Measuring the mortality reductions produced by Irish and Danish breast-cancer screening programs. CASI, Dundalk, Ireland, 2019.05.15.

Good afternoon. 8 years ago I convinced this agency to fund this neglected but big-ticket measurement question. Amy Liu's thesis developed a new statistical model and applied it to trials of colon and lung cancer screening. But for breast cancer we had to look to population data. 

I will start with the traditional measures, and then give you some sharper more meaningful estimands that provide more useful answers 21 / 68

We know a lot about the COSTS mammography screening programs: the financial outlays and the individual harms. The easiest BENEFIT to measure should be the number of breast cancer deaths averted, but even on this measure analysts cannot agree. One big reason is the arbitrariness of their estimands. There are contemporary population-level data from countries that staggered the introduction of their organized programs. But the Big country-level Data that are easily obtained are not sharp enough, and dilute the reductions. 

The sharper studies use diagnosis dates from the cancer registry to define the women targeted by the screening program. Let's start with Denmark. Copenhagen, here, was the first area to introduce a screening program. 

I will focus on the province of Funen, here, which began in 1993 , again well before most of the rest Denmark. In 2015, Sisse and colleagues compared the mortality in the relevant woman-years 14 years before and after it started. In case this Lexis diagram is new to you, COHORTS proceed along the diagonal, and become 1 year older in AGE every calendar YEAR; all three critical elements – age, period and cohort – are shown in one diagram. The shaded areas are the woman years that would be impacted if screening was from age 50 to 69. Some of the pre-post difference in

mortality rates might be due to improved management and treatments over time, rather than screening per se, and so they used the pre-post difference in the still-not-screening regions of Denmark to estimate this and calculated a double difference to measure the portion attributable to screening.

Ireland's BreastCheck program began in 2000 in these 11 eastern counties. it was extended to these 3 in these years, and the last 12 at the end of 2007. We focused on the earliest and latest. Different from most programs, screening in BreastCheck used to end at 64 rather than 69 (the extension 69 is being phased in now). 59 / 391

First, Sisse's 2015 analysis and results for FUNEN and the 8 times bigger non-screening comparison experience. 16 / 407

The cross-product ratio of breast cancer mortality rates gives an adjusted hazard ratio of 0.78, or a 22% 'reduction' that they (cautiously of course) attributed to the screening program. i.e., there were an estimated 22% fewer breast cancer deaths than there would have been if they hadn't screened for these 14 years. 52 / 459

Now to Ireland. Recall the basic comparison, between 2 regions that started screening almost 8 years apart. But what if these two regions had different mortality rates even in the absence of screening? The Irish Cancer Registry did not begin until the mid 1990s so we could not use the same type of historical comparison that was used in Denmark. So we opted to stay entirely in the 21st century, and for a 'control', use the experience of women who were already OLDER than the upper screening age of 64 when screening was first introduced in 2000. These woman-years allowed us

to check if there were differences in the background cancer death rates in the 2 regions and to correct for them. ↓ 123 / 582

Let's look first at these older WY – lived by women born before 1936. As you can, see the death rates are very close, but slightly lower in the first (eastern) region to start screening that the western region where women 50-54 had to wait. So when we compare the rates in the screen-eligible WY, we will have to handicap the east just a tad. 65 / 647

What happened in the same 14 years in the woman-years targeted by screening? As you can see the 2 rates were 12% lower in the region that started first. 29 / 676

So when we take the ratio of the hazard ratios so as to handicap the East, we get a corrected HR of 0.91, ie. a 9% difference. We can interpret this as saying that the almost 8 years' more screening led to 9% fewer deaths in East in the 14 years. But what if we asked the more relevant counterfactual comparison: how would the rates in the East have looked relative to those we would have seen there if the program had not been introduced at all? Or if there were a full 14 year gap between the East and the West, like in Denmark. The 9% is merely a lower bound. Because of

the delays before the full results in the East are realized, we can only conjecture as to how much more than 9% it would have been. 140 / 816

