cause of death was not related to that specific cancer. However, the complexity of these studies and the decreasing proportion of autopsies performed worldwide make this type of study rarer today. Systematic reviews and individual-level autopsy studies are available for prostate,³⁶ lung,³⁷ colorectal,³⁸ thyroid,³⁹ breast⁴⁰ and kidney⁴¹ cancers. The higher prevalence of cancer in autopsy suggests a higher likelihood of the presence of overdiagnosis.

Cross-sectional diagnostic surveys and the first rounds of screening trials can also provide information on the prevalence of asymptomatic tumours in the population, especially with recent advances in diagnostics that offer a resolution to capture smaller lesions.^{42 43} Cross-sectional studies have several important limitations: they are not always representative, and the confirmation of malignancy in these studies requires a biopsy, an intervention potentially leading to adverse effects. Also, longer-term follow-up

is needed to prove that small asymptomatic tumours represent true overdiagnosis. Recent lung and prostate screening trials have shown a substantial number of lesions detected by modern imaging but have also introduced solutions to identify detected nodule features and volume doubling time to avoid unnecessary biopsy, highlighting how understanding the natural history of diseases can help shape screening programmes.⁴⁴

Lastly, the prevalence of any disease depends on healing rates, which can be spontaneous. This phenomenon was described for several malignant tumours (eg, melanoma, kidney cancer), but the evidence is scarce, and rates of spontaneous regression are unclear.^{45 46} Theoretically, increased diagnostic activities are likely to cause overdiagnosis in tumours where spontaneous regression is more common, and this phenomenon should not be dismissed.

Changes in treatment

When discussing the benefits of earlier cancer diagnosis, the ultimate goal is to reduce mortality. Overdiagnosis is an unavoidable consequence of earlier cancer detection and must be weighed against possible benefits. As mortality reduction and overdiagnosis affect individuals differently, the balance of benefits and harms is difficult to measure and subjective. This highlights the need for informed and shared decision-making. Mortality rates for many cancers have declined in high-income countries⁴⁷ coinciding with both improved diagnostics and new treatments. Distinguishing these effects is essential for evidence-based, sustainable screening policies.

Although treatment progress reduces mortality for certain types of cancer,⁵ in reality, this process is often not fast. The introduction of new drugs can affect mortality years later. Assessing the results of cancer treatment of advanced disease can help identify overall progress in cancer management. For example, in advanced breast cancer, survival after metastatic recurrence increased from 1.9 years in 2000 to 3.2 years in 2019.⁴⁸ These changes should be interpreted cautiously as stage migration and better follow-up may inflate survival due to significant diagnostic advancements.

Another way that treatment information can contribute to overdiagnosis research is through the implementation of watchful waiting and active surveillance strategies, indicating that certain cancer cases may not require treatment and should not have been detected in the first place. Predicting the malignant potential of detected tumours is difficult, so introducing additional markers and diagnostic strategies can help identify non-progressive tumours before any invasive intervention begins. Watchful waiting strategies have been introduced for thyroid⁴⁹ and prostate cancer.⁵⁰ The ongoing discussion on the treatment of precancers in the cervix uteri⁵¹ and breast⁵² not only highlights possible adverse events linked to overdiagnosis but also paves the way for reevaluating disease categories in the future.

A somewhat underused source of data is studies that describe patients who refused conventional treatment. These studies often focus on the effect of complementary or alternative treatments, which are seen as a threat in light of the progress of conventional treatments. However, they also show potential for tumour progression when detected relatively early. For example, in a study from the National Cancer Database in the USA, a substantial proportion of patients with prostate (86%), breast (58%) and even lung (20%) cancer diagnosed between 2004 and 2013 were alive after 5 years without any specific treatment.⁵³

Sometimes, the effects of both diagnostic and treatment interventions are assessed in simulation exercises. For example, the combined effect of screening and treatment was shown to be associated with the reduction of breast cancer mortality in

Table 2 Evidence categories, studies and measures that inform overdiagnosis research

