tumours from grade 2 tumours, and grade 2 tumours from grade 3 tumours. We are currently studying inter-
observer variations in histological grade, scoring of oestrogen-receptor expression, progesterone-receptor expression, and ERBB2 expression for all patients from the 16 participating centres enrolled in the RASTER (MicroarRAy prognoSTics in Breast CancER) study and are comparing the original assessments for these parameters with findings obtained at central review at the Netherlands Cancer Institute.

Finally, Neven and colleagues ask whether the 70-gene prognostic signature is any better than traditional clinical prognostic markers. The RASTER study was a feasibility study assessing the use of the 70-gene prognostic signature in community hospitals and was not designed to answer this important question. The MINDACT trial (Microarray in Node-Negative Disease may Avoid Chemotherapy trial), which is currently recruiting patients, has the optimum design to answer this question in a prospective setting. Meanwhile, we can only make some general comments on this issue. A major problem with traditional clinical prognostic markers is a paucity of standardisation in the techniques used to assess these factors, most notably for the histological grading. In the original Nature publication, most of the well-established prognostic factors mentioned by Neven and colleagues have been assessed in various multivariate models. However, the 70-gene prognostic signature remained the strongest independent prognostic factor in all of these comparative calculations.

*Sabine C Linn, Jolien M Bueno-de-Mesquita, Marc J van de Vijver
Department of Medical Oncology, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands
s.linn@nki.nl
MJvdV is a named inventor on a patent application for the 70-gene signature used in the RASTER study. All other authors declared no conflicts of interest.

had received mammograms during the randomised phase of the trial, and Biesheuvel and colleagues refer to data collected up to 6 years after the trial ended, which are diluted. The very low estimates of over-detection of breast cancer that Biesheuvel and colleagues provide for this trial, about 2% and 7%, are, therefore, also substantially downwardly biased.

A further shortcoming in the review by Biesheuvel and co-writers is that they adjusted the increased incidence of breast cancer for lead-time (ie, the mean time that screening advances diagnosis) based on theoretical models. Thus, they assume that the increased incidence during screening is followed by a decrease in incidence when screening stops, and the longer the lead-time, the bigger and longer-lasting the decrease in incidence would be. Lead-times are usually estimated to be around 2–5 years but this has never been observed in trials. Boer and co-workers predicted that the incidence of cancer should fall by about 50% immediately after age 69 years and approach the same incidence as that of the general, non-screened population after about 10 years. However, for all countries we have examined, the reduction in breast cancer incidence after age 69 years has been small or non-existent (this includes countries that have a low level of opportunistic screening after age 69 years). In the figure, we illustrate this for Norway. We obtained age-specific incidence for breast cancer from the Cancer Registry of Norway for the period 1991–2005 for four Norwegian counties (Akershus, Oslo, Rogaland, and Hordaland) where screening was first introduced (screening was introduced much later in the other counties, ie, from 2000-2004). About 40% of the population of Norway live in these four counties; 376 693 women who were aged 40–79 years in 2005 and 11 957 women with breast cancer were registered in the observation period. The red line in the figure shows the incidence of breast cancer for the period 1991-1995 before the nationwide screening programme started. The prevalence of cancer detected as a result of the first screening round in 1996-1997 is omitted from the figure. The other lines represent the incidence of breast cancer detected during the second (1998-1999), third (2000-2001), fourth (2002-2003), and fifth screening rounds (2004-2005). Only a small decline in the incidence of breast cancer after age 69 years was noted. This suggests that there is something fundamentally wrong with lead-time models, and that adjusting for lead-time will seriously underestimate over-detection of breast cancer.

Biesheuvel and co-writers state that Zahl and colleagues did not adjust for lead-time. However, formal adjustment is inappropriate. Instead, Zahl and colleagues noted that the 12% decline in incidence in the 70–74 year age group in Norway accounted for only 3% of the 54% increase in incidence in the 50–69 year age group that was eligible for screening.

