Critical assessment of new risk factors for breast cancer: considerations for development of an improved risk prediction model

Richard J Santen, Norman F Boyd, Rowan T Chlebowski, Steven Cummings, Jack Cuzick, Mitch Dowsett, Douglas Easton, John F Forbes, Tim Key, Susan E Hankinson, Anthony Howell and James Ingle

Abstract

The majority of candidates for breast cancer prevention have not accepted tamoxifen because of the perception of an unfavorable risk/benefit ratio and the acceptance of raloxifene remains to be determined. One means of improving this ratio is to identify women at very high risk of breast cancer. Family history, age, atypia in a benign biopsy, and reproductive factors are the main parameters currently used to determine risk. The most powerful risk factor, mammographic density, is not presently employed routinely. Other potentially important factors are plasma estrogen and androgen levels, bone density, weight gain, age of menopause, and fracture history, which are also not currently used in a comprehensive risk prediction model because of lack of prospective validation. The Breast Cancer Prevention Collaborative Group (BCPCG) met to critically examine and prioritize risk factors that might be selected for further testing by multivariate analysis using existing clinical material. The BCPCG reached a consensus that quantitative breast density, state of the art plasma estrogen and androgen measurements, history of fracture and height loss, BMI, and waist–hip ratio had sufficient priority for further testing. As a practical approach, these parameters could be added to the existing Tyrer–Cuzick model which encompasses factors included in both the Claus and Gail models. The BCPCG analyzed potentially available clinical material from previous prospective studies and determined that a large case/control study to evaluate these new factors might be feasible at this time.
**Introduction**

Investigative studies and clinical trials over the past two decades have improved the recurrence free and overall survival rates of breast cancer patients by focusing on early diagnosis and improved treatment strategies. However, patients and their physicians have increasingly recognized the substantial cost and emotional burden resulting from the diagnosis of breast cancer and its treatment. Even if a tumor is diagnosed at an early stage, women must undergo surgery, followed usually by radiotherapy and adjuvant chemo- or endocrine therapy. These considerations have highlighted the importance of breast cancer prevention.

Six randomized, controlled trials have established the efficacy of selective estrogen receptor modulators (SERMs) such as tamoxifen or raloxifene to prevent breast cancer (Cuzick et al. 2003, Vogel et al. 2006). However, a low percentage of candidates for these prevention strategies choose this option after considering existing risk/benefit statistics (Port et al. 2001, Bober et al. 2004, Melnikow et al. 2005, Nekhlyudov et al. 2005, Paterniti et al. 2005). An important factor in the prevention of any disease is the number of patients who will require treatment to prevent one case (i.e. NNT, number needed to treat; Gail et al. 1999). In the NSABP-P1 tamoxifen prevention trial, ~1 in 56 women taking this agent for 5 years experienced prevention of an invasive breast cancer over that period and during 7 years of follow-up (Fisher et al. 2005). Another factor is the number of women who will experience side effects or toxicity from the preventative agent. Adverse effects in the NSABP-P1 trial included endometrial cancer in ~1 of 91 women and veno-thrombotic events (deep venous thrombosis and pulmonary embolism) in ~1 per 217. Analyzed in this way, the benefit to risk equation for the majority of women is not sufficient for them to choose tamoxifen for prevention (Paterniti et al. 2005). Three studies indicated that only 5, 18, and 51% of eligible women choose the tamoxifen prevention strategy (Port et al. 2001, Bober et al. 2004, Melnikow et al. 2005). The results of the STAR (Study of Tamoxifen and Raloxifene) trial have recently been reported and demonstrated that raloxifene produced a reduction in invasive breast cancer that was similar to tamoxifen but with less impact on non-invasive breast cancer (Vogel et al. 2006). Raloxifene was associated with a lower incidence of thromboembolic events and endometrial cancers. Overall, there did not appear to be a difference in quality of life (Land et al. 2006). Whether there will be a better acceptance of raloxifene as a preventative measure remains to be seen.

