Screening for multiple cancers: aggregate measures are only a start

Thomas Callender[1,2], Katherine Payne[3], Nora Pashayan[4], Anne Mackie[1]

[1] Department of Health and Social Care, London, UK
[2] Division of Medicine, University College London, London, UK
[3] Manchester Centre for Health Economics, Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, UK
[4] Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Correspondence to:
Dr Tom Callender
Division of Medicine,
University College London,
WC1E 6JF
t.callender@ucl.ac.uk

Word count: 2,199

KEY MESSAGES

- A range of multi-cancer detection tests are being developed to simultaneously screen for multiple cancers.
- As the natural history of different cancers varies substantially, even were such tests to reduce all-cancer mortality, the use of a single screening strategy for multiple cancers will have different cancer-specific benefits and harms that are both clinically and policy relevant.
- To allow policymakers to effectively consider multi-cancer screening programmes, we think that research that provides robust evidence of cancer-specific benefits and harms will be required.

Tom Callender and colleagues argue that aggregate outcomes amongst groups of cancers are insufficient for the evaluation of multi-cancer detection tests for cancer screening and that research providing cancer-specific outcomes will be necessary.

Multi-cancer detection tests (MCDs) using blood-based biomarkers for the concurrent screening of multiple cancer types in asymptomatic individuals are currently being trialled in the UK and elsewhere [1]. This has led to talk of their adoption in early cancer population screening programmes.

Considerable attention has been paid to the potential benefits that MCDs could present by improving cancer survival, particularly that of rarer cancers, through early detection. Balanced against these benefits has been a focus on the possible harms from false positives; inaccurate predictions of, or indeed inability to, predict the cancer site; and cases where a cancer is detected by an MCD but is too small to be seen on imaging [2,3].

Most discussion has considered the overall – aggregate – benefits and harms of an MCD programme. However, cancers behave remarkably differently both between individuals and cancer types. Despite this, the benefits and harms accruing to each type of cancer with strategies that uses a single multi-cancer screening test has had relatively little emphasis.

In this analysis, we discuss how differences in disease natural history could affect one potential harm, overdiagnosis, when screening for multiple cancers simultaneously. We argue that a focus on aggregate, all-cancer, metrics to evaluate MCDs will obscure differences between cancers that are clinically and policy relevant. Building on this, we discuss the benefits and challenges of approaches that balance the translation of a promising MCD screening programme against the further research that may be required to inform robust policy.

Early cancer detection with multi-cancer tests

Cancer underdiagnosis remains a challenge – cancers cause 28% of all deaths from England, and 42% of premature deaths amongst those aged 40-70 years [4]. Approximately two-thirds of deaths from cancer in men and half of those in women are from cancers for which there are no screening programmes [4]. Because of the relationship between stage at diagnosis and survival, early detection through screening is a research and policy priority.

Developing screening tests and cost-effective screening programmes for single rarer cancers is challenging. As the number of screening programmes and ensuing tests

expands, the burden these programmes place on individuals increases, as does the likelihood of suffering a harm such as a false positive [5] (Panel 1). Correspondingly, the prospect of a single blood-based MCD screening test could be attractive and there is broad enthusiasm amongst the public for MCD screening, particularly amongst high-risk groups [6,7].

MCDs analyse patterns of tumour-derived biomarkers, most commonly in a blood sample, to detect a cancer signal. They may have uses for asymptomatic screening as well as symptomatic diagnosis and disease monitoring [8]. On presenting a positive signal, the MCD might be able to predict the cancer site, allowing for subsequent guided diagnostic imaging. Notably, this is not always possible. Managing situations where a positive signal occurs but the cancer site is unclear is an active area of research and, in some cases, whole-body imaging may be needed. Our focus here is on screening asymptomatic populations for multiple, including less common, cancers simultaneously.

Overdiagnosis in cancer screening

Overdiagnosis occurs when a cancer is detected through screening that would otherwise not have been diagnosed or have caused harm during an individual's lifetime [9]. This happens when a cancer would never have presented clinically – would have spontaneously regressed or is indolent – had we not looked for it, or where a cancer would have eventually presented clinically had the individual not died from another cause beforehand. It is a natural consequence of screening.