Categories	Information obtained for overdiagnosis	Methods and measures	Strengths	Limitations	Examples
Cancer surveillance	Help identify meaningful changes in incidence, mortality and survival that can point to overdiagnosis	Trend analysis: breakpoint regression, evaluation of breakpoint and annual per cent changes. Relative rates or risks between cohorts and periods in age–period–cohort analysis. Stratification by age and stage is often applied.	Population-based studies can use national cancer registry data, which typically include almost all patients with cancer.	An ecological study design that cannot be directly used for causal inference or exploring reasons for incidence change. Requires high-quality cancer registry data, which is only available in some countries.	The estimation of the magnitude of prostate cancer in Australia, ⁶ mortality and incidence trends in European countries, ⁴ age–period–cohort analysis for kidney cancer rates in 16 countries worldwide. ¹⁵
Cancer screening evaluation and diagnostic utilisation	Helps to identify the changes in use and effect of diagnostic interventions potentially associated with cancer overdiagnosis. Also helps evaluate the balance between benefits (mortality reduction) and harms (excess incidence in the intervention arm) of specific diagnostic interventions.	Randomised trials, cohort and case–control studies that report relative risks or absolute risk differences in incidence and mortality. Cross-sectional diagnostic survey report on the use of diagnostic methods. Screening registries can help identify test characteristics in the population.	Pragmatic randomised trials can provide trustworthy evidence about overdiagnosis. Diagnostic surveys can provide reliable information about the spread of diagnostic technologies. Cancer register can reliably monitor the real-world implementation of screening.	Randomised trials may have limited applicability. Contamination in screening trials may bias. Diagnostic utilisation studies do not always report screening as an indication, and registers may miss information from private healthcare providers.	Estimating the rate of overdiagnosis from prostate cancer screening in Finland, ⁵⁶ performance indicators for mammography screening in Europe. ²⁵
Natural history of cancer	Helps to identify reservoirs of tumours for detection and overdiagnosis.	Cross-sectional studies, the first round of screening trials or programmes and autopsy studies that report the prevalence of indolent cancers.	Provides a biological basis for overdiagnosis research and helps estimate the prevalence of conditions in the populations.	Observational cross-sectional studies that may be biased and not representative of the general population, eg, might underestimate or overestimate true prevalence.	Autopsy study of occult papillary carcinoma of the thyroid in Finland, ²¹ the prevalence of renal cancer detected by abdominal ultrasonography (meta-analysis of studies from Japan, Canada, Germany, USA, Turkey, Finland, Bahrain, UK). ⁴²
Treatment patterns and effectiveness	Helps to identify the effect of different treatment options and possible effects on cancer mortality and clarify the effects of diagnostics.	Randomised trials, active surveillance studies and studies that can inform of the effects of refusal to receive conventional treatment.	Randomised trials provide the most trustworthy evidence of the treatment effect and watchful waiting/active surveillance strategies.	Observational studies may be biased, and distinguishing the effects, treatment and diagnostics may not be possible, especially when both are introduced simultaneously. Simulation studies require all assumptions to hold.	Changes in survival in metastatic breast cancer with treatment advances in USA, ⁴⁸ changes of practices in the diagnosis of thyroid cancer in France. ⁴⁹

US women.⁵⁴ However, while simulation studies are useful, it is still difficult to prove that all assumptions hold and results have external validity without data from trials.

Conclusion

This article outlines evidence categories that can support a systematic, proactive approach to identifying and synthesising evidence on overdiagnosis. We discuss several interrelated evidence categories that can help evaluate the balance of benefits and harms of diagnostics and study cancer overdiagnosis. Additionally, we outline the strengths and weaknesses of each category, with examples of studies provided in table 2. Figure 1 schematically illustrates possible relationships between the evidence categories described.

While conducting systematic reviews, we suggest researchers formulate search strategies to cover a broad range of relevant evidence categories. Research papers relevant to synthesising

evidence on overdiagnosis may not explicitly use terms like ‘overdiagnosis’ or ‘overdetection’ and thus may be inappropriately dismissed. To address this, we urge researchers to consider broadening their search criteria and scope, expanding it beyond surveillance studies and screening trials to include studies on the natural history of cancer, research exploring reservoirs of subclinical undetected cancers, and studies that cover diagnostics utilisation and treatment effects. Including diverse study types may provide a more comprehensive understanding of overdiagnosis and its relationship to mortality reduction.

There are also several ways researchers can use evidence categories to design future screening trials. First, trials should assess overdiagnosis and mortality reduction as critical outcomes of potential screening interventions, aiming to minimise overdiagnosis while maximising mortality reduction. Second, when strong evidence for cancer overdiagnosis exists, deimplementation studies—traditionally focused on treatment and less frequently on

