Clinical trials for evaluating the mortality benefits of MCD testing

Presenter: Hormuzd Katki, Senior Investigator, Biostatistics Branch, National Cancer Institute

October 28, 2024

NIH NATIONAL CANCER INSTITUTE

Need for RCTs when screening benefits are unknown, as is the case for MCD

- Long been known that those who choose to get screened outside of screening guidelines tend to be “health-seekers” with unexplainably lower mortality than population average

- Observational studies, even “target-trial emulation”, can only partially control for confounding by who chooses to get screened
  - Gold-standard target-trial emulation observational study in SEER-Medicare of mortality reductions from mammograms past age 70 found that mammograms reduced *endometrial cancer* mortality by 17% (CI: -7% to 50%)
    - Garcia-Albeniz et al, *Ann Intern Med* 2020

- What are the characteristics of those who will, and will not, choose MCD testing?
  - Unknown and likely even harder to predict than for mammography
  - Observational studies will tend to overestimate mortality reductions from MCD screening because propensity score model only partially accounts for health-seeking behaviors in those who got an MCD test
  - Well-conducted randomized controlled trials (RCTs) are internally valid and avoid this confounding issue

Limitations of RCTs for screening

- To obtain internal validity, RCTs tend to sacrifice external validity
  - Volunteers are both self-selected, and chosen by trial entry-criteria, to be healthier than the general population
  - RCTs tend to be conducted at major medical centers, which tend to have above-average outcomes

- Trial estimates of mortality benefits may not translate to the general population considered for screening
  - NNS to prevent 1 death was 320 in NLST, but if we apply the NLST 20% mortality reduction to the US NLST-eligible population of 2010-2012, the NNS would be 220
    - NLST-eligibles in the US have substantially higher lung-cancer mortality risk than NLST members
    - NNS is increased by about 50% in the NLST (Katki et al, *JAMA*, 2016)

- Trial estimates of screening harms may not translate to the general population of medical centers
  - Trials tend to take extra care to avoid harms
  - Harms are well-known to be lower at major medical centers with greater experience and the best equipment for conducting medical procedures

Statistical modeling of potential cancer mortality reductions in an MCD screening trial

- **Ping Hu**
- Hormuzd Katki
- Philip Prorok

Ping Hu

Hu et al, *JNCI*, 2024 (online early)

Base-case Hu-Zelen model parameters that obtain high statistical reproducibility ("90% statistical power") for cancer mortality reduction

- Model enables calculating the expected number of cancer mortality outcomes in control vs. screening arms for yearly screening with fixed number of screens
  - Calculate the power for a mortality reduction
  - Model assumes disease is purely progressive between 3 states: healthy -> preclinical –> clinical
  - Ages 60-74, SEER incidence, 100,000 control arm and 100,000 screening arm, 5 annual screens

- Consider mortality from 9 cancers
  - Lung, CRC, Pancreas, Liver, Esophagus, Stomach, Head/Neck, Lymphoma, Ovary

- Per-screen sensitivity during pre-clinical phase (β) set to be lower than for single-screen tests
  - For LDCT lung screening, β=80%, but we set lung MCD screening β=50%
  - Set CRC β=40% and all other cancer organ sites β=30%

- Set stage shift from 4 $\rightarrow$1 that is set lower than for single-screen tests
  - For NLST CT-arm had 35% reduction in stage-4 proportion (48%-13%), we set MCD lung stage-shift to 23% (48%-25%)
  - Stage shift for CRC and Liver: 10% (b/c not much stage-4 cancer for either), and 22% for all other 6 cancers

Base case statistical power is driven by lung cancer

		Follow-up year		
		7	8	9
All 9 cancers	deaths prevented	138	159	173
	mortality reduction (%)	10	10	9
	NNS	724	628	578
	power (%)	87	89	88
Lung	deaths prevented	91	104	111
	mortality reduction (%)	14	13	11
	NNS	1098	961	900
	power (%)	83	85	83
CRC	deaths prevented	11	14	17
	mortality reduction (%)	8	8	8
	NNS	9090	7142	5882
	power (%)	17	19	21
other 7 cancers combined	deaths prevented	36	41	45
	mortality reduction (%)	7	6	5
	NNS	2777	2439	2222
	power (%)	31	31	31
% deaths prevented from lung cancer		66%	65%	64%
% deaths prevented from CRC		8%	9%	10%
% deaths prevented from 7-cancers		26%	26%	26%

- 85-90% power is achieved in years 7-9 following 5 screens
  - 9-10% mortality reduction from the 9 cancers
  - NNS of 578-724 is competitive with mammograms
- 64-66% of prevented cancer deaths at each year are lung cancers
  - 11-14% lung cancer mortality reduction is substantial, but less than NLST (20%)
  - Lung cancer mortality by itself has ~83-85% power
- Rest of cancer sites still matter, although no power to isolate their mortality reductions
  - 8-10% of prevented cancer deaths are CRC cancers
    - 8% mortality reduction
  - 25-26% are the other 7 cancers
    - 5-7% mortality reduction

Other lessons from Hu-Zelen modeling of MCD trials

- 90% power was achieved in a relatively short 7-9 year time span
  - This relatively short time span is driven by the predominance of prevented lung-cancer deaths, for which early mortality reductions have been observed in NLST and other lung screening trials
  - A relatively short projected time span when using cancer-mortality endpoints may alleviate calls to accelerate MCD RCTs by using surrogate endpoints

- Some cancers are more/less amenable for early-detection screening
  - Lung-cancer is most amenable because it has by far the most common (best NNS of 900-1100), currently most lung-cancer is detected at stages-3/4 (which have poor survival), and stage-1 cancer has good enough survival
    - Second most important was CRC
  - Next came stomach, ovary, and esophagus
    - All have excellent stage-1 survival yet currently very few of those cancers are detected at stage-1
  - Some cancers have characteristics inhibiting effective early-detection screening
    - Poor stage-1 survival (liver, pancreas) or excellent stage-4 survival (head/neck, lymphoma)

- No power in MCD RCTs to isolate mortality reductions for cancers we currently do not screen for in general SEER-risk populations
  - Can we confidently recommend use of MCD tests to screen for the currently unscreened cancers, in general SEER-risk populations?