The precise numbers overdiagnosed depends on the details of a screening programme: when and how often individuals are screened, the stage-specific sensitivity of the combined screening and subsequent diagnostic cascade of tests, along with the natural history of a tumour prior to clinical presentation. Nevertheless, the age-specific proportion of screen-detected cancers overdiagnosed can be estimated theoretically from the mean sojourn time (MST) [10]. The MST is the average time that a cancer could be detected with a particular screening test before it would be diagnosable clinically (Appendix Figure 1).

The MST is unknown for most cancers [3]. However, where estimates are available, they indicate that there is substantial variation both within and between cancer types. The MST for lung cancer has been estimated at between one and six years, varying between stage and histology [11,12]. Equivalent figures for breast cancer range up to 6.5 years [13], whilst the MST for prostate cancer could be ten or more years [14].

In setting the time between screening rounds a cancer's estimated MST has an important role, supporting a balance between ensuring that curable cancers are detected by screening and not in the intervals between screening rounds, whilst minimising the harms and costs of screening too frequently. This is a reason why not all screening programmes have the same interval between screening episodes.

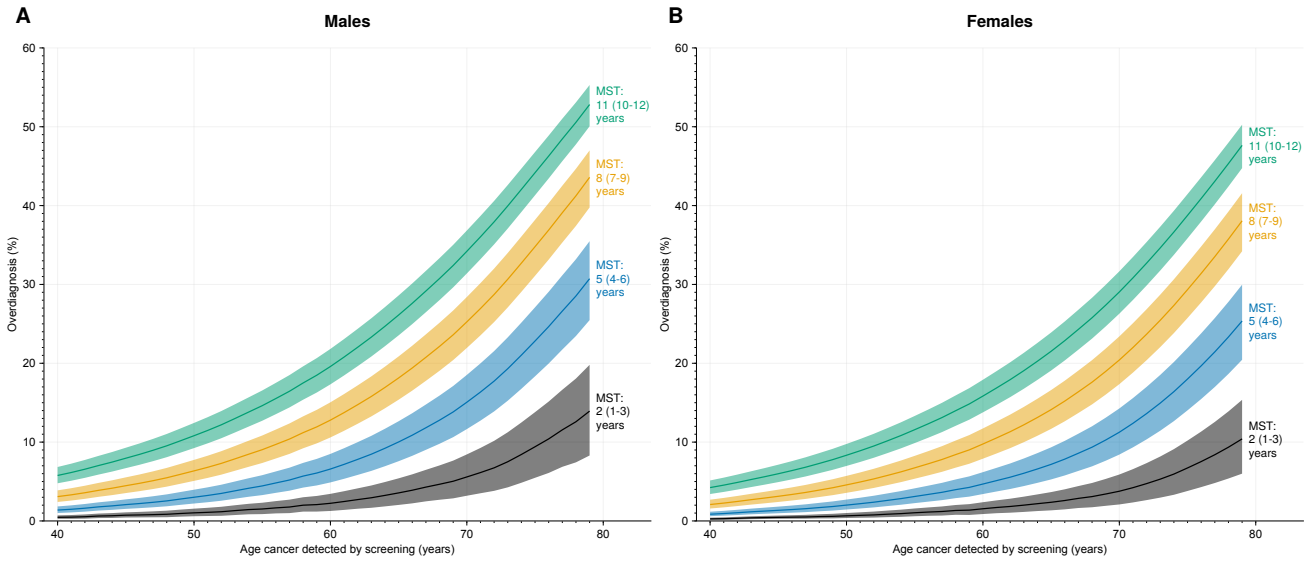

Figure 1: Age-specific percentage of screen-detected cancers overdiagnosed by sojourn time in men (A) and women (B). The mean sojourn time reflects the interplay of the characteristics of a specific screening pathway and the natural history of a particular tumour. Sex-specific differences reflect the different life expectancy of men and women. Estimates use the distribution of life-expectancies in the UK population from Office for National Statistics lifetables [4] and the lead time approach of Paci and colleagues [10,15]. Further details are available in the Appendix. Abbreviations: MST, mean sojourn time.

Relatively small differences in the MST translate into clinically relevant differences in the age-specific percentage of screen-detected cancers overdiagnosed (Figure 1). For a faster-growing cancer with an MST of three years, the percentage of screen-detected cancers overdiagnosed is small at age 50 (≤2%). However, at age 70, these values rise to 6% of women and 8% of men. With an MST of 10 years, between 7% and 9% of 50-year-olds and between one-quarter and one-third of 70-year-old women and men, respectively, are likely to have an overdiagnosed screen-detected cancer. Beyond the age of 70, overdiagnosis rises rapidly. At age 74, approximately 10% of screen-detected cancers would be overdiagnosed

with an MST of 3 years, reaching up to 40% of cancers with an MST of 10 years. The steep rise in overdiagnosis over the age of 70 reiterates the importance of upper age recommendations across all cancer screening programmes.