A logical conclusion regarding acceptance of the SERMs (and other potential hormonal preventative agents such as aromatase inhibitors) is that the benefit/risk ratio must be substantially improved before women will elect to accept prevention strategies. This ratio can be improved by developing more effective and safer preventative agents which is an emphasis of the current research. A parallel and equally important approach is to identify women at higher risk of developing breast cancer as candidates for prevention. To accomplish this, a major requirement is to improve the power of risk prediction methodology. While valid methods are currently available for risk assessment (Freedman et al. 2005, Costantino et al. 1999), several factors are not included in these models which potentially could enhance the power and identify women at very high risk of breast cancer. The major factors missing are measurement of mammographic density, plasma androgens and estrogens, bone density and BMI, and a history of weight gain, age of menopause, and fracture. Recent studies have included breast density and BMI to enhance the power of risk prediction but these factors have not yet been generally applied (Chen et al. 2006a).

A group of interested investigators met in St Gallen, Switzerland to critically analyze new risk factors for potential use in more powerful risk prediction models for use in postmenopausal women. The reason for limitation to postmenopausal women was to facilitate use of hormone measurements and to include only women at higher risk of breast cancer because of age. This team critically analyzed, prioritized, and selected risk factors that could be further examined by multivariate analysis in future studies. To determine if such an analysis would be feasible using existing samples from the previously conducted, randomized, prospective trials, they identified such trials and conducted a preliminary analysis to determine what patient samples are currently available. At the meeting, the panel agreed to form an ongoing group called the Breast Cancer Prevention Collaborative Group or BCPCG.

**Critical assessment of breast cancer risk factors: rationale for approach**

The BCPCG adopted a pragmatic approach to the critical analysis of various components which could be added as new factors for development of an improved risk prediction model. Those factors with potential for high predictive power but lacking general acceptance because of methodological problems, incomplete data,
Mammographic density

Background

The radiological appearance of breast tissue varies among individuals because of differences in cellular distribution of adipocytes, stroma and epithelium, and the X-ray attenuation properties of each component (Johns & Yaffe 1987). Fat appears dark on a mammogram, while epithelium and stroma appear light or white, an appearance referred to as ‘mammographic density’. No commonly accepted, standardized method exists for classifying these variations in the radiological appearance of breast tissue and both qualitative and quantitative approaches are in use (Boyd et al. 2007).

Review of methodology

In 1976 Wolfe proposed a system comprised of four categories: N1, in which the breast was mainly fat and risk of breast cancer was lowest, DY for breasts that were mostly dense and in which risk was highest, and P1 and P2 in which there were linear densities of different extent and in which risks were intermediate (Wolfe 1976a,b). Most well-designed studies have found that women with the P2 or DY patterns are at higher risk of breast cancer than those with P1 or N1 patterns, although risk gradients have in general been smaller than those originally described by Wolfe. Later the qualitative Breast Imaging Reporting and Data Systems (BIRADS) classification was devised with four categories: ‘extremely fatty’, ‘scattered density’, ‘heterogeneous density’, and ‘extremely dense’ (Harvey & Bovbjerg 2004). Few studies have used BIRADS to predict risk but the ‘extremely dense category’ predicts a significantly higher breast cancer incidence (Wolfe 1976a, Harvey & Bovbjerg 2004). A recent study involving 2 392 998 women reported a relative risk of 4.09 (95% confidence intervals, CI = 3.58–4.59) for those in the extremely dense category (Barlow et al. 2006).

Quantitative approaches include visual estimation by radiologists as well as planimetric and computer-assisted methods. Fifteen independent studies (ten case–control studies and five cohort or case–control studies nested in cohorts) with a total of 6274 cases of breast cancer and 11 638 controls have used quantitative approaches (Boyd et al. 1982, 1995, 2005, Brisson et al. 1982, 1984, 1989, Wolfe et al. 1987, Saftlas et al. 1991, Byrne et al. 1995, Kato et al. 1995, van Gils et al. 1998, 2000, Maskarinec & Meng 2000, Thomas et al. 2002, Ursin et al. 2003, Torres-Mijia et al. 2005; Table 1). All of them found an increased risk of breast cancer associated with more extensive density. The relative risks found in these studies varied from 1.8 to 6.0 when comparing the most and least extensive categories of density. Quantitative approaches have yielded more consistent results and larger gradients in risk than the qualitative methods. A recently published study involving 1744 white women found that those with breast density in the 75–100% category had a relative risk of breast cancer of 3.79 (Chen et al. 2006a).