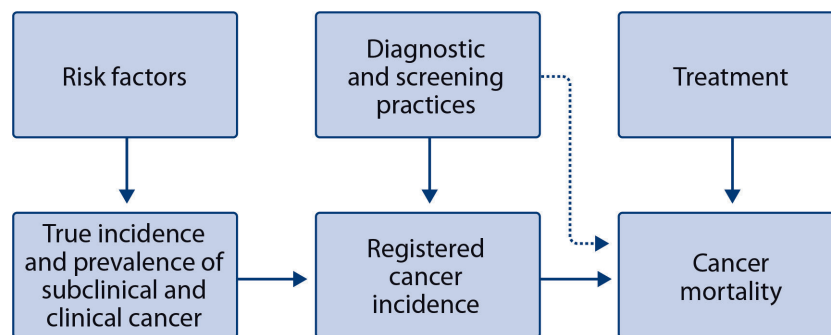


Figure 1 The role of diagnostic practices, screening and treatment in the pathway from risk factors to cancer incidence and mortality (diagram illustrates the flow from risk factors to cancer mortality: risk factors influence the true incidence and prevalence of cancer, which, along with diagnostic and screening practices, determines the registered cancer incidence; treatment impacts cancer mortality rates and diagnostic practices (screening) might also affect mortality).

diagnostics or screening⁵⁵—should be designed to reduce the harms of unnecessary diagnostics across various healthcare settings.

This review aims to encourage a broader, inclusive discussion that can lead to a more structured appraisal of the available evidence in overdiagnosis research. Overdiagnosis is a controversial topic where heated arguments are common. It often becomes prone to speculations without a consensual approach to the available evidence. As diagnostic tools become more sensitive, overdiagnosis is likely to increase across more cancer types. A systematic approach to evidence on overdiagnosis can help detect those changes early enough to prevent adverse effects and maximise the benefits of cancer diagnostics.

X Anton Barchuk @AntonBarchuk, Alex L E Halme @halmealex and Kari A O Tikkinen @KariTikkinen

Contributors AB, NKN, ALEH and KAOT planned and designed the outline of the submitted manuscript. AB drafted the first version of the manuscript. NKN, ALEH and KAOT critically reviewed and edited the manuscript. All authors approved the manuscript. AB acted as guarantor.

Funding KAOT reports funding from the Research Council of Finland (353026), Helsinki University Hospital State Research Funding (TYH2023236), Sigrid Jusélius Foundation and Vyborg Tuberculosis Foundation.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Anton Barchuk <http://orcid.org/0000-0002-4629-3326>
 Niko K Nordlund <http://orcid.org/0009-0007-6681-9675>
 Alex L E Halme <http://orcid.org/0000-0001-6597-818X>
 Kari A O Tikkinen <http://orcid.org/0000-0002-1389-8214>

References

- Dunn BK, Woloshin S, Xie H, *et al*. Cancer overdiagnosis: a challenge in the era of screening. *J Natl Cancer Cent* 2022;2:235–42.
- Bjørch MF, Gram EG, Brodersen JB. Overdiagnosis in malignant melanoma: a scoping review. *BMJ Evid Based Med* 2024;29:17–28.
- Brodersen J, Schwartz LM, Heneghan C, *et al*. Overdiagnosis: what it is and what it isn't. *BMJ Evid Based Med* 2018;23:1–3.
- Karim-Kos HE, de Vries E, Soerjomataram I, *et al*. Recent trends of cancer in Europe: a combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. *Eur J Cancer* 2008;44:1345–89.
- Welch HG, Kramer BS, Black WC. Epidemiologic Signatures in Cancer. *N Engl J Med* 2019;381:1378–86.
- Glasziou PP, Jones MA, Pathirana T, *et al*. Estimating the magnitude of cancer overdiagnosis in Australia. *Med J Aust* 2020;212:163–8.
- Autier P, Boniol M, Koechlin A, *et al*. Effectiveness of and overdiagnosis from mammography screening in the Netherlands: population based study. *BMJ* 2017;359:j5224.
- Sankatsing VDV, van Ravesteyn NT, Heijnsdijk EAM, *et al*. The effect of population-based mammography screening in Dutch municipalities on breast cancer mortality: 20 years of follow-up. *Int J Cancer* 2017;141:671–7.
- Boyle P. Faecal occult blood testing (FOBT) as screening for colorectal cancer: the current controversy. *Ann Oncol* 2002;13:16–8.
- Tikkinen KAO, Dahm P, Lytvyn L, *et al*. Prostate cancer screening with prostate-specific antigen (PSA) test: a clinical practice guideline. *BMJ* 2018;362:k3581.
- Adami HO, Kalager M, Bretthauer M. The Future of Cancer Screening—Guided Without Conflicts of Interest. *JAMA Intern Med* 2023;183:1047–8.
- Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer 'epidemic'—screening and overdiagnosis. *N Engl J Med* 2014;371:1765–7.
- Vaccarella S, Franceschi S, Bray F, *et al*. Worldwide Thyroid-Cancer Epidemic? The Increasing Impact of Overdiagnosis. *N Engl J Med* 2016;375:614–7.
- Li M, Dal Maso L, Pizzato M, *et al*. Evolving epidemiological patterns of thyroid cancer and estimates of overdiagnosis in 2013–17 in 63 countries worldwide: a population-based study. *Lancet Diabetes Endocrinol* 2024;12:824–36.
- Znaor A, Laversanne M, Bray F. Less overdiagnosis of kidney cancer? an age-period-cohort analysis of incidence trends in 16 populations worldwide. *Int J Cancer* 2017;141:925–32.
- Ryzhov A, Corbex M, Piñeros M, *et al*. Comparison of breast cancer and cervical cancer stage distributions in ten newly independent states of the former Soviet Union: a population-based study. *Lancet Oncol* 2021;22:361–9.