Overdiagnosis remains unknown for multi-cancer screening

Overdiagnosis may turn out to impact relatively few individuals using an MCD. This could be because the sensitivity of several MCDs for early stage (stage I-II) cancers has been relatively low in studies to date [16], although this could also imply that these MCDs are unlikely to be useful for screening. It has also been suggested that MCDs may preferentially detect more aggressive tumours [17], although this remains debated [18]. Perhaps the MCD is detecting cancers that already have micro-metastatic disease and so are wrongly being classified as early rather than advanced stage at diagnosis [19]. Nevertheless, these data are preliminary, and the field is rapidly developing.

Fundamentally, should an MCD have sufficient sensitivity for early-stage tumours whose natural history allows systematic screening to translate into mortality benefit, some overdiagnosis will occur. The proportion overdiagnosed could be relatively low overall. But by screening for multiple cancers simultaneously using a single test and screening interval, all of which have a different mean sojourn time, our examples highlight how aggregate outcomes from MCD studies will hide potentially important variation in overdiagnosis between cancers.

Balancing the benefits and harms of screening requires more than aggregate measures

Overdiagnosis is just one possible consequence of screening. Indeed, qualitative studies have highlighted patient concerns related to test accuracy and false positives in the context of MCDs [20]. Yet, overdiagnosis has a key role in determining the overall impact of screening. Many screening harms derive from the sequelae of managing cancer – the magnitude of these harms is directly affected by overdiagnosis. For example, 1 in 5 men who have their prostate removed after a prostate cancer diagnosis could suffer urinary incontinence [21]. The greater the cancer-specific overdiagnosis, the larger the proportion of harms suffered by individuals who will ultimately derive no benefit from their screening. Considered in conjunction with other harms from screening, such as the cancer-specific probability of false positives over time, an understanding of cancer-specific benefits and harms of any MCD will be necessary for policy making.

Understanding the cancer-specific benefits and harms of a multi-cancer screening programme

A wide range of MCD tests are currently under development for screening [8]. To understand their aggregate clinical effectiveness, adequately powered randomised controlled trials with long-term follow-up are necessary [22]. An initial such trial, NHS Galleri [1], is underway. In addition, we need a wider conversation around the desirability, methods, and infrastructure that would allow us to sustainably gather data at scale over sufficient timescales to enable cancer-specific assessment of harms and benefits. Pragmatic, registry-based, solutions have been put forward to balance the realistic assumption that MCDs may be used prior to definitive trial evidence in some jurisdictions [3]. Such registries will support the production of necessary real-world evidence to study the benefits, harms – including overdiagnosis – and the cost-effectiveness of these technologies. But, to paraphrase Collins and colleagues, there is a magic to randomisation [23].

Despite the interest in expanding screening, screening causes harm as well as having the potential to do good [24]. There is a balance to be struck between waiting for long-term cancer-specific benefit and harm outcomes and ensuring that as many people as possible can benefit from a potentially useful test. Yet, this need not come at the expense of randomisation, which has proved the most reliable mechanism by which relatively unbiased and authoritative results can be obtained [23,25]. Furthermore, evidence from prospective screening trials is necessary for modelling pre-clinical natural history that can be used subsequently to underpin simulation studies.

In the context of MCDs, should results from initial randomised trials such as NHS Galleri [1], or in future Vanguard [22], prove encouraging, research could be expanded as a randomised health service intervention [25] prior to a formal recommendation to screen. Any national programme would take time to develop and implement, so further comparative evaluations – such as a cluster randomised pilot programme – should be conducted in the interim and on an ongoing basis to guide policy [25]. To be implemented, we suggest this would require efficient linkage of trial data with health records and national registries. Gradually, this will allow us to reach trials of sufficient power to provide cancer-specific analysis of benefits and harms, whilst acknowledging that such a pilot programme may need to be stopped, or the cancers included altered. Such an approach will also allow us to study MCDs within complex, risk-tailored, screening programmes which may ultimately prove the most clinically- and cost-effective approaches to multi-cancer screening.

