Screening for cervical, prostate and breast cancer: interpreting the evidence

Stacy M Carter, Jane Williams, Lisa Parker, Kristen Pickles, Gemma Jacklyn, Lucie Rychetnik, Alexandra Barratt (2015)

Cancer screening is well-established in high income countries, but its evidence base is constantly evolving and often contentious. This leaves physicians and policymakers in a difficult position, forced to act in the context of methodological complexity and substantive disagreement.[1,2] Three cases of screening for cancer or cancer risk are considered: cervical, prostate and breast screening. The unique characteristics of the disease, test and program in each case are outlined in Table 1. Tables 2-4, catalogue sources of controversy in each case; these are discussed in more depth below. The concluding section presents five common themes that may help explain the ongoing controversies.

The aim is not to synthesize the evidence, but to provide the 'backroom' story of the evidence on cancer screening, and so illuminate why experts so often disagree.

Cervical screening

Cervical screening is one of the best-supported and least controversial forms of cancer screening. Nonetheless, there are potentially contentious features of the cervical screening evidence base. These are:

1) Dependence on observational data;
2) Understanding, communicating and managing the balance of benefit and harm; and
3) The uncertain future impact of new technologies.

The first challenge in the cervical screening evidence base is the status of the existing evidence. Screening was established in parts of Europe and North America between the late 1940s and early 1960s and data from those programs, rather than from controlled trials, provide the evidence base for cervical screening effectiveness. Observational studies compared screened and unscreened populations and showed reduced cervical cancer incidence and mortality in the former.[3-5] This evidence base clearly shows that cervical

screening reduces morbidity and mortality: what is less clear is who to screen, when, and how to optimize benefit and minimize harm.

The cervical screening evidence base is susceptible to the well-known biases of any observational study.[1] It is not clear how these likely biases should be taken into account. In addition, the observational data about cervical screening crosses jurisdictions in which there are substantially different programs and reporting standards. This means that these observational data from different settings may not be as easily comparable as is often assumed (Table 1). To minimize bias, meta-analysis of Randomized Controlled Trial (RCT) evidence is the preferred method for estimating benefit and harm in screening. RCT evidence of different screening technologies, and combinations of technologies, is emerging. This may add more certainty to the cervical screening evidence base, although some of the findings from RCTs in low and middle income countries (LMICs) may not be transferable to other settings.[6-9]

Table 1: Disease, test and program characteristics in each case

	Cervical cancer	Prostate cancer	Breast cancer
Tests used	Pap smear using conventional &/or liquid based cytology +/- computer-assisted reading. HPV DNA testing increasing +/- cytology. Visual inspection with acetic acid/liquid iodine (VIA/VILI) in LMICs.	PSA test. New testing methods, including use of biomarkers, are being developed. DRE also used.	Mammogram. Fixed or mobile mammogram unit; recently widely upgraded to digital technology.
When test was invented	Pap test developed late 1930s.	First commercial PSA test released in 1986.	X-ray used for breast disease 1910s; 1st screening RCT 1963-75.
When test was first used for screening	Used to screen asymptomatic women from the 1940s.	USFDA approved PSA test for prostate cancer screening in 1994.	Ad hoc screening from mid 20th century;[60] population screening programs 1980s onwards (based on publication of results from early RCTs).
What test is designed to detect	Abnormal cells on the cervix (cytology, VIA/VILI) OR Presence of oncogenic HPV strains (HPV test).	Raised serum PSA levels.	Variations in soft tissue radiolucency. Originally diagnostic.
Relationship between test and target disease?	HPV-caused lesions are potential precursors for cervical cancer.	Poor. Test not developed to screen for cancer. Elevated PSA may not indicate cancer risk.	Cancers have characteristic (often subtle) soft tissue appearances on x-ray.
What results of screening are reported	Lesions: nature & severity (grade) of changes. Reporting standards differ. HPV reported by type.	Prostate Specific Antigen levels, expressed as nanograms of PSA per millilitre (ng/ml) of blood.	Apparent presence of masses and lesions suspicious for invasive and/or in situ cancer.
Contention over test itself	Cytology is prone to human error. Terminology & reporting standards vary. Sensitivity & specificity estimates vary widely.[61]	There is no meaningful 'normal range' for the PSA test in screening.	There is variation in what degree of suspicion constitutes a positive screen.

Variations between jurisdictions that may change the evidence base regarding benefit and/or harm	IARC recommends 3-yrly cytology screening from 25yrs. Evidence base pools data from widely varied programs:[5] start-age ranges from 18-30yrs, interval 1-5yrs; reporting standards, terminology & treatment vary.	Differences in target age, recommended finishing ages, screening intervals, thresholds definition of "abnormal", biopsy thresholds.	Differences in target age, screening intervals, thresholds for recall & biopsy. Service studies may differ in: participant population age (and therefore underlying cancer risk), follow-up, out-of-study screening.
Developments in the test	Tests that detect oncogenic-type HPV may supersede cytology as primary screening test.	New test rules in development. Variations proposed (free: total PSA ratio, PSA density, velocity, doubling time, prostate health index) for clinical significance. No evidence these improve health outcomes [2].	Increasing use of tomosynthesis (integrated 2/3D mammography) and MRI which may contribute to both benefits and harms.
Table Legend: DRE Digital Rectal Examination; HPV Human Papillomavirus; IARC International Agency for Research on Cancer; LMICs Low and Middle Income Countries; MRI Medical Resonance Imaging; PSA Prostate Specific Antigen; RCT Randomized Controlled Trial; USFDA US Food and Drug Administration; VIA visual inspection of the cervix using acetic acid to highlight precancerous lesions; VILI visual inspection of the cervix using Lugol's iodine to highlight precancerous lesions.			