Planimetric and the computer-assisted method have the advantage over radiologists’ classifications (i.e. the Wolfe or the BIRADS systems) with respect to objectivity and to the generation of a continuous rather than a categorical measure. Examples of mammograms of varying percent density and the relative risks of breast cancer associated with these categories are shown in Fig. 1. However, subjective qualitative measures may still be better for assessing changes, since the eye adjusts automatically to conformal differences. Further, these methods provide an absolute measure of the projected area of dense tissue and the total area of the breast in the mammogram. Although results are generally expressed as the percentage of dense area as a function of the total, the area of dense tissue alone is also associated with differences in risk of breast cancer. However, in the two published papers that provide risk estimates for both dense area and percent density, percent density was associated with larger gradients in risk in each (Byrne et al. 1995, Ursin et al. 2003). The computer-assisted method has been shown to be highly reproducible, with test–retest reproducibility > 0.9 in most studies (Brisson et al. 1989, Boyd et al. 1995, 2005). Based upon these considerations, the consensus favored use of the computer-assisted method of Boyd et al. (1995, 2001).
<table>
<thead>
<tr>
<th>Age (years)</th>
<th>n</th>
<th>Type of measurement</th>
<th>Partition&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Odds ratio&lt;sup&gt;b&lt;/sup&gt; (95% CI)</th>
<th>Trend&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Adjustments</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nested case–control in cohort</td>
<td>NR 1880 patients</td>
<td>Planimetry</td>
<td>0 vs ≥75%</td>
<td>4.3 (3.1–6.1)</td>
<td>Yes</td>
<td>Weight, age at birth of first child, family history, years of education, alcohol use, previous benign biopsy sample, and number of reproductive years</td>
<td>Byrne et al. 1995</td>
</tr>
<tr>
<td>Case–control</td>
<td>35–64 2152 controls 622 patients</td>
<td>Computer assisted</td>
<td>&lt;1% vs &gt;75%</td>
<td>5.2 (1.7–16.1)</td>
<td>NR</td>
<td>Age, body-mass index, age at menarche, family history, number of full-term pregnancies, menopausal status, hormone use, and age at first full-term pregnancy</td>
<td>Ursin et al. 2003</td>
</tr>
<tr>
<td>Cohort</td>
<td>40–80 443 controls 111 patients</td>
<td>Computer assisted</td>
<td>0.5% vs &gt;46%</td>
<td>3.49 (1.4–5.2)</td>
<td>Yes</td>
<td>Age, education, parity, height, and body-mass index</td>
<td>Torres-Mijia et al. 2005</td>
</tr>
<tr>
<td>Nested case–control in cohort</td>
<td>40–59 3100 controls 354 pairs</td>
<td>Estimation by observer and computer assisted</td>
<td>0 vs ≥75%</td>
<td>6.0&lt;sup&gt;c&lt;/sup&gt; (2.8–13.0)</td>
<td>Yes</td>
<td>Age, parity, age at birth of first child, weight, height, number of births, age at menarche, and family history</td>
<td>Boyd et al. 1995</td>
</tr>
<tr>
<td>Case–control</td>
<td>40–65 183 pairs</td>
<td>Estimation by three observers</td>
<td>&lt;10% vs ≥75%</td>
<td>4.0&lt;sup&gt;d&lt;/sup&gt; (2.1–7.7)</td>
<td>Yes</td>
<td>Age at birth of first child, parity, and family history</td>
<td>Boyd et al. 1982</td>
</tr>
<tr>
<td>Case–control</td>
<td>20–69 408 patients</td>
<td>Estimation by observer</td>
<td>0 vs ≥60%</td>
<td>5.4&lt;sup&gt;f&lt;/sup&gt; (2.5–11.4)</td>
<td>Yes</td>
<td>Parity, age at birth of first child, family history, age at menopause, and hormone use</td>
<td>Brisson et al. 1984</td>
</tr>
<tr>
<td></td>
<td>1021 controls</td>
<td></td>
<td></td>
<td>3.8&lt;sup&gt;g&lt;/sup&gt; (1.6–8.7)</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Age (years)</th>
<th>n</th>
<th>Type of measurement</th>
<th>Partition</th>
<th>Odds ratio (95% CI)</th>
<th>Trend</th>
<th>Adjustments</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case–control NR</td>
<td>362 patients 686 controls</td>
<td>Estimation by observer 0 vs ≥60%</td>
<td>4.4 (2.5–7.9)</td>
<td>Yes</td>
<td>Weight and height</td>
<td>Brisson et al. 1982</td>
<td></td>
</tr>
</tbody>
</table>
| Case–control 40–67 | 290 patients | Estimation by observer 0 vs ≥60% | 4.6 (2.4–8.5)
(5.5 (2.3–13.2)) | Yes | Age, parity, education, weight, and height | Brisson et al. 1989 |
| Case–control 60 (mean) | 647 pairs | Computer assisted <10% vs >50% | 1.8 (1.1–3.0) | No | Age at menarche, menopausal status, parity, age at birth of first child, family history, hormone use, and breast problems | Maskarinec & Meng 2000 |
| Case–control <50 | 547 patients 472 controls | Planimetry <26.7% vs >70.3% | 4.4 (3.0–6.7) | NR | Age and study | Thomas et al. 2002 |
| Case–control >35 | 108 patients | Computerised <5% vs >25% | 3.3 (1.5–7.2) | NR | Age, year of screening, menopausal status, and body-mass index | van Gils et al. 1998 |
| Case–control >35 | 139 patients 400 controls | Computerised <5% vs >25% | 2.9 (1.6–5.6) | NR | Age and parity | van Gils et al. 2001 |
| Case–control 30–85 | 160 pairs 197 patients | Planimetry <20% vs ≥70% Upper vs lower | 4.3 (1.8–10.4) | No | Parity | Wolfe et al. 1987 |
| Nested case–control in cohort 35–65 | 197 patients | Planimetry Upper vs lower | 3.6 (1.7–7.9) | Yes | Body-mass index, parity, and menopause | Kato et al. 1995 |
| Nested case–control in cohort 35–74 | 266 patients 521 controls | Planimetry <5% vs ≥65% | 2.1 (1.1–3.8) | Yes | Age, weight, and parity | Saftlas et al. 1991 |
| Nested case–control in cohort 35–74 | 266 patients 521 controls | Planimetry <5% vs ≥65% | 4.3 (2.1–8.8) | Yes | Age, weight, and parity | Saftlas et al. 1991 |