The potential benefits of an MCD for different cancers are thought to vary significantly [26]. In Appendix Table 1, we highlight crude, illustrative, estimates of the sample sizes that might be necessary under individual randomisation to reliably detect the 5-year mortality reductions thought possible with existing simulation studies. These range from approximately 100,000 in the case of colorectal cancer to >18 million for stomach cancer receiving an MCD (with an equivalent number of controls). This variation reflects the relative prevalence of the conditions, and the relative mortality benefits an MCD might produce. The exact numbers would depend on methodological advances, the test, trial design, approach to calculating sample size, cancer-specific effect size, time scale over which benefits are measured, and population in which the test was to be used. However, they serve to highlight that such an evaluation would need to be conducted at the scale of a live service and the fact that different approaches may be needed for different constituent cancers within an MCD.

Randomised interventions within live screening services have precedent, having been performed for both cervical and breast cancer in both Finland and the UK [25,27,28]. Learning from experience and from innovations in statistical methods [29] and trial designs [30], could support solutions to specific challenges – particularly regarding scale, consent and uptake, contamination of control arms, and evolving treatment paradigms – in generating randomised evidence on cancer-specific outcomes with MCDs. For instance, high-levels of uptake were coupled with individual patient consent in the TASTE trial through a randomised registry design [30]. This enabled the investigators to randomise 61% of all patients across Sweden and Iceland referred for percutaneous coronary intervention to manage their acute myocardial infarction for ~10% of the cost of similar trials [30]. By contrast, the AgeX trial used a cluster randomised approach to evaluate extending routine breast cancer screening across 5/6th of women in the UK [27]. As these women can request a mammogram beyond the end of the routine UK breast cancer screening programme, AgeX will provide experience in managing contamination of control arms and in disentangling the screening benefits over long follow-up periods during which treatments have improved. In the US, the Vanguard MCD trial [22] will support evaluation of innovations in statistical methods; specifically, the use of “Intended Effect” or Targeted approaches that reduce necessary sample sizes – of particular importance for cancer-specific outcomes – through selective testing of specimens from controls [29]. However, the requirement for baseline samples in both screening and control arms would need consideration at the point of study design when working at the scale of a live service.

Cancers are not one condition. To maintain the rigour with which we make policy on screening, we now need to consider the process, infrastructure, and funding approaches that would allow us to generate appropriate cancer-specific outcome data to evaluate these new screening technologies at scale and over the long term.

Panel 1: Screening tests for single or multiple cancers

In many countries, single cancer screening programmes are available for breast, cervical, colorectal, and lung cancer. Should multi-cancer tests prove promising, the future of cancer screening will likely include a mix of single and multi-cancer approaches; their advantages and disadvantages are highlighted in the following table.

	Advantages	**Disadvantages**
Single cancer tests	• May have greater sensitivity and thus more benefits. • Can be tailored to the natural history of the cancer, with more suitable starting and stopping ages and intervals, minimising the harms of screening, including overdiagnosis. • Have more potential for personalisation. For example, through risk-based eligibility or dynamic adjustment of screening intervals.	• Multiple screening programmes run concurrently increases the cumulative risk that an individual will suffer a harm from screening. • For rarer cancers, single cancer tests may not prove cost-effective or feasible. Single cancer tests are not available at present for many cancers.
Multi-cancer detection tests	• Could enable a large expansion of screening, particularly for rarer cancers. • Using a single test with a fixed specificity will lead to fewer false positives than multiple tests. • Likely to facilitate the delivery of screening and may consequently improve screening uptake overall and in underserved communities. • Could be used as an adjunct to existing screening programmes. • MCDs may prove more cost-effective than multiple single cancer screening programmes for rarer cancers.	• The benefits and harms of screening will vary between cancers included. • Screening cannot be tailored to the natural history of different cancers. • Opportunities to personalise multi-cancer programmes are complex and may need to focus on the groups of cancers included in a particular test, rather than using risk factors specific to different cancer types. • The cancers included in different MCDs differ, making comparisons between MCDs complex.

In most situations, established prospective trial methodologies can be extended to comparisons of single versus multi-cancer tests. In situations where a single cancer screening programme is already in place, such a programme would reflect current practice and a control arm. By contrast, were an MCD programme in place and a new single cancer test developed, the MCD programme would constitute a control arm. In both situations, we suggest that evaluations will eventually need to be powered for single cancer outcomes. As with existing screening programme evaluations, there are several key considerations:

Screening objective(s)

When comparing a single test against an MCD, improving outcomes from a specific cancer is the objective. If comparing two MCDs, overall and cancer-specific outcomes will be needed, particularly when the MCDs screen for different cancers.

