# Solution 6: Intervention studies

## Section 1: The New York Trial

**Question 1.** A randomised controlled trial (although we cannot be completely sure since the method of allocation was not fully described).

**Question 2.** The aim of the study was to measure the effect of screening on breast cancer mortality (death).

While the **actual** incidence of breast cancer would be similar across randomly allocated study arms, we expect the **measured** incidence of breast cancer over 7 years to be higher in the screening arm than in the control arm due to improved diagnosis.

**Question 3.** The main reason to use randomisation is to ensure balance on all known and unknown confounders (alternative risk factors) for breast cancer death.

**Question 4.** At randomisation, there likely were similar numbers of women with a history of breast cancer in both groups. However, the manner in which the investigators subsequently excluded women likely resulted in an imbalance in participants with a history of breast cancer across study arms, such that more of these women were in the control arm than in the mammography arm. Also, women with a history of breast cancer are at higher risk of subsequent breast cancer and breast cancer death. This might result in an increased breast cancer mortality rate in the control arm vs intervention arm, and we are not able to exclude selection bias as an explanation when we interpret the results of this study. To minimize this kind of bias, it is very important to assess eligibility and obtain consent BEFORE randomisation to ensure that only chance determines which group a trial participant enters.

**Question 5.** When randomisation has been done properly the null hypothesis that the two groups come from the same population is true i.e. any observed difference must be due to chance. We therefore do not recommend using p-values to compare baseline characteristics between randomised groups. The best way to judge the adequacy of the randomisation is to consider the method that was used and make a judgement as to whether or not this is adequate (such as for example telephone randomisation) or inadequate (such as for example alternation, which means that the next assignment can be predicted, and is therefore discordant with the principle of randomisation).

If the groups were not adequately randomised then there may be differences between the groups other than whether or not treatment was received. This means that it is difficult to judge whether differences in outcomes (e.g. breast cancer death) between the groups are a result of the intervention or other baseline differences.

**Question 6.** The fact that researchers assessing cause of death knew the screening status of the women may have introduced bias since this knowledge may make them more or less likely to record a death as a breast cancer death. This would be an example of differential misclassification of outcome (see Lecture 5: Chance and bias).

We cannot tell from the information presented whether the difference was due to bias or whether it was an effect of the screening programme. It seems possible that the bias occurred as more women in the control group who had a “dubious” status were classified as breast

cancer deaths than the women in the intervention arm. This bias could have overestimated the number of breast cancer deaths in the control arm and so made it seem as if intervention reduced the risk of breast cancer death.

**Question 7.** An analysis according to the group to which the women were randomised (usually called an intention-to-treat analysis) should provide a closer estimate of the 'real-world' effect of a policy of **offering** breast cancer screening. An intention-to-treat analysis is usually preferable to an analysis by actual screening status ('per protocol' analysis), which provides a closer – albeit potentially confounded – estimate of the effect of policy implementation with perfect adherence.

**Question 8.** Women who had mammography screening had 0.65 times the risk of dying from breast cancer, as compared to those who did not have mammography screening. The 95% C.I. for the risk ratio does not include one, and so the result provides strong evidence that the screening programme reduces the risk of breast cancer death over 7 years of follow-up. However, we have seen a lot of potential for bias so whether we believe this result is a matter for judgement!

**Question 9.**

The risk difference is 124/31000 - 81/31000 = 0.40%-0.26%=0.14% or 1.4 per 1000 over 7 years of follow-up. The number needed to treat (NNT) is 1/risk difference = 1/0.0014 ≈714. Note: Without rounding, the risk difference is 0.00138709…% and the NNT will be ≈721.

This means that 714 women need to undergo mammography screening to prevent one breast cancer death, according to these data. NB: The NNT figure assumes the causal relationship is established, and at the magnitude estimated by the risk difference.

**Question 10.** It would be important to consider whether there is genuine clinical uncertainty – i.e. equipoise - as to whether screening is beneficial or not. Given the potential for bias in this study and the few other studies of screening (this was "one of the first"), it could be considered ethical to undertake further well designed and conducted randomised controlled trials.

## Section 2: Meta-analysis

**Question 11.** Within trials conducted in low and intermediate burden areas there was no evidence of a difference in weight associated with deworming. In contrast, in trials conducted in high burden areas, the mean difference in weight was 0.57kg (0.08 to 1.06kg), with two trials reporting a mean weight gain of more than one kg. Across all groups ('total') the mean difference in weight between arms was 0.23 kg (0.05-0.42), favouring deworming.

**Question 12.** In their discussion, the authors explain that the (few) trials that saw large effects on weight were conducted more than twenty years ago, and the effects have not been in the numerous larger studies conducted recently. Accordingly, the certainty of evidence was graded 'Very low' due to inconsistency (in addition to risk of bias, "as most trials did not adequately describe allocation concealment").

While the authors do not make an explicit recommendation either way for mass deworming of children, they conclude by saying "We would caution against selecting only the evidence from

these older studies as a rationale for contemporary mass treatment programmes as this ignores the recent studies that have not shown benefit."

Regular deworming of all children has become an increasingly polarised debate in recent years. While recognising that evidence in terms of weight gain and other important health outcomes is of "very low to moderate quality" WHO guidelines continue to recommend deworming of all children in areas where the prevalence of soil-transmitted helminths is greater than 20%, emphasising the benefit of treatment to those who are infected, the safety and acceptability of treatment, and relatively low cost of delivery.

## Section 3 (To do in your own time): The Malmo Study

**Question 13.** We are told that the age distribution is similar but we cannot conclude from this that randomisation was adequate – once again we need to consider carefully how randomisation was done.

**Question 14.** It would be very difficult or impossible to prevent women from knowing if they had been screened or not but this would not bias an assessment of the cause of death which was made by an observer. The crucial difference between the Malmo study and the New York study is that in the Malmo study the researchers did not know the screening status of the woman when assigning cause of death, whereas in the New York study the researchers were aware of the screening status of the deceased woman. The woman's awareness or not of her screening status should not influence the researcher's assessment of the cause of her death.

**Question 15.** The 95% Confidence Interval includes the null value of 1.0, so we are unable to conclude that there is truly a 4% reduced risk of breast cancer death associated with screening.

## References


Kunz R, Oxman A. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ 1998; 317:1185-1190.

Altman D, Bland M. How to randomize. *BMJ* 1999; 319:703-704.

Hewitt C, Torgerson D. Is restricted randomisation necessary? *BMJ* 2006 332: 1506-1508.

Roberts C, Torgerson D. Baseline imbalances in randomized controlled trials. *BMJ* 1999; 319:185-185.