Increasing power in screening trials by testing stored specimens in the control arm: Application to MCD screening

- Hormuzd Katki
- Paul Pinsky
- Philip Prorok
- Philip Castle
- Lori Minasian

Katki et al, *JNCI*, 2024

Increasing power in screening trials by testing stored specimens in the control arm: Application to Multicancer Detection (MCD) screening

- Randomized controlled trials (RCTs) for novel cancer screening tests have historically required large sample-sizes (~50,000 per arm) and long time-horizons (~7-12 years) to achieve 90% power for cancer-specific mortality outcomes

- To improve the feasibility of screening trials, we describe a design and analysis based on the concept that screening only affects the primary outcome in those who ever have a positive screen.
  - This approach reduces the “noise” of events in “never-positives” by exploiting information gained by testing stored control-arm specimens, which is suited for blood-based screening tests, such as MCD tests
  - This approach, which we call the **Intended Effect (IE) design and analysis**, could substantially increase statistical power, which could be used to either reduce sample size or accelerate the time to 90% power
  - Commentaries have noted variants of this approach (Weiss, *J Clin Epi*, 2013), particularly for MCD trials (Hackshaw and Berg, *Lancet Oncol*, 2021; Weiss, *J Natl Cancer Inst*, 2024)

Katki et al, *JNCI*, 2024

Standard vs. Intended Effect (IE) design and analysis

- Standard design and analysis
  - Screen-arm subjects have blood drawn, tested in real time, and are informed of their result
  - Control-arm subjects are simply followed up for outcomes – no blood or material is stored
  - Compare the everyone in the screen arm to everyone in the control arm with the relative risk (RR) or risk difference (RD)

- The IE design
  - Screen-arm subjects have blood drawn, tested in real time, and are informed of their result
  - Blood is also drawn and stored from all control-arm participants according to a common protocol
  - All control-arm specimens would be tested towards the end of trial follow-up to ensure there is no effect on control-arm outcomes

- The IE analysis
  - Calculate the RR and RD *only among participants in both arms who test positive on at least 1 screen ("ever-positive")*
    - The justification is that trial arm assignment should have *no intended effect* on outcomes for those who never test positive on any screen ("never-positives")
    - Never-positives in the screen-arm never experience diagnostic procedures that are triggered by their screening test result and thus their cancers could not have been detected early
    - Hence never-positives should experience no intended effect from screening
  - Removing everyone whose outcomes are unaffectable by screening (which is the IE analysis) should increase the relative mortality reduction and reduce the p-value

Example MCD trial: standard vs IE analysis

Standard trial analysis table

	screen	control	
D+	900	1,000	1,900
D-	49,100	49,000	98,100
	50,000	50,000	100,000

1,000-900 = 100 prevented deaths in trial (risk difference [RD] = 0.2%), with relative risk (RR = 900/1,000 = 0.90; *P* = .019). Power for the standard analysis is 65%.

95% never test positive on any screen: IE analysis removes these people

Never-positive table

	screen	control	
D+	250	250	500
D-	47,250	47,250	94,500
	47,500	47,500	95,000

No effect of screening for never-positives: 250-250 = 0 prevented deaths

RD_{neg} = 0% and RR_{neg} = 1

Remove never-positives because screening had no intended effect on their outcomes

IE analysis:
Analyze only the 5% who test positive on any screen (P_{pos} = 5%)

- $P_{EV\text{-}pos}$ = 750/1000 = 75% of control D+ events in ever-positives

Ever-positive table

	screen	control	
D+	650	750	1,400
D-	1,850	1,750	3,600
	2,500	2,500	5,000

All 750-650 = 100 prevented deaths concentrate in ever-positives (risk difference among ever-positives RD_{pos} = 4%).

Relative risk among ever-positives RR_{pos} = 650/750 = 0.867 is stronger than the overall RR = 0.90

P = 0.0014 is 14 times smaller than overall *P* = 0.019

Power for the IE analysis is 89%

Intended Effect (IE) design/analysis issues for MCD trials

- Statistical power gains from the IE design (which can be used to reduce sample sizes or to reduce p-values) that were shown in the previous slide hold generally under IE assumptions
  - Most important: Non-compliance in control-arm members providing blood samples is comparable to non-compliance in screen-arm members
    - Equivalent non-compliance across arms could be assured if trial participants were blinded to arm assignment

- IE can increase statistical power for any outcome
  - Cancer mortality, or reduction in stage-4 incidence, for all cancers or individual cancers
    - In particular for currently non-screened cancers

- IE design is in accordance with principles of medical ethics
  - No harm is done to participants and all participants are encouraged to receive standard-of-care screenings
  - Question of appropriateness of testing the stored specimens of participants in the control arm without returning the results
    - Currently, it is not known whether acting on MCD results (ie, working up the patient for a presumed cancer diagnosis) will result in an overall health benefit, especially reducing cancer-specific mortality in the screen arm over the control arm
    - This equipoise suggests that there is no ethical imperative to return results, which would be many years old and not medically actionable
    - At the end of the study, control-arm members can be offered a free up-to-date MCD test, which could be medically actionable

Thank you