Biesheuvel and co-writers concluded that the least biased estimates of over-detection of breast cancer ranged from –4% to 7.1% for women aged 40–49 years, 1.7% to 54% for women aged 50–59 years, and 7% to 21% for women aged 60–69 years. We believe it is unhelpful to readers to provide such broad ranges of estimates. Furthermore, they are not correct—screening cannot lead to less cancers but must lead to more cancers because it is well-known that a large reservoir of slow-growing cancers and early lesions such as carcinoma-in-situ exist that are detected by screening.
If Biesheuvel and co-workers had addressed the biases more carefully, they could have reduced their ranges considerably. We have data from a large number of publications and databases that describe changes in breast cancer incidence that have occurred after screening was introduced in many countries and with long follow-up. Taken together, these data strongly suggest that the amount of over-detection of breast cancer in practice is actually higher than the 30% we first estimated based on the randomised trials.1

Per-Henrik Zahl, Karsten Juhl Jørgensen, Jan Maehlen, *Peter C Gøtzsche

The Nordic Cochrane Centre, Copenhagen, Denmark (KJJ, PCG); Norwegian Institute of Public Health, Oslo, Norway (PHZ); Department of Pathology, Ulleval University Hospital, Oslo, Norway (JM)
p.c.gotzsche@cochrane.dk

The authors declared no conflicts of interest.

3  Gøtzsche PC, Jørgensen KJ. Estimate of harm/benefit ratio of mammography screening was five times too optimistic. http://bmj. bmjournals.com/cgi/eletters/332/7543/691 (accessed Feb 12, 2008).

Authors’ reply

We agree with Gøtzsche and colleagues that over-detection of invasive breast cancer as a result of breast screening programmes is important, and might be substantial in magnitude. For example, if the extent of over-detection is at the top end of the best estimates that we have identified it could be about 50% for women in their 6th decade and 20% for women in their 7th decade. Such over-detection would have profound implications for breast screening programmes.

However, in regard to the criticisms of our review we would like to note the following. In our systematic review,1 we identified both low screening participation in screening populations (and some screening participation in non-screening [ie, control] populations) and insufficient consideration of lead-time as common and important factors that can lead to bias in studies of breast-cancer screening. We noted that the effect of the first of these biases would be estimates of cancer over-detection that are too low, and the effect of the second type of bias would usually be estimates of over-detection that are too high. The over-detection estimates reported by Zackrisson and co-workers on the Malmö trial2 and the estimates reported by Moss and colleagues on the Canadian trials3 are indeed affected by non-compliance and are, therefore, biased downwards. However, these estimates are the least biased of all published estimates derived from randomised clinical trials of breast-cancer screening.

Lead-time is an inevitable consequence of screening, even if it is not possible to observe it directly. If the cumulative-incidence method is used to calculate over-detection, not adjusting for lead-time will result in estimates of over-detection that are too high, unless the duration of follow-up after screening ended exceeds the lead-time. The authors of the Cochrane review4 did not adequately adjust for lead-time. Because they used the cumulative-incidence method, they could simply have allowed adequate follow-up time after the trials ended before estimating over-detection. However they did not do this, nor did they adjust statistically for lead-time. Therefore their over-detection estimates are biased, and thus are too high.

Gøtzsche and colleagues comment that we accounted for lead-time twice. If the incidence-rate method is used to calculate over-detection, we indeed recommend excluding the initial breast-cancer screening rounds and recommend adjustment for lead-time. Excluding the initial screening rounds adjusts for the lead-time of prevalent cancers, and further adjustment is then only needed to account for lead-time of incident cancers detected in subsequent rounds (as illustrated in figure 3 in our systematic review5). This method will not be able to adjust for prevalent cancers that were never destined to become clinically relevant and therefore this method produces lower estimates of over-detection than the cumulative-incidence method. This is an inherent problem with the incidence-rate method and, therefore, the cumulative-incidence method is better to estimate over-detection.

The Norwegian data, which Gøtzsche and colleagues say fail to show an important dip in incidence for women aged 70–74 years who finished screening at age 69 years,5 are also biased as we have noted in our systematic