The second challenge in this evidence base concerns understanding, communicating and managing the balance of benefit and harm: this problem has several dimensions. It is easy to inadvertently overstate the mortality benefit of cervical screening, particularly in high-income countries. This is because mortality from cervical cancer in high-income countries is considerably lower than for cancers such as breast and prostate. This was true even prior to widespread Pap-smear testing. For example the age-standardized mortality rate from cervical cancer in the UK was approximately 8/100,000 in 1971, compared to 37.5/100,000 for breast cancer and 20/100,000 for prostate cancer.[10] Thus, even substantial proportional (or relative risk) reductions in mortality attributed to screening may represent only small reductions in the absolute number of deaths prevented in well-resourced countries (Table 2). Cervical cancer, however, remains a significant burden and leading cause of cancer mortality in some low-income regions.[11]

Table 2: Main issues in cervical cancer screening

Issue	Explanation
Incidence and mortality of cervical cancer is low in high income countries	The incidence of cervical cancer is much lower than e.g. breast or prostate cancer, so number needed to screen over many years to avoid one death is high.[12]
Cervical screening reduces morbidity and mortality from cervical cancer.	Early Nordic observational studies suggest a mortality benefit from screening using the Pap test. Organised programs confer greater benefit than opportunistic screening.[5]
There is no RCT evidence from high income countries	Because Pap test screening for cervical cancer was introduced so early, it was not possible or ethical to conduct an RCT of its effectiveness.
RCTs are being conducted in LMICs	These will be a useful evidence base for LMICs.
It is easy to overstate the benefits of cervical cancer screening because the	Because incidence is low, number needed to screen is high and absolute risk reduction low. Statements of benefit may obscure the relatively small absolute number of

underlying mortality rate is low	people affected. For example mortality is often said to have halved in the decade following commencement of organised screening in Australia: this is accurate, but the absolute change was from only 4/100,000 to 2/100,000 women.
Most cervical lesions regress	It has been recognised since the 1970s that most cervical lesions will not progress to cervical cancer.
It is not clear what proportion of lesions regress, or which lesions will regress	It may never become clear which lesions will regress or what proportion of them will regress. CIN3 progression to cancer has been estimated at 12%,[62] 20%[12] and 30%[63] in different studies.
Overtreatment is difficult to measure and to manage	The majority of treatment is overtreatment, but as it is not possible to identify which lesions will regress this may not be resolvable with the technology currently available. There are vastly more abnormal results than there are invasive cancers, especially in women <25. E.g. in Australia in 2010 the incidence of invasive cancer in women <25 was 1.5/100,000, but 40,000 out of the 250,000 screens in women <25 returned an abnormal result.[64] Perinatal morbidity in treated women is the main iatrogenic harm of concern.[15]
The evidence base is affected by differences in program design between countries	Evidence about cervical screening comes mostly from monitoring data from screening programs. However, different countries run their programs differently. They use different tests, screening ages and screening intervals. They classify and report on their programs using different terminology and standards. Then the data from these very different contexts are combined. This has implications for the evidence base.
Screening technology is changing	Due to HPV vaccination a move away from cytology seems likely; an alternative future might be mass HPV screening with cytological examination of those with positive HPV tests. It is unclear what the incremental benefits and costs of these new technologies over existing screening programs will be. This is a rapidly evolving part of the evidence base in cervical screening.

Table legend: CIN3 Cervical intra-epithelial neoplasia; HPV Human Papilloma Virus; LMICs Low and Middle Income Countries; RCT Randomized Controlled Trial.

In addition, the treatments triggered by screening may be unnecessary and harmful in some cases. Cervical screening reduces cancer incidence as well as mortality. This is because it detects cellular abnormalities on the cervix, or pre-cancerous lesions, caused by human papillomavirus (HPV—Table 1). Cervical cancer is a rare outcome of persistent infection over a long time. However cellular abnormalities are common: there is an estimated lifetime incidence of 40% in women born since 1960.[12] Also, progression appears to be less linear than originally thought,[13] and most HPV infections regress spontaneously. This means 4/5 women with dysplasia may be treated unnecessarily;[12] but at present it is not possible to identify which individual high-grade lesions will regress (and so can be left untreated) or will progress (so require treatment–Table 2).

The evidence does suggest a solution however: to focus on minimizing harm, particularly in women under age 25. The evidence shows that: a) HPV infection is most likely to spontaneously regress in this group; b) paradoxically, these women also experience more abnormal cytology, treatment, and cervical incompetence and perinatal morbidity as a result of treatment; and crucially c) there is no mortality benefit in screening this age group.[14, 15] As a result, many countries are delaying commencement of screening until the age of 25 (Table 1) and/or recommending screening thereafter only every 3-5 years.[16, 17]

Although this change is supported by the evidence, in many jurisdictions women continue to be screened earlier and more often than these guidelines would support.[5, 18, 19]

Finally, it is important to anticipate the future impact of new technologies on the evidence base and on practice.[20, 21] Research increasingly supports screening women aged ≥30 using an oncogenic-type HPV test instead of or in addition to cytology.[6] The US Preventive Services Taskforce (USPSTF), for example, now recommends that women aged 30 to 65 years can screen with a combination of cytology and HPV testing every 5 years if they wish, rather than with cytology alone every 3 years.[7] The FDA has recently approved the use of HPV testing alone as a primary screening test,[22] which seems likely to result in further revision of recommendations. The recommendations are somewhat ahead of the evidence – with the exception of an Indian cluster RCT,[8] primary HPV testing has not yet shown mortality benefit. Similarly, comparative benefits and harms of different sequential combinations of HPV and cytology testing are not yet clear. However RCTs of newer screening technologies (e.g. HPV tests, including self-testing and testing in vaccinated populations, and computer assisted cytology reading) are underway. HPV vaccination will further reduce underlying risk in the population and so potentially reduce the relevance of the existing evidence on cervical screening.

