

## Editorial

# The screening mammography paradox: better when found, perhaps better not to find

DA Berry<sup>1</sup>

<sup>1</sup>Division of Quantitative Statistics, Department of Biostatistics, University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Box 447, Houston, TX 77030, USA

British Journal of Cancer (2008) 98, 1729–1730. doi:10.1038/sj.bjc.6604349 www.bjcancer.com

Published online 27 May 2008

© 2008 Cancer Research UK

In this issue, Wishart *et al* (2008) address whether ‘screen-detected breast cancer confers additional prognostic benefit to the patient, over and above that expected by any shift in stage at presentation’. Their conclusion is affirmative. Although this observation is not new and indeed is increasingly understood and accepted by the research community (Joensuu *et al*, 2004; Shen *et al*, 2005), their study is an important confirmation. One application of the observation has been in statistical models that make inferences about the relative contributions of screening and adjuvant therapy to the decrease in breast cancer mortality that has occurred over the last 20 years in the United States (and in other countries as well) (Berry *et al*, 2005, 2006b; Fryback *et al*, 2006). A conclusion of these models is that, taking therapy into account, the observed breast cancer incidence and mortality would be inconsistent unless some cancers preferentially detected by screening mammography had, in the terminology of one study (Fryback *et al*, 2006), ‘limited malignant potential’.

Wishart *et al* (2008) point out a clinical implication: the relatively good prognosis of screen-detected cancers means that clinicians are overtreating some patients. Indeed. The problem is that although we are learning rapidly about the biology of breast cancer and its impact on treatment (Bazell, 1998; Paik *et al*, 2004, 2006; Berry *et al*, 2006a; Hayes *et al*, 2007), our understanding is not yet sufficiently advanced to make clear which tumours need treatment, or which treatments are best for which biological subtypes. Nor do we know the proportion of patients in either detection category who will benefit from treatment. Knowing that fewer screen-detected than symptomatically detected tumours need treatment is not of much help unless we can identify which ones they are.

That screen-detected tumours have better prognoses is an important consideration in clinical research. Method of detection of breast cancer is not a standard prognostic factor recorded in clinical trials. This omission is regrettable. In some settings, method of detection may be as important as lymph node status in predicting disease recurrence. Happily, the research community is beginning to recognise this oversight. Trials such as TAILORx in the United States (Sparano, 2006; National Cancer Institute, 2008; Sparano and Paik, 2008) have incorporated method of detection on patient case report forms.

In their analyses of the importance of method of detection, Wishart *et al* (2008) correctly adjust for stage of disease and for the

Nottingham Prognostic Index (NPI) more generally. The question of clinical interest is whether given the clinical, pathological, and demographic characteristics of the patient and her tumour, it helps to know the method of detection as well? Their answer is yes.

How can this be? It has always been clear that there is a stage shift associated with screening mammography: screening tends to find cancer earlier than otherwise. But it is surprising that a tumour clinically and pathologically identical to another tumour is fundamentally different, simply because of the way it was found. A possible explanation for the additional benefit of screen-detected tumours is a ‘within-stage shift.’ For example, some node-negative tumours harbour the potential to become metastatic (or already are metastatic!) while others do not. But as I discuss below, other explanations are possible.

The title of the Wishart *et al* (2008) article is ‘Screen-detected *vs* symptomatic breast cancer—is improved survival due to stage migration alone?’ This could be read as implying that screening improves survival. In the text they say, ‘Our results confirm a strong survival advantage of screening compared with symptomatic detection’. This is literally and incontrovertibly true. Their words will be seen by many readers to imply that screening mammography improves survival. But their study cannot address the question of screening effectiveness, nor do they claim to address this question.

There is an inherent aspect of the Wishart *et al* (2008) study that makes it difficult to state conclusions without having them misinterpreted. I speak from experience. There was substantial press coverage for my observation that method of detection is an important independent prognostic factor in breast cancer (Shen *et al*, 2005). However, despite my protestations, most reporters interpreted the study as meaning that screening was effective (which is the reason for the press coverage!). Perhaps my abilities to explain such matters are limited; after prolonged discussion with one reporter, she wrote an article with this headline in large lettering: ‘Mammograms Boost Survival Odds.’ In a nod towards accuracy, she added a subtitle, in much smaller lettering: ‘The screenings often detect slower-growing tumours, a new study finds.’

Wishart *et al* (2008) make clear that their observation is at least partially explained by lead-time and length biases. Lead-time bias is the easier to understand of these two biases. Suppose screening finds breast cancer an average of 3 years before it would become

symptomatic. Then, this 'lead time' is added to the lives of women whose tumours are screen-detected and this additional 3 years may seem to be a benefit. But it is pure bias. Any benefit of screening has to be above and beyond this 3-year advantage. To get an unbiased estimate of screening effectiveness in a nonrandomised setting, one might subtract 3 years from the survival of every woman whose tumour is detected by screening. The problem is that in any particular setting we do not know that 3 years is the right adjustment: the true bias might be larger or smaller.

