**University of Utah Internal Medicine Journal Club Facilitators Guide:**

Updated: Brian Locke 2/25/2020

**Preventive (e.g. screening) RCT:**

* Article Title
* Study question, design, and justification for screening: (Is this condition important? Is there early/asx disease to detect? Is treatment better when found early? Is this test potentially accurate enough to use?)
* Patients included: (Where and how were patients enrolled? Inclusion/Exclusion criteria? What is their baseline risk for the disease? How did they determine their sample size?)
* What was the workup and treatment in the intervention and control arms? (Was there a protocol for how to follow-up abnormal findings, or treat disease when found? If not, what actually occurred?)
* Outcomes:(Focus on the primary outcome: Patient centered, or surrogate outcome? Blinded assessment? Objective? Appropriate duration of follow-up? What secondary / adverse outcomes were assessed?)
* What are the Results? (including absolute effect size and precision in the estimate. What were the benefits, and what were the harms in each group? Loss to follow-up?)
* Critique: (Are there threats to the internal validity - such as bias or chance - or external validity aka generalizability?)
* Can I apply the results to my patient? How?

Justification:

“Do Physicians Understand Cancer Screening Statistics? A National Survey of Primary Care Physicians in the United States

…

Conclusion: Most primary care physicians mistakenly interpreted improved survival and increased detection with screening as evidence that screening saves lives.

Published: Ann Intern Med. 2012;156(5):340-349.”

Study design:

**What assumptions have to be made for a screening test to work?**

1. The condition must be an important health problem
2. There must be a recognizable latent or early symptomatic stage
3. There must be a benefit to early treatment (over late treatment)
4. There must be a test with high enough accuracy (sensitivity and specificity to detect disease that will ultimately progress to impact the patient) to be used in the target population (generally, low prevalence of the condition)

**Why are randomized trials particularly important in screening?**

In addition to balancing confounding (both measured and unmeasured) between groups, the unified assessment of time 0 combats length and lead time bias (below)

Patient selection:

**Why do screening trials have to be so large?**

* Low event rates = enrolling healthy volunteers
* Low absolute treatment effect (because it’s harder to make people who are already healthy more healthy)
* Long delay between time of screening and occurrence of clinically important outcomes (time to diagnosis is not relevant, so generally death is the only meaningful end-point

**Did the authors discuss how they decided who to include in the screening intervention**? Does this population have a high enough baseline risk to justify screening? Are there expected changes in incidence of this disease that might impact this in the future?

**Did the study reach it’s goals size?** If not, why not. Under-recruitment is particularly important if the study did not reject the null (ie. Negative study)

Outcomes:

**What benefits did the authors report?** If it is anything other than overall mortality, did they justify how their chosen end-point is expected to correlate with overall mortality (or, did they have some other patient-centered end-point, such as quality of life?)

**Is disease-specific mortality an acceptable surrogate for what we care about?**

Argument for: it is a more proximal end-point, thus takes smaller / shorter study to demonstrate a potentially important effect. (often up to 20x as large, e.g. DOI: <https://doi.org/10.3399/bjgp18X696545>).

Argument against: attributing the cause of death is difficult, and susceptible to biases (where as overall mortality is not). Empirically, the overall death rate and disease-specific death rate have not been congruent in many prior trials – possibly because the harms of the screening tests and harms of treatment for overdiagnosed disease.(see below summary) (<https://doi.org/10.1093/jnci/94.3.167>)

A screenshot of a cell phone

Description automatically generated

Probably not, if our goal is to infer if we are ‘saving lives’.

**What harms did the authors report? Are there other (non-reported) harms that would likely be present if the screening protocol were implemented?**

* True positive but inconsequential (=Overdiagnosis): harm of labeling as disease, investigations and treatments for a disease that would never have impacted them
* False positive: harms associated with further investigation of the abnormality, anxiety relating to investigations and treatment
* False negative: false reassurance, delayed presentation
* All: cost

**Blinding:** still possible, and quite important for disease-specific mortality – there is a lot of ambiguity in how deaths are coded. Not so important for all-cause mortality

Did they assess patient-important non-mortality end-points, such as quality of life?