Screening for prostate cancer

Unlike cervical screening, prostate specific antigen (PSA) testing for prostate cancer risk is intensely contested;[23] this includes contention over the relationship between evidence and practice. Important issues include:

1) There is inconsistency between the findings of different trials (and tension over the interpretation of observational findings);
2) Tests and thresholds for abnormality vary within and between studies; and
3) The evidence suggests that the PSA test performs poorly for screening purposes.

The first challenge is the quality and interpretation of research about the efficacy and effectiveness of PSA testing. Observational data from highly-screened communities is sometimes used to argue that testing reduces prostate cancer mortality.[24-26] However, as noted earlier, findings from observational studies may be misleading because of characteristic biases such as lead time, length time, and selection bias.[2, 27] Early RCTs were of poor quality (Table 3).[2, 27] Since then, two ongoing RCTs have reported results: the European Randomized Study of Screening for Prostate Cancer (ERSPC), and the USA Prostate Lung Colorectal and Ovarian Cancer (PLCO) trial. PLCO has shown no effect on prostate cancer-specific or all-cause mortality.[28] ERSPC reported reduced prostate cancer mortality in screened men but no change in all-cause mortality.[29] There is considerable controversy over trial design (Table 3). Although difficult to quantify, frequency of testing and follow-up, and type of treatment provided after diagnosis, are likely to affect outcomes reported from trials.[30-32]

Table 3 Main issues in prostate cancer screening

Issue	Explanation
Most prostate cancer is not life threatening	Although prostate cancer can be life threatening, the vast majority of cases are indolent.
Early trials of PSA screening were of poor quality	Early trials—which reported very positive findings—had serious methodological problems, including low participation in screening, failure to randomise, and failure to analyse by intention to screen.
Large RCTs are currently underway	ERSPC trial, and the USA PLCO trial have made interim reports but are ongoing. These are the only large, methodologically sound trials of PSA screening conducted to date.
There is controversy over the design of the current large RCTs	ERSPC included different countries using different screening tests and procedures. Those screened in the trial were more likely to be treated in a University hospital. The Swedish subset of ERSPC compared volunteer screenees (probably a healthier group) to whole-population controls (particularly significant because Sweden was one of only two, out of seven, subgroups to report statistically significant reductions in prostate cancer mortality after 11 years. These patterns likely to bias results in favour of screening. In PLCO, >50% of controls were screened during the trial, 44% of participants had previously been screened. Methodologists disagree on whether these biases are fatal to the results of the trials.
PSA screening may decrease prostate cancer death	Some trials suggest reductions in incidence of prostate cancer death. Observational studies in highly-screened populations suggest lower prostate cancer mortality.
PSA screening is unlikely to decrease all-cause mortality	Only ERSPC has reported a mortality benefit, which was very small in absolute terms. 1055 men would have to be screened to prevent one death from prostate cancer over 11 years.[30]
The PSA test is not prostate cancer-specific	PSA test has poor sensitivity and specificity for detecting prostate cancer. A PSA >4.0ng/ml produces a 6.2% false positive rate but detects only 20.5% of cancer cases.[65] PSA test cannot distinguish increased cancer risk from other common conditions e.g. benign prostatic hyperplasia, prostatitis. Certain medications (e.g. finasteride), ejaculation, and prostate manipulation can also increase PSA levels.
PSA test manufacturers and PSA thresholds vary between studies, laboratories, and clinicians	Studies and laboratories employ more than one kind of PSA test and different abnormal thresholds. The evidence base is thus hard to interpret due to lack of comparability. Conventional threshold for further investigation is 4ng/mL, but men with PSA levels 4-10ng/ml may not have prostate cancer,[66] and men with results <4ng/mL can show histological evidence of prostate cancer.[25, 67] Lowering the threshold below 4ng/mL would increase overdiagnosis and overtreatment of clinically unimportant disease.[27, 68] A meaningful threshold for screening may not exist because of the test's poor sensitivity and specificity i.e. the PSA test has little utility as a screening tool for prostate cancer. There is currently no alternative test available.
PSA screening can increase the likelihood of receiving treatment	In the USA, for example, up to 90% of men with prostate cancer diagnosed as a result of PSA testing receive treatment.[56]
Prostate cancer treatment can produce considerable negative consequences	Treatment can result in erectile dysfunction or impotence, anxiety, urinary incontinence, bowel dysfunction, or death.
Table Legend: ERSPC European Randomized Study of Screening for Prostate Cancer; PLCO Prostate Lung Colorectal and Ovarian Cancer trial; PSA Prostate Specific Antigen.	

Expert bodies increasingly advise against PSA screening. The USPSTF concluded that the mortality benefit is very small and outweighed by risk of harm.[33] The American College of Preventive Medicine has similarly concluded that populations should not be routinely screened with the PSA test, due to insufficient evidence.[34] The Australian National Health and Medical Research Council evidence guideline on PSA testing in asymptomatic men has recently concluded that there is no effect of PSA testing on all-cause mortality, and that no conclusions can be drawn about prostate cancer mortality.[35] These decisions are consistent with the evidence, which suggests that PSA testing may reduce the short term risk of dying from prostate cancer by a very small amount, at the cost of a much greater risk of harm, including from false positive results, overdiagnosis and overtreatment. The question this raises is: if a screened man will not die any later than an unscreened man, is it meaningful to prevent him from dying of prostate cancer in particular? And at what cost (harms to the man as well as expense to the man and the health system) should this goal be pursued? This question seems to divide experts, not least according to whether they care for men with the disease or have experienced it themselves.