- 17 Jørgensen KJ, Gøtzsche PC, Kalager M, *et al.* Breast Cancer Screening in Denmark: A Cohort Study of Tumor Size and Overdiagnosis. *Ann Intern Med* 2017;166:313–23.
- 18 Ellis L, Woods LM, Estève J, *et al.* Cancer incidence, survival and mortality: explaining the concepts: incidence, survival and mortality. *Int J Cancer* 2014;135:1774–82.
- 19 Leenhardt L, Bernier MO, Boin-Pineau MH, *et al.* Advances in diagnostic practices affect thyroid cancer incidence in France. *Eur J Endocrinol* 2004;150:133–9.
- 20 Black WC, Welch HG. Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *N Engl J Med* 1993;328:1237–43.
- 21 Harach HR, Franssila KO, Wasenius VM. Occult papillary carcinoma of the thyroid. A 'normal' finding in Finland. A systematic autopsy study. *Cancer* 1985;56:531–8.
- 22 Albarqouni L, Arab-Zozani M, Abukmail E, *et al.* Overdiagnosis and overuse of diagnostic and screening tests in low-income and middle-income countries: a scoping review. *BMJ Glob Health* 2022;7:e008696.
- 23 Zhang L, Mosquera I, Lucas E, *et al.* CanScreen5, a global repository for breast, cervical and colorectal cancer screening programs. *Nat Med* 2023;29:1135–45.
- 24 Kregting LM, Olthof EMG, Breekveldt ECH, *et al.* Concurrent participation in breast, cervical, and colorectal cancer screening in the Netherlands. *Eur J Cancer* 2022;175:180–6.
- 25 Armaroli P, Riggi E, Basu P, *et al.* Performance indicators in breast cancer screening in the European Union: A comparison across countries of screen positivity and detection rates. *Int J Cancer* 2020;147:1855–63.
- 26 Weinstock MA, Lott JP, Wang Q, *et al.* Skin biopsy utilization and melanoma incidence among Medicare beneficiaries. *Br J Dermatol* 2017;176:949–54.
- 27 Cho YJ, Kim DY, Park E-C, *et al.* Thyroid fine-needle aspiration biopsy positively correlates with increased diagnosis of thyroid cancer in South Korean patients. *BMC Cancer* 2017;17:114.
- 28 Yoo S, Yoon E, Boo D, *et al.* Transforming Thyroid Cancer Diagnosis and Staging Information from Unstructured Reports to the Observational Medical Outcome Partnership Common Data Model. *Appl Clin Inform* 2022;13:521–31.
- 29 Ripping TM, Ten Haaf K, Verbeek ALM, *et al.* Quantifying Overdiagnosis in Cancer Screening: A Systematic Review to Evaluate the Methodology. *J Natl Cancer Inst* 2017;109:djx060.
- 30 Tsodikov A, Gulati R, Heijnsdijk EAM, *et al.* Reconciling the Effects of Screening on Prostate Cancer Mortality in the ERSPC and PLCO Trials. *Ann Intern Med* 2017;167:449–55.
- 31 Lindberg A, Talala K, Kujala P, *et al.* Bias-corrected estimates of effects of PSA screening decisions on the risk of prostate cancer diagnosis and death: Analysis of the Finnish randomized study of screening for prostate cancer. *Int J Cancer* 2019;145:632–8.
- 32 Nieminen P, Kallio M, Anttila A, *et al.* Organised vs. spontaneous Pap-smear screening for cervical cancer: A case-control study. *Int J Cancer* 1999;83:55–8.
- 33 Jr JED, Sprunt DH. Uterine Cancer Case Finding by Vaginal Cytology: Memphis and Shelby County, Tennessee. *Public Health Rep* 1955;70:341.
- 34 Feinleib M, Zelen M. Some Pitfalls in the Evaluation of Screening Programs. *Arch Environ Health Int J* 1969;19:412–5.
- 35 Bailar JC III, Smith EM. Progress against Cancer? *N Engl J Med* 1986;314:1226–32.
- 36 Bell KJL, Del Mar C, Wright G, *et al.* Prevalence of incidental prostate cancer: A systematic review of autopsy studies. *Int J Cancer* 2015;137:1749–57.