Population included

The populations targeted by a single test and one or more MCDs may differ; this will need to be considered in any evaluation.

Clinical- and cost-effectiveness

The benefits, harms, and cost-effectiveness of different screening approaches requires comparison.

Feasibility

The clinical, social, and ethical impacts of different screening tests require evaluation, along with the feasibility of implementing a screening strategy.

Contributors

TC is a public health registrar in the UK National Screening Committee secretariat and a senior clinical research fellow at University College London. KP is professor of health economics at the University of Manchester, Deputy Director of the Division of Population Health, Health Services Research and Primary Care, and a member of the UK National Screening Committee's Research and Methodology Group. NP is professor of the epidemiology of ageing at the University of Cambridge and an honorary consultant in Public Health. AM is Director of Screening in the UK Department of Health and Social Care. All authors are involved in research and/or national policy related to cancer screening. TC acts as article guarantor.

Acknowledgements

We are grateful to members of the UK National Screening Committee and its subcommittees, as well as Professors David Baldwin and Matthew Callister, for their valuable comments on earlier drafts of this paper.

Report of patient involvement

Patients/public members were not specifically involved in this work.

Conflicts of Interest

TC is an NHS employee seconded to the UK National Screening Committee Secretariat in the UK Department of Health and Social Care. TC is partially supported by an EU Horizon 2020 Grant (848294). KP is a member of the research and methodology group of the UK National Screening Committee. NP is supported in part from NIHR Grant 206132 - Policy Research Unit on Cancer Awareness, Early Detection and Screening. AM is employed by the UK National Screening Committee, hosted by the UK Department of Health and Social Care. The opinions are those of the authors and not the UK National Screening Committee, Department of Health and Social Care, NIHR, or NHS, nor do they constitute official policy.

References

1 Neal RD, Johnson P, Clarke CA, *et al.* Cell-Free DNA-Based Multi-Cancer Early Detection Test in an Asymptomatic Screening Population (NHS-Galleri): Design of a Pragmatic, Prospective Randomised Controlled Trial. *Cancers* . 2022;14. doi: 10.3390/cancers14194818

2 Olivier T., Gill J., Prasad V. Multi-Cancer Screening Tests: Communicating About Risks Should be Prioritized. *Am J Med*. 2022;135:413–5. doi: 10.1016/j.amjmed.2021.09.012

3 Etzioni R., Gulati R., Weiss N.S. Multicancer Early Detection: Learning from the Past to Meet the Future. *J Natl Cancer Inst*. 2022;114:349–52. doi: 10.1093/jnci/djab168

4 Office for National Statistics. Mortality statistics - underlying cause, sex, and age 2017-2019. https://www.nomisweb.co.uk/ (accessed 12 January 2023)

5 Croswell JM, Kramer BS, Kreimer AR, *et al.* Cumulative incidence of false-positive results in repeated, multimodal cancer screening. *Ann Fam Med*. 2009;7:212–22. doi: 10.1370/afm.942

6 Schmeising-Barnes N, Waller J, Marlow LAV. Intention to have blood-based multi-cancer early detection (MCED) screening: a cross-sectional population-based survey in England. *Br J Cancer*. Published Online First: 27 August 2024. doi: 10.1038/s41416-024-02822-4

7 Adi-Wauran E, Clausen M, Shickh S, *et al.* 'I just wanted more': Hereditary cancer syndromes patients' perspectives on the utility of circulating tumour DNA testing for cancer screening. *Eur J Hum Genet*. 2024;32:176–81. doi: 10.1038/s41431-023-01473-y

8 Carr DJ, Welch HG. Assessing the Clinical Utility of Liquid Biopsies Across 5 Potential Indications From Therapy Selection to Population Screening: A Review. JAMA Intern. Med. 2023;183:1144–51.

9 Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst*. 2010;102:605–13. doi: 10.1093/jnci/djq099

10 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. doi: 10.1093/jnci/djx060

11 Wu D, Rai SN, Seow A. Estimation of Preclinical State Onset Age and Sojourn Time for Heavy Smokers in Lung Cancer. *Stat Interface*. 2022;15:349–58. doi: 10.4310/21-sii696

12 Ten Haaf K, van Rosmalen J, de Koning HJ. Lung Cancer Detectability by Test, Histology, Stage, and Gender: Estimates from the NLST and the PLCO TrialsLung Cancer Detectability by Test/Histology/Stage/Gender. *Cancer Epidemiol Biomarkers Prev*. 2015;24:154–61.