The second problem in the PSA testing evidence is interpretability and comparability of PSA results. This is an issue for many screening tests (see Table 1), but especially for the PSA test. Manufacturers and laboratories employ divergent PSA calibrations, producing different PSA readings from the same sample.[36] Even when identical methods are used, thresholds set to separate 'normal' from 'high risk' PSA levels often differ. Within and between studies different standards are often combined, potentially invalidating conclusions.[25, 37] Tests and thresholds used by different countries participating in large trials often vary (Table 3), and trial study groups have been unable to identify acceptable PSA cut-off points for prostate cancer screening. This makes it difficult to compare study results and apply them to real-life settings.

The final problem with interpreting the evidence about PSA testing is addressing the potential for harm. The evidence suggests that sensitivity and specificity of the test are poor (Tables 1 & 3), which means cancers are missed (poor sensitivity) and false positives are common (poor specificity). The evidence suggests that PSA testing increases diagnosis of indolent disease, frequently cascades to diagnostic biopsies and follow-up treatments, and produces physical and psychological harms and costs: for every life saved by the PSA test, up to 48 men may be overtreated (Table 3).[38] Determining whether this is acceptable requires difficult debate over the nature of a good outcome, and what harm or expense that outcome might justify.

Screening for breast cancer
Like the evidence for PSA testing, the evidence for breast screening has been controversial. Important features of this evidence base include:

1) Uncertainty regarding the extent of breast cancer mortality reduction benefit;
2) Uncertainty regarding the extent of harm; and
3) Disagreement about managing in situ disease.

The first challenge for the evidence on breast screening is that despite a considerable body of research, the degree to which breast screening reduces breast cancer mortality remains unclear. The evidence base includes 11 RCTs (1971-2006), numerous observational studies, and mathematical models. It is probable that an invitational program of breast screening by mammography offers a population breast cancer mortality benefit, particularly for women aged 50-70. If poorer-quality RCTs are removed from meta-analyses this benefit is reduced. By how much is unclear (Table 4). Absolute and relative benefits are lower in women aged <50.[39] Also, treatment has greatly improved in recent decades, so including RCTs from the 1970s–1990s may overstate the benefit of screening (Table 4).[1, 40-44] The degree to which widely-observed declines in breast cancer mortality are attributable to improvements in treatment remains contested.[45] It is unclear how this can be resolved. Incremental changes in technology—from film mammography to digital mammography, tomosynthesis (integrated 2/3D mammography) and MRI to screen high risk women—may also affect the balance of screening benefits and harms.[46, 47]

Table 4 Main issues in breast cancer screening

Issue	Explanation
Mortality benefit exists	Most studies show mortality benefit from organised mammographic screening, especially for women aged 50-70, of approximately 20%.[40-43]
The extent of mortality benefit is contentious	Estimates of benefit vary considerably. Different study types are used including RCTs, observational studies and modelling. Meta-analysis of RCTs is widely regarded as the best way to identify population benefits, but different meta-analyses include or exclude different RCTs due to differing judgements about study quality.[40-43]
Mortality benefit is less than originally thought	Recent meta-analyses of RCTs suggest that benefit is lower than suggested by the earliest studies. This can be partly attributed to problems in quality with some of the RCTs. It has been hypothesised that treatment improvements in recent decades may leave less room for screening to have an effect and make older trial data less relevant.[40-43]
The harm from false positive screening tests varies between programs and populations	The rate of false positives varies as a result of factors such as: • Test factors e.g. equipment quality; skill of the clinicians reading the mammograms. • Differing policies and standards regarding acceptable levels of false positives and false negatives. • Frequency of screening in the program (increased frequency tends to increase the absolute number of false positives). • Individual participant factors (e.g. greater breast density in some women, including pre-menopausal women and women taking hormone replacement therapy (HRT)) which can make mammogram interpretation more challenging (and false positives more common). • Population factors: the frequency of false positives in part depends on the positive predictive value of the test, which depends on the prevalence of disease in the screened population. This depends on population risk profile (e.g. younger women have lower incidence).[48]
The extent of overdiagnosis is contentious	Estimates of overdiagnosis vary as a result of factors including the population studied, research questions asked (e.g. total cancer or invasive cancer only), methods used (e.g. comparing incidence in intervention and control arms of RCTs, comparing observational annual incidence data, comparing observational cumulative incidence data, using

	simulated population models), correction for possible biases such as lead time, and fundamental assumptions when estimating overdiagnosis in models. [1, 15, 44, 50]
Biological consequences of in situ disease is unclear	Before the onset of screening, in situ disease was mostly diagnosed in conjunction with an invasive cancer. It was not anticipated to be a common isolated finding on screening. It is unclear what the right response to increased diagnosis of in situ disease should be. Knowledge of the natural history of in situ breast diseases is improving but still incomplete. Diagnosis and management are controversial, especially for less aggressive diseases (e.g. low grade DCIS) where risk of death is only slightly increased but surgery to negate the risk may be extensive. [40-43]
There are small radiation harms of screening	Harms from radiation during mammography are generally agreed to be real, and may be greater in women screened more often (e.g. those identified as carrying potentially harmful mutations in the *BRCA1* or *BRCA2* genes). [69, 70] However in screening of the general population, these risks are extremely small, and likely to be further reduced by the implementation of digital mammography.
BRCA1/BRCA2 BReast CAncer susceptibility gene 1 and 2; DCIS Ductal Carcinoma In Situ; RCT Randomised Controlled Trial.	