Two of the problems with the meta-analyses of the old trials, and even the better population-based comparisons, is that they largely ignore the delays before mortality reductions show up, ie that hazards are inherently non-proportional, and the variation in numbers of invitations. Amy Liu's thesis took the fundamental parameters to be the effect of 1 round, and used them to built up a bathtup shaped HR function over the trial follow-up time. She only applied her model to trials, but we dont have recent ones in mammography. 87 / 903

Remember the FUNEN data I showed you earlier. 8 / 911

In 2015, I contacted Sisse and suggested we try Amy's simple parametrization but add the age-dimension to the program-year dimension. 20 / 931

She was keen, and updated the follow-up to the 22 years shown in this compact Lexis diagram that drops the pre phase. The black dots every 2 years are the invitations to Funen women, stopping at 69. (This right hand wall is when the next part of Denmark started screening). Those aged 69 when the program started got just 1 invitation. Other birth cohorts got many more. The Rest of Denmark is white dots every year. 76 / 1007

The easiest way to understand the 2 fundamental parameters in our model is to think of an unscreened population, and women whose cancers that proved fatal in say 2019. Then ask oneself, if these women could have been offered just one screen, when in the past would have been optimal and what percentage of them would have had these deaths averted because of the earlier detection and treatment? 68 / 1075

In the blue curve in the diagram on the right, the sweet spot (tau) is about 7 years earlier and the maximum percentage (delta) is about 8%. The blue curve is the probability of being helped if the 1st and only screen were x = 0, 1, ..., 22 years before the cancers would(otherwise) have proved fatal. The black curves show the probabilities for 2, 3, 4 .. rounds and can be thought of as convolutions or amalgamations of the benefits of multiple screens. I will leave the details on the left to question time. 95 / 1170

On the left are what the data look like, one row per Lexis cell. The first row is for those aged 87 in 2014. 11 died of breast cancer in the Rest of Denmark, and 1 in Funen, among approx 17,000 and 2,000 respectively. The Funen women had received 2 invitations, 20 and 18 years earlier, when they were 67 and 69. Those in the second row had had 4 and those in the third had had 7. The no. and timing of the invitations are the x's in the HR regression function.

Here are the fitted mortality deficits or % reductions, based on convoluting the fitted parameters and invitation histories. Those in the uppermost diagonal had just 1 invitation, so the top numbers are the fitted blue curve for 1 round of screening, reaching a nadir of 8% at year 7. The lower down ones had more invitations and so the trough is deeper and longer. The overall reduction is about 19%, an average of reductions ranging from 0% to 30%. 79 / 1342

FUNEN's 14 year time gap made it a bit easier to fit the 1-round parameters. When I initially tried the model on the Irish data, with less than an 8 year gap, and treated the west as entirely unscreened, I had trouble, so the PLOS article only had the less-meaningful overall 9% difference. I have since refined the data-analysis to allow for the second startup and am now able to report the 2 fitted parameters and the fitted HR function over the Lexis space. 84 / 1426

Here are the 2 sets of fitted mortality deficits, one for each region. Women in the uppermost diagonal in the FIRST region (where the larger bold numbers are) had just 1 invitation – in 2000; the fitted curve for 1 round of screening reaches a nadir of 6% at year 6. In the lower down ones the troughs are deeper and longer. The largest reduction is about 19%. In the second region, women in the several uppermost diagonals had no invitations, so had 0 benefit; in the most-often invited, the fitted reductions have only reached 10%. The 2 sets of Hazard ratios explain why the

average difference between the regions was only 9%, and it will get smaller as the follow-up is extended. 123 / 1549

we have more detail here [I will repeat this at end] 11 /

1560

So, to summarize... Part of the reason the monetary costs and harm have been better quantified than the benefits is that basic cancer screening principles have been ignored. Our parametrization is minimalist, and leads to non-PH HR functions that are more realistic and more meaningful. It applies both to trial and population data. As for Breastcheck in particular: When they started, this was their stated goal. to reduce breast cancer mortality by 20% in 10 years. For those cohort of women who on the main diagonal, i.e., invited from age 50 onwards, we think that

close to a 20here are some references, 4 / 1842

and the website containing all of our work 8 / 1850

Thank you to my collaborators, to my funder over the last 8 years, as well as the Institute that paid my Air Canada ticket when I went to the University of Waterloo 50 years ago this September. 37 / 1887