13 Lange J, Zhao Y, Gogebakan KC, *et al.* Test sensitivity in a prospective cancer screening program: A critique of a common proxy measure. *Stat Methods Med Res*. 2023;32:1053–63. doi: 10.1177/09622802221142529

14 Pashayan N, Duffy SW, Pharoah P, *et al.* Mean sojourn time, overdiagnosis, and reduction in advanced stage prostate cancer due to screening with PSA: implications of

sojourn time on screening. *Br J Cancer*. 2009;100:1198–204. doi: 10.1038/sj.bjc.6604973

15 Paci E, Warwick J, Falini P, *et al.* Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? *J Med Screen*. 2004;11:23–7. doi: 10.1177/096914130301100106

16 Leevan E., Pinsky P. Predictive Performance of Cell-Free Nucleic Acid-Based Multi-Cancer Early Detection Tests: A Systematic Review. *Clin Chem*. 2024;70:90–101. doi: 10.1093/clinchem/hvad134

17 Chen X, Dong Z, Hubbell E, *et al.* Prognostic Significance of Blood-Based Multi-cancer Detection in Plasma Cell-Free DNA. *Clin Cancer Res*. 2021;27:4221–9. doi: 10.1158/1078-0432.CCR-21-0417

18 Avanzini S, Kurtz DM, Chabon JJ, *et al.* A mathematical model of ctDNA shedding predicts tumor detection size. *Sci Adv*. 2020;6:2020.02.12.946228. doi: 10.1126/sciadv.abc4308

19 Callister M.E.J., Crosbie E.J., Crosbie P.A.J., *et al.* Evaluating multi-cancer early detection tests: an argument for the outcome of recurrence-updated stage. *Br J Cancer*. 2023;129:1209–11. doi: 10.1038/s41416-023-02434-4

20 Schmeising-Barnes N, Waller J, Marlow LAV. Attitudes to multi-cancer early detection (MCED) blood tests for population-based screening: A qualitative study in Great Britain. *Soc Sci Med*. 2024;347:116762. doi: 10.1016/j.socscimed.2024.116762

21 US Preventive Services Task Force, Grossman DC, Curry SJ, *et al.* Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2018;319:1901–13. doi: 10.1001/jama.2018.3710

22 Minasian LM, Pinsky P, Katki HA, *et al.* Study design considerations for trials to evaluate multicancer early detection assays for clinical utility. *J Natl Cancer Inst*. 2023;115:250–7. doi: 10.1093/jnci/djac218

23 Collins R, Bowman L, Landray M, *et al.* The Magic of Randomization versus the Myth of Real-World Evidence. *N Engl J Med*. 2020;382:674–8. doi: 10.1056/NEJMsb1901642

24 Gray JAM, Patnick J, Blanks RG. Maximising benefit and minimising harm of screening. *BMJ*. 2008;336:480–3. doi: 10.1136/bmj.39470.643218.94

25 Hakama M, Malila N, Dillner J. Randomised health services studies. *International Journal of Cancer*. 2012;131:2898–902. doi: 10.1002/ijc.27561

26 Sasieni P, Smittenaar R, Hubbell E, *et al.* Modelled mortality benefits of multi-cancer early detection screening in England. *Br J Cancer*. 2023;129:72–80. doi: 10.1038/s41416-023-02243-9

27 AgeX trial protocol. https://www.ceu.ox.ac.uk/research/agex-trial/trial-documentation/agex-trial-protocol (accessed 10 January 2025)

28 Lim AWW, Deats K, Gambell J, *et al.* Opportunistic offering of self-sampling to non-attenders within the English cervical screening programme: a pragmatic, multicentre, implementation feasibility trial with randomly allocated cluster intervention start dates (YouScreen). *EClinicalMedicine*. 2024;73:102672. doi: 10.1016/j.eclinm.2024.102672

29 Katki HA. Testing stored control-arm specimens could dramatically increase statistical power yet reduce costs in cancer screening trials. *J Natl Cancer Inst*. 2024;djae293. doi: 10.1093/jnci/djae293

30 James S, Rao SV, Granger CB. Registry-based randomized clinical trials—a new clinical trial paradigm. *Nat Rev Cardiol*. 2015;12:312–6. doi: 10.1038/nrcardio.2015.33