The second concern is the extent of harm caused. Invitational mammography programs cause harm, including from false positives and overdiagnosis. The absolute rate of false positives can vary according to equipment used, skill and experience of film readers, test thresholds and screening frequency (Table 4).[48] Although the rate of false positives per screen may be low, they accumulate, so the chance of false positive recall or biopsy over a lifetime is much higher. Increasingly, evidence suggests that breast screening produces overdiagnosis of both invasive and in situ breast cancer. Although experts agree that mammography screening causes overdiagnosis, there is disagreement on its extent. A recent meta-analysis suggests that, in women invited to screening, there is an 11% lifetime risk of overdiagnosis as a proportion of cancers diagnosed, and a 19% risk during the active screening period.[40, 43] Harms, especially overdiagnosis, may tend to outweigh benefits in women >70 as they age.[49] However the relevant evidence is highly contentious for methodological and other reasons explained in Table 4.[44, 50]

The final challenge in this evidence base concerns ductal carcinoma in situ (DCIS), which represents approximately 17% to 34% of screen-detected cases and 20%–25% of all newly diagnosed cases of breast cancer in the USA.[51] Women are rarely diagnosed with DCIS because they experience symptoms: DCIS is diagnosed almost entirely as a result of screening. Overdiagnosis of DCIS is considered by many an important harm of mammographic screening. However the evidence is not clear on either the natural history of DCIS or how aggressively DCIS should be treated. More research is needed to evaluate treatments for in situ disease.[40, 44]

What characteristics of the screening evidence base could explain expert disagreement?
In high-income countries, cancer screening is a familiar feature of preventive medical care. Screening is expected—with good reason—to be informed by evidence. Across these three cases, there are two less-often discussed tensions and three more explicit tensions that help to explain why interpreting the evidence is such a difficult task.

Tensions in the evidence base that are less-often discussed

Two tensions in the evidence base are under-examined: the comparability of data between studies and contexts, and the impact of technological developments. These tensions are also difficult to resolve and potentially destabilizing.

Data from very different contexts may not be comparable

Data from different contexts may not be comparable, particularly for observational data from monitoring studies. As shown, the evidence base contains data from different times, countries and programs, and from populations with varying event rates (Table 1). Transferability of this evidence is difficult for several reasons. Because screening trials are particularly large and need long follow-up to show effects, they can be especially susceptible to the passage of time. When early trials were conducted, screening techniques were less developed, treatments less effective, cancer incidence lower and cancer mortality often higher. (Breast screening evidence, for example, includes decades-old trials: treatment has progressed substantially since they were conducted.) Evidence from screening trials is also susceptible to local variation (e.g. in disease biology, event rates and age distribution), not least because screening is applied to whole populations, not just people who are ill. (As HPV vaccination is implemented differently around the world, for example, the underlying event rate for cervical cancer will change dramatically.) The resource intensiveness of cancer screening trials also means that: 1) few trials are done (leaving less evidence to interpret); 2) trials are often funded by industry (changing the research questions asked); and 3) trials are somewhat dependent on local screening and treatment practices (e.g. target age, screening intervals, testing techniques, follow-up time, available treatment). The variability and transferability of screening evidence is a challenge for methodologists; even more so for clinicians and policymakers, as the characteristics on which the evidence depends are not always made clear in reporting.

Screening technologies affect evidence quality, evidence must evolve with changing technologies

Cancer screening relies on complex cascades of technology for collecting, imaging, analysing and interpreting possible changes in human bodies. Without the technology, there is no screening, but as technology evolves, it potentially makes existing evidence obsolete.[52]

The evidence on PSA is hampered by poor technology. The PSA test has limited sensitivity and specificity, studies and laboratories use multiple test types and different thresholds, there is no meaningful 'normal range', and new test rules do not appear to change patient outcomes. Some propose using test results only within, rather than between, patients, but the poor test characteristics of PSA make even this problematic. It is understandable that clinicians want to retain some tool to measure prostate cancer risk.[53] But given the test

characteristics of the PSA, it may not be possible to generate a meaningful evidence base about its use in populations.

The cervical screening evidence base is shifting because of changing technology; tests that detect oncogenic-type HPV may become the primary form of screening in vaccinated populations. Mammography remained relatively constant in the 20th Century, changing only incrementally from film to digital mammography. In the 21st Century we face substantial technological change, with moves to tomosynthesis (integrated 2/3D mammography) and MRI screening of high risk women. Although tomosynthesis is receiving considerable attention in the lay press and peer reviewed literature, attempts to estimate its effects have been based on opaque assumptions and limited evidence. It seems possible that both MRI and tomosynthesis will enhance both the benefits and harms of screening, but at present this is unknown.[46, 47]

Acknowledged tensions in the evidence base
Three other, more explicit, tensions are over the quality of evidence of benefit, the relatively new evidence regarding screening harms, and risk communication.

The quality of evidence of benefit varies, and the implications are contested
When one expert says to another 'you are wrong about the evidence on screening', she is likely to mean this: 'I disagree with the criteria that you have used to separate good-quality studies, which should be included, from poor-quality studies, which should be excluded. I therefore disagree with your conclusion.'
The cancer screening evidence base contains observational studies, RCTs, and modelling, of widely varying quality, and with disparate results. Early studies of screening generally suggested greater benefit, and later studies less benefit, which may be because early trials were poorly designed (e.g. PSA) or because recent treatment improvements leave less room for screening to provide benefit (e.g. breast screening). Even new trials contain methodological flaws (e.g. PLCO, ERSPC), and methodologists often disagree about study design, particularly over whether screened and unscreened groups are comparable.