Length bias is more important than lead-time bias, at least in breast cancer. But neither its importance nor the concept itself is easy to understand. 'Length' refers to the tumour's presymptomatic period when the tumour is mammographically detectable. The length of this period is the tumour's sojourn time. Sojourn time varies from one tumour to another. (There is an obvious relationship between lead time and sojourn time; lead time is shorter because it requires actually finding the tumour during the presymptomatic period.) Sojourn time is typically positive, but it is negative for tumours that become symptomatic without being detectable on a mammogram. Breast tumours are heterogeneous, even after accounting for stage and other known clinical and biological characteristics. Aggressive tumours have shorter sojourn times because they grow faster. Indolent tumours have longer sojourn times. Screening finds tumours in proportion to their sojourn times, and therefore longer times and slower growing tumours are preferentially selected. This is length bias. (There are

many analogues: when you look in the sky and see a shooting star, it is more likely to be one with a longer arc; when you reach into a newly opened bag of potato chips and select one, it is more likely to be big.)

A special case of length bias is overdiagnosis, when screening finds a tumour with a sojourn time so long that the tumour would not kill the woman even if it was never found.

Again, Wishart *et al* (2008) make clear that their observation is at least partially explained by lead-time and length biases. Their Figure 1 encompasses both, as they understand. Adjusting for NPI and other factors at least partly removes lead-time bias. But it does not remove length bias. Some tumours grow slowly. By their very nature such tumours have better prognoses than do rapidly growing tumours. Fast growing tumours will eventually become symptomatic, and among women who participate in screening mammography programs, they are likely to be the ones detected symptomatically between screenings (so-called interval cancers). Some tumours grow so slowly that they would never be detected were it not for screening. The consequence is overdiagnosis.

Therefore, although the authors are correct in worrying that screen-detected cancers may be overtreated, a greater concern is that some screen-detected cancers should never have been detected in the first place! The rub is that just as with treatment, we do not yet have a good understanding regarding which cancers we do not want to detect. Mammography is too crude a tool to make this distinction.

## REFERENCES

- Bazell R (1998) *Her-2: The Making of Herceptin, a Revolutionary Treatment for Breast Cancer*. New York: Random House
- Berry DA, Cirincione C, Henderson IC, Citron ML, Budman DR, Goldstein LJ, Martino S, Perez EA, Muss HB, Norton L, Hudis C, Winer EP (2006a) Estrogen-receptor status and outcomes of modern chemotherapy for patients with node-positive breast cancer. *JAMA* **295**: 1658–1667, Erratum in: *JAMA* 295: 2356, doi: 10.1001/jama.295.14.1658
- Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JDF, Feuer EJ, Cancer Intervention and Surveillance Modeling Network (CISNET) Collaborators (2005) Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* **353**: 1784–1792
- Berry DA, Inoue L, Shen Y, Venier J, Cohen D, Bondy M, Theriault R, Munsell MF (2006b) Modeling the impact of treatment and screening on US breast cancer mortality: a Bayesian approach. *J Natl Cancer Inst Monogr* **36**: 30–36
- Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL (2006) The wisconsin breast cancer epidemiology simulation model. *J Natl Cancer Inst Monogr* **36**: 37–47
- Hayes DF, Thor AD, Dressler LG, Weaver D, Edgerton S, Cowan D, Broadwater G, Goldstein LJ, Martino S, Ingle JN, Henderson IC, Norton L, Winer EP, Hudis CA, Ellis MJ, Berry DA, Cancer and Leukemia Group B (CALGB) Investigators (2007) HER2 and response to paclitaxel in node-positive breast cancer. *N Engl J Med* **357**: 1496–1506
- Joensuu H, Lehtimäki T, Holli K, Elomaa L, Turpeenniemi-Hujanen T, Kataja V, Anttila A, Lundin M, Isola J, Lundin J (2004) Risk for distant recurrence of breast cancer detected by mammography screening or other methods. *JAMA* **292**: 1064–1073
- National Cancer Institute, US National Institutes of Health (2008) *The TAILORx Breast Cancer Trial*, <http://www.cancer.gov/clinicaltrials/digestpage/TAILORx> Accessed 02/11/2008
- Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* **351**: 2817–2826
- Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J, Costantino JP, Geyer Jr CE, Wickerham DL, Wolmark N (2006) Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* **24**: 3726–3734
- Shen Y, Yang Y, Inoue LT, Munsell MF, Miller AB, Berry DA (2005) Role of detection method in predicting breast cancer survival: analysis of randomized screening trials. *J Natl Cancer Inst* **97**: 1195–1203
- Sparano JA (2006) TAILORx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer* **7**: 347–350
- Sparano JA, Paik S (2008) Development of the 21-gene assay and its application in clinical practice and clinical trials. *J Clin Oncol* **26**: 721–728
- Wishart GC, Greenberg DC, Britton PD, Chou P, Brown CH, Purushotham AD, Duffy SW (2008) Screen-detected vs symptomatic breast cancer—is improved survival due to stage migration alone? *Br J Cancer* **98**: 1741–1744