Intervention:

**Did the authors discuss how they determined the screening frequency?** In general, more frequent screening increases the cumulative sensitivity, but also the cumulative exposure to harms.

**How would the studied effect change as treatments for the disease change?** Important to note that we are testing a treatment process – the effect of testing, the current diagnostic process, followed by the current treatment process. In general, as treatments for the disease gets better, the screening gets less effective (unless the treatment is only effective in early stage disease, e.g. a surgical advancement) – e.g. testicular cancer, where we no longer recommend screening because it’s still 97% curable if diagnosed when metastatic.

Results

**Why might relative risks (as opposed to absolute) be important to report?** They are less influenced by the baseline risk of the population – and so may represent the effect size of the treatment better – but are less applicable to patient care. The lower the risk in the control group, the larger the difference between relative risk reduction and absolute risk reduction.

**How much loss to follow-up was there?** In general, if over 20% this may be problematic (though the consequence of missing data depends greatly on why the data is missing. Death records (all-cause mortality) should be very complete.

Statistical concepts relating to screening RCTs:

How does **‘overdiagnosis’** differ from ‘false-positive’? Overdiagnosis is correctly diagnosing disease – such as prostate cancer - (or ‘psuedodisease’ – such as a colon polyp, that may or may not become ‘disease’) that would have never impacted the patient. It is not possible to know, except in hindsight, whether a diagnosis was ‘overdiagnosis’ for an individual. However, at the population level, increasing incidence of a disease without a change in mortality from the disease suggests its present. False positive is making an incorrect determination about the presence of a disease (or an abnormality requiring further investigation) – which can be determined in real-time at the patient level. Another tip-off is when the increase in diagnoses of ‘early stage’ disease is not off-set by a decrease in late stage disease.

Interesting aside, when determining the sensitivity and specificity of the testing, we really should be considering detection of clinically realvent disease (e.g. not including overdiagnosis cases). This is often not done in practice, because you can’t determine for indiv case until post-mortem (From H Gil Welch DOI: 10.1056/NEJMsr1905447)

**A close up of a map

Description automatically generated**

**The following two biases are addressed by RCTs assessing overall mortality and are the reason why other evidence is insufficient unless bolstered by arguments against these:**

**-Lead time bias**: even if the screening intervention does not influence clinical course, by finding patients earlier in their course, it will appear that they live longer from the time of diagnosis.

**-Length time bias:** screening catches mostly slow growing disease (because it will be present at more time points, thus if you diagnose relatively more slow growing disease, outcomes will appear improved.

Analogy: Trying to fence in animals - birds (rapid growth, can’t be caught = aggressive, screening wouldn’t have mattered), turtles (grow slow, but wouldn’t leave the fenced area = indolent tumors disproportionally caught by screening), rabits (fence actually works = the cancer we want to pick up on screening).

From Jonathan Howard “Critical Thinking In Medicine” - https://twitter.com/JHowardBrainMD/status/1142443094144626689

|  |  |
| --- | --- |
| A screenshot of a cell phone  Description automatically generated | A screenshot of a cell phone  Description automatically generated |
| A screenshot of a cell phone  Description automatically generated | A screenshot of a cell phone  Description automatically generated |

A screenshot of a cell phone

Description automatically generated

Shared decision-making:

Is there an ethical difference between considering medical care (treatment) of an ill-patient vs medical care (screening) for a health patient? (would argue so – we should have a higher bar of certainty.

How do you get ‘informed consent’ on screening, when it is difficult to anticipate what the consequences of screening will be for an individual? (both because the workup/treatment may vary significantly, and because medical practice is likely to change).

What values-preferences might cause a patient to forego treatment that has been shown to reduce overall mortality?

**Is the demonstrated benefit large enough to warrant any adverse effects / costs associated with the intervention** (which is almost always going to depend largely on what the patient’s baseline risk is for the outcome we’re trying to prevent)?