New RCTs are expensive and logistically challenging, so are rare. Thus new conclusions generally arise from reanalyses of existing research findings rather than from new trials. Researchers performing meta-analyses must decide on criteria for including and excluding studies. The recent Marmot review of the evidence on breast screening demonstrates that this is possible,[40] even in high-profile situations, but disagreement over criteria is likely to remain. And when new analyses produce new findings, those whose settled beliefs are challenged may perceive the chosen criteria as arbitrary or incorrect. This highlights the importance of transparency regarding how and why meta-analyses are conducted.

Evidence about harms is relatively new, there are gaps in that evidence, and disagreement about what it means

Initially, cancer screening researchers focused on measuring screening benefits; they have only recently turned to potential harms. For all three cases—cervical, prostate and breast cancer (including DCIS)—there is limited evidence about which instances of disease or pre-disease are aggressive, so require treatment, and which will be indolent or regress. Because of this, many people will be over-treated, and may be harmed. Researchers are trying to address this gap, by studying the mortality benefit of treatment for small, grade 1, node negative breast cancers, for example, or the genetic profile of aggressive versus indolent prostate cancers. This work may assist in future. In the meantime, existing knowledge suggests opportunities to reduce harm. For example, there is currently no way to determine which cervical lesions will regress or progress. However epidemiological data demonstrates that women 18-25 are most likely to have unnecessary treatment, experience harms from treatment, and fail to benefit from treatment. This has led some jurisdictions to restrict cervical screening to women aged over 25.

Even when evidence about screening harms emerges, experts often disagree about what it means and how to respond. This may be in part because public health and medical professionals have learned to think in a particular way, and have taught citizens to think similarly, of cancer and pre-cancer as progressive and life-threatening, and screening as one of few defences against this threat. For the first several decades of screening research, harms were rarely measured. Although later research suggested that screening may harm, it may be difficult for this evidence to reach public attention given the powerful cultural meaning of cancer death.[54, 55] New facts about screening harms are hotly contested, with regard both to their accuracy and their implications. And screening programs continue to be evaluated primarily against increasing participation targets, rather on the likely balance between benefit and harm achieved.

For example, it is generally accepted that prostate biopsies and prostate cancer treatments are likely to produce harms. This is taken as a fact, but that fact is interpreted very differently. Some argue that most screen-detected prostate cancers are indolent, so most diagnosis is overdiagnosis, and most harm done is unnecessary harm. They conclude that insurers or policymakers should constrain clinicians who test healthy men, thus preventing harm. Others take a different view, that without PSA testing clinicians have no way of diagnosing tumours that would develop or metastasize. These experts tend to take the view that insurers or policymakers should leave testing open to clinicians, and allow the possibility of harm to be dealt with via more judicious decisions about treatment. Their opponents might counter with studies showing that men diagnosed with prostate cancer generally proceed to treatment rather than 'watching and waiting'.[56] Although each party can present data of some kind to support their claims, it is worth remembering that data become evidence only through interpretation, and that experts are susceptible to biases in this interpretive process.[57, 58]

Evidence about outcomes is often poorly communicated, despite the evidence about communication

Researchers and programs tend to express outcomes using relative risks, which incorporate baseline risk and so are easier to generalize across contexts. However research shows that relative risks encourage lay people and clinicians to overemphasize benefits and minimize harms. This has been acknowledged as ethically problematic, potentially biasing or manipulating people's perceptions, misleading them, and undermining their autonomy.[59] If experts are obliged to communicate honestly with citizens—an obligation that seems supportable—this becomes an urgent issue to address for all forms of cancer screening.

Conclusions

The benefits and harms of screening are often finely balanced; more than anticipated when screening was established. There are both unique and shared characteristics of cervical, prostate and breast screening that help to explain the challenge of balancing benefit and harm. These include the incomparability of data from different times, places and programs, the instability of the very technology on which screening is based, disagreement on which studies are well-enough designed to be taken seriously, gaps in knowledge, and disagreement about how to understand newly emerging evidence of harm. This suggests five principles for evaluating and using the evidence:

1) attend closely to transferability;
2) consider the influence of technologies on the evidence base;
3) query the design of meta-analyses;
4) ensure harms are defined and measured; and
5) improve risk communication practices.

However even more fundamental are questions about the purpose of screening, and who should make decisions about screening. Should insurers or policymakers leave screening options open for clinicians and patients to choose? Or should they be directive, promoting some forms of screening and limiting others to minimize harm? Should community engagement and deliberation guide screening policy and practice? And what should the purpose of screening be? There are many potential aims of cancer screening, including preventing cancer death, reducing all-cause mortality, minimizing anxiety, maximizing cost efficiency and/or minimizing avoidable harm. These different aims reflect different values, values that may differ between patients, clinicians, funders and policymakers. Questions about the evidence base need resolution. This should be complemented with clear thinking about the aims of screening. Only when the aims of screening are clear will researchers be able to generate an evidence base sufficient to assist decision-making, and clinicians be able to best support their patients to make good screening decisions.

References

1.   Bretthauer M, Kalager M. Principles, effectiveness and caveats in screening for cancer. Br J Surg. 2013;100(1):55-65. Epub 2012/12/06.

2.   Moyer VA. Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2012;157(2):120-34.

3.   Hakama M, Chamberlain J, Day N, Miller A, Prorok P. Evaluation of Screening Programs for Gynecological Cancer. Br J Cancer. 1985;52:669-73.

4.   IARC. Screening For Squamous Cervical Cancer: Duration Of Low Risk After Negative Results Of Cervical Cytology And Its Implication For Screening Policies. BMJ. 1986;293(6548):659-64.