- 37 Bonney A, Togawa K, Ng M, *et al.* Prevalence of subclinical lung cancer detected at autopsy: a systematic review. *BMC Cancer* 2023;23:794.
- 38 Wang S, Yang Z, Sha F, *et al.* Prevalence of incidental colorectal cancer and polyps in autopsies of different populations: a systematic review with meta-regression analysis. *Eur J Epidemiol* 2023;38:939–55.
- 39 Furuya-Kanamori L, Bell KJL, Clark J, *et al.* Prevalence of Differentiated Thyroid Cancer in Autopsy Studies Over Six Decades: A Meta-Analysis. *J Clin Oncol* 2016;34:3672–9.
- 40 Thomas ET, Del Mar C, Glasziou P, *et al.* Prevalence of incidental breast cancer and precursor lesions in autopsy studies: a systematic review and meta-analysis. *BMC Cancer* 2017;17:808.
- 41 Mindrup SR, Pierre JS, Dahmouch L, *et al.* The prevalence of renal cell carcinoma diagnosed at autopsy. *BJU Int* 2005;95:31–3.
- 42 Rossi SH, Hsu R, Blick C, *et al.* Meta-analysis of the prevalence of renal cancer detected by abdominal ultrasonography. *Br J Surg* 2017;104:648–59.
- 43 van den Akker-van Marle ME, Reep-van den Bergh CMM, Boer R, *et al.* Breast cancer screening in Navarra: interpretation of a high detection rate at the first screening round and a low rate at the second round. *Int J Cancer* 1997;73:464–9.
- 44 de Koning HJ, van der Aalst CM, de Jong PA, *et al.* Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N Engl J Med* 2020;382:503–13.
- 45 Blessing K, McLaren KM. Histological regression in primary cutaneous melanoma: recognition, prevalence and significance. *Histopathology* 1992;20:315–22.
- 46 Radha G, Lopus M. The spontaneous remission of cancer: Current insights and therapeutic significance. *Transl Oncol* 2021;14:101166.
- 47 Santucci C, Carioli G, Bertuccio P, *et al.* Progress in cancer mortality, incidence, and survival: a global overview. *Eur J Cancer Prev* 2020;29:367–81.
- 48 Caswell-Jin JL, Sun LP, Munoz D, *et al.* Analysis of Breast Cancer Mortality in the US-1975 to 2019. *JAMA* 2024;331:233–41.
- 49 Lebouleux S, Tuttle RM, Pacini F, *et al.* Papillary thyroid microcarcinoma: time to shift from surgery to active surveillance? *Lancet Diabetes Endocrinol* 2016;4:933–42.
- 50 Baboudjian M, Breda A, Rajwa P, *et al.* Active Surveillance for Intermediate-risk Prostate Cancer: A Systematic Review, Meta-analysis, and Metaregression. *Eur Urol Oncol* 2022;5:617–27.
- 51 Tainio K, Athanasiou A, Tikkinen KAO, *et al.* Clinical course of untreated cervical intraepithelial neoplasia grade 2 under active surveillance: systematic review and meta-analysis. *BMJ* 2018;k499.
- 52 Glencer AC, Miller PN, Greenwood H, *et al.* Identifying Good Candidates for Active Surveillance of Ductal Carcinoma In Situ: Insights from a Large Neoadjuvant Endocrine Therapy Cohort. *Cancer Res Commun* 2022;2:1579–89.
- 53 Johnson SB, Park HS, Gross CP, *et al.* Use of Alternative Medicine for Cancer and Its Impact on Survival. *JNCI* 2018;110:121–4.
- 54 Plevritis SK, Munoz D, Kurian AW, *et al.* Association of Screening and Treatment With Breast Cancer Mortality by Molecular Subtype in US Women, 2000–2012. *JAMA* 2018;319:154–64.
- 55 Raudasoja AJ, Falkenbach P, Vernooij RWM, *et al.* Randomized controlled trials in de-implementation research: a systematic scoping review. *Implement Sci* 2022;17:65.
- 56 Walter SD, Hu J, Talala K, *et al.* Estimating the rate of overdiagnosis with prostate cancer screening: evidence from the Finnish component of the European Randomized Study of Screening for Prostate Cancer. *Cancer Causes Control* 2021;32:1299–313.