5.   IARC. Cervix Cancer Screening. International Agency for Research on Cancer, editor. Lyon, France: IARC Press; 2005.

6.   Ronco G, Dillner J, Elfström KM, Tunesi S, Snijders PJF, Arbyn M, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. The Lancet. 2014;383(9916):524-32.

7.   Moyer VA, US Preventive Services Task Force. Screening for Cervical Cancer: US Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2012;156(12):880-91.

8.   Sankaranarayanan R, Nene BM, Shastri SS, Jayant K, Muwonge R, Budukh AM, et al. HPV Screening for Cervical Cancer in Rural India. N Engl J Med. 2009;360(14):1385-94.

9.   Vesco KK, Whitlock EP, Eder M, Lin J, Burda BU, Senger CA, et al. Screening for Cervical Cancer: A Systematic Evidence Review for the U.S. Preventive Services Task Force. Portland, Oregon: Oregon Evidence-based Practice Centre, 2011.

10.   Cancer Research UK. Cancer mortality statistics. [Updated 15 September 2014; cited 30 July 2013] Available from: http://www.cancerresearchuk.org/cancer-info/cancerstats/mortality/.

11.   Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010;127(12):2893-917.

12.   Raffle AE, Alden B, Quinn M, Babb PJ, Brett MT. Outcomes of screening to prevent cancer: analysis of cumulative incidence of cervical abnormality and modelling of cases and deaths prevented. BMJ. 2003;326(7395):901.

13.   Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. The Lancet. 2007;370(9590):890-907.

14.   Sasieni P, Castanon A, Cuzick J. Effectiveness of cervical screening with age: population based case-control study of prospectively recorded data. BMJ. 2009; 339: b2968.

15.   Kyrgiou M, Arbyn M, Martin-Hirsch P, Paraskevaidis E. Increased risk of preterm birth after treatment for CIN. BMJ 2012; 345: e5847.

16. Bastos J, Peleteiro B, Gouveia J, Coleman MP, Lunet N. The state of the art of cancer control in 30 European countries in 2008. Int J Cancer. 2010;126(11):2700-15.

17. Vesco KK, Whitlock EP, Eder MP, Burda BU, Senger CA, Lutz KM. Risk Factors and Other Epidemiologic Considerations for Cervical Cancer Screening: A Narrative Review for the U.S. Preventive Services Task Force. Ann Intern Med. 2011;155(10):698-705.

18. Corbelli J, Borrero S, Bonnema R, McNamara M, Kraemer K, Rubio D, et al. Differences among primary care physicians' adherence to 2009 ACOG guidelines for cervical cancer screening. J Womens Health. 2014;23(5):397-403.

19. Henderson JT, Yu JM, Harper CC, Sawaya GF. U.S. clinicians' perspectives on less frequent routine gynecologic examinations. Prev Med. 2014;62(0):49-53.

20. Meyerson BE, Crosby RA, Van Der Pol BJ, Zimet GD. Thinking differently about cervical cancer screening in high-risk populations. Am J Prev Med. 2012;43(2):221-4.

21. Perkins RB, Anderson BL, Sheinfeld Gorin S, Schulkin JA. Challenges in cervical cancer prevention: a survey of US obstetrician-gynecologists. Am J Prev Med. 2013;45(2):175-81.

22. Food and Drug Administration (FDA). FDA approves first human papillomavirus test for primary cervical cancer. 2014 [Updated 24 April 2014; cited 17 June 2014]; Available from: http://www.fda.gov/newsevents/newsroom/pressannouncements/ucm394773.htm.

23. American Academy of Family Physicians. AAFP, USPSTF Issue Final Recommendation Against Routine PSA-based Screening for Prostate Cancer. AAFP Breaking News 2012 [Updated 5 December, 2013]; Available from: http://www.aafp.org/news-now/health-of-the-public/20120522psascreenrec.html.

24. Wallner LP, Jacobsen SJ. Prostate-Specific Antigen and Prostate Cancer Mortality: A Systematic Review. Am J Prev Med. 2013;45(3):318-26.

25. Loeb S, Chan DW, Sokoll L, Kan D, Maggiore J, Mikolajczyk SD, et al. Prostate specific antigen assay standardization bias could affect clinical decision making. J Urol. 2008;180(5):1959-62.

26. Jemal A, Ward E, Wu X, Martin HJ, McLaughlin CC, Thun MJ. Geographic patterns of prostate cancer mortality and variations in access to medical care in the United States. Cancer Epidemiol Biomarkers Prev. 2005;14(3):590-5.

27. Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Cochrane Database Syst Rev. 2013 Jan 31;1:CD004720. doi: 10.1002/14651858.CD004720.pub3

28. Chapman S, Barratt A, Stockler M. Let sleeping dogs lie? What men should know before getting tested for prostate cancer. Sydney: Sydney University Press; 2010. Available from: http://ses.library.usyd.edu.au/bitstream/2123/6835/3/Let-sleeping-dogs-lie.pdf.

29. Roobol MJ, Kerkhof M, Schroder FH, Cuzick J, Sasieni P, Hakama M, et al. Prostate cancer mortality reduction by prostate-specific antigen-based screening adjusted for nonattendance and contamination in the European Randomised Study of Screening for Prostate Cancer (ERSPC). Eur Urol. 2009;56(4):584-91.

30. Schröder FH, Hugosson J, Roobol MJ, Tammela TLJ, Ciatto S, Nelen V, et al. Prostate-Cancer Mortality at 11 Years of Follow-up. N Engl J Med. 2012;366(11):981-90.

31. Andriole GL, Crawford ED, Grubb RL, 3rd, Buys SS, Chia D, Church TR, et al. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med. 2009;360(13):1310-9.

32. Chou R, Croswell JM, Dana T, Bougatsos C, Blazina I, Fu R, et al. Screening for Prostate Cancer: A Review of the Evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2011;155 (11):762-71.

33. US Preventive Services Task Force. Screening for Prostate Cancer. US Preventive Services Task Force; 2012 [cited 24 June 2014]; Available from: http://www.uspreventiveservicestaskforce.org/prostatecancerscreening/prostatefinalrs.htm .

34. Lim LS, Sherin K. Screening for prostate cancer in US men: ACPM position statement on preventive practice. Am J Prev Med. 2008;34(2):164-70.

35. NHMRC. Prostate-specific Antigen (PSA) testing in asymptomatic men. Canberra: NHMRC; 2014. Available from: www.nhmrc.gov.au/guidelines/publications/men4

36. Zhu X, Albertsen PC, Andriole GL, Roobol MJ, Schroder FH, Vickers AJ. Risk-based prostate cancer screening. Eur Urol. 2012;61:652-61.

37. Sokoll L, Chan D. Total, free, and complexed PSA: analysis and clinical utility. J Clin Ligand Assay. 1998;21:171-9.

38. Andriole G, Djavan B, Fleshner N, Schroder F. The case for prostate cancer screening with prostate-specific antigen. European Urology Supplements. 2006;5:737-45.

39. Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. The Lancet. 2006;368(9552):2053-60.

40. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. The Lancet. 2012; 380(9855):1778-1786.

41. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. Ann Intern Med. 2009;151(10):727-37.

42. IARC Handbooks of Cancer Prevention Volume 7: Breast Cancer Screening. IARC Press, Lyon, 2002.

43. Gøtzsche P, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2011 Jan 19;(1):CD001877. doi:10.1002/14651858.CD001877.pub4.

44. Autier P, Esserman L, Flowers C, Houssami N. Breast cancer screening: the questions answered. Nature Reviews Clinical Oncology. 2012;9(10):599-605.

45. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005;353(17):1784-92.

46. Lord S, Lei W, Craft P, Cawson J, Morris I, Walleser S, et al. A systematic review of the effectiveness of magnetic resonance imaging (MRI) as an addition to mammography and ultrasound in screening young women at high risk of breast cancer. Eur J Cancer. 2007;43(13):1905-17.

47. Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. The Lancet Oncology. 2013;14(7):583-9.

48. Elmore JG, Armstrong K, Lehman CD, Fletcher SW. Screening for breast cancer. JAMA. 2005;293(10):1245-56.

49. Royce TJ, Hendrix LH, Stokes WA, Allen IM, Chen RC. Cancer screening rates in individuals with different life expectancies. JAMA Internal Medicine. 2014;174(10):1558-65.

50. Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. Lancet Oncol. 2007;8(12):1129-38.

51. Kerlikowske K. Epidemiology of Ductal Carcinoma In Situ. JNCI Monographs. 2010;2010(41):139-41.

52.   Irwig L, Houssami N, Armstrong B, Glasziou P. Evaluating new screening tests for breast cancer. BMJ. 2006;332(7543):678-9.

53.   Han PKJ, Coates RJ, Uhler RJ, Breen N. Decision making in prostate-specific antigen screening National Health Interview Survey, 2000. Am J Prev Med. 2006;30(5):394-404.

54.   Hoffman RM, Elmore JG, Fairfield KM, Gerstein BS, Levin CA, Pignone MP. Lack of Shared Decision Making in Cancer Screening Discussions. Am J Prev Med. 2014;47(3):251-9.

55.   Rosenbaum L. Invisible Risks, Emotional Choices — Mammography and Medical Decision Making. N Engl J Med. 2014;371(16):1549-52.

56.   Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. J Natl Cancer Inst. 2009;101:1325-9.

57.   Finkel AM. Perceiving Others' Perceptions of Risk. Ann N Y Acad Sci. 2008;1128(1):121-37.

58.   Sjöberg L. The Allegedly Simple Structure of Experts' Risk Perception: An Urban Legend in Risk Research. Science, Technology & Human Values. 2002;27(4):443-59.

59.   Shaw D, Elger B. Evidence-based persuasion: an ethical imperative. JAMA. 2013;309(16):1689-90.

60.   Gershon-Cohen J, Berger S. Detection of breast cancer by periodic X-ray examinations: a five-year survey. JAMA. 1961;176(13):1114-6.

61.   Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap Test Accuracy. Am J Epidemiol. 1995;141(7):680-9.

62.   Ostor A. Natural History of Cervical Intraepithelial Neoplasia: A Critical Review. Int J Gynecol Pathol. 1993;12(2):186.

63.   McCredie MRE, Sharples KJ, Paul C, Baranyai J, Medley G, Jones RW, et al. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. The Lancet Oncology. 2008;9(5):425-34.

64.   Australian Institute of Health and Welfare. Cervical screening in Australia 2010-2011. Canberra: 2013.

65.   Saenger A. PSA Standardization. Mayo Foundation for Medical Education and Research; 2013 [updated May 23 2013]; Available from: http://www.mayomedicallaboratories.com/articles/hottopics/transcripts/2009/2009-10b-psa/10b-9.html.

66.   Harris R, Lohr KN. Screening for Prostate Cancer: An Update of the Evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2002;137(11):917-29.

67.   Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. N Engl J Med. 2004;350(22):2239-46.

68.   Welch HG, Schwartz LM, Woloshin S. Prostate-specific antigen levels in the United States: implications of various definitions for abnormal. J Natl Cancer Inst. 2005;97(15):1132-7.

69.   Berrington de Gonzalez A. Estimates of the potential risk of radiation-related cancer from screening in the UK. J Med Screen. 2011;18(4):163-4.

70.   Moyer VA. Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer in Women: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2014;160(4):271-81.