NR, not reported.

Categories of least and most widespread density from which odds ratios were calculated.

Significantly increased risk of breast cancer across all categories of density analysed in study.

Area of density estimated by radiologist.

Areas of density calculated by computer-assisted measurement.

Results from individual observers.

Data for homogeneous density.

Data for modular density.

Data for total density.

Data for pre menopausal patents.

Data for post menopausal participants.

Adapted with permission from Lancet Oncology, Boyd NF et al. 2005, 6, 798–808 with permission from Elsevier.
Pitfalls

Use of mammographic density for prediction of breast cancer risk has been controversial in the past because of the problems of ‘masking’ and non-reproducibility. ‘Masking’ of breast cancer by dense breast tissue confounds initial mammographic readings and is associated with an increased risk of breast cancer upon subsequent examinations. (Ma et al. 1992). Masking might inflate the risks associated with mammographic density in cohort studies, as cancers missed in the first mammogram, because of dense tissue, would eventually be detected during the subsequent follow-up. However, the increased risk associated with extensive density persists in screening studies and in cohort studies (Boyd et al. 1995). No attenuation of risk occurs for a period of follow-up for at least 7–10 years (Boyd et al. 1995, 2001, 2005, Byrne et al. 1995). Critical review of these data led to the consensus that long-term breast cancer risk associated with high breast density cannot be explained by masking. The problem of non-reproducibility was considered to be less important by the group since the advent of computerized techniques, a conclusion based upon analysis of multiple studies.

Choice of methodology

The group concluded that accurate quantitation of mammographic density should include application of a computer-aided system, as described extensively by Boyd and Yaffe (Boyd et al. 1995). Films are digitized and the area of the breast and total area of breast parenchyma (dense area) are outlined using interactive thresholding. The percentage of density is then calculated by dividing the total parenchymal area by the total breast area. Many published studies have collected mammograms from multiple centers and shown highly consistent results. These findings provide confidence that any variations that may exist between centers in the quality or technical performance of mammography is not a limiting factor in studying risk. The potential use of volumetric techniques was discussed but not considered sufficiently validated for inclusion in a model at the present time.

Plasma hormone levels and breast cancer risk

Review of data

The ability to predict breast cancer risk with plasma hormone levels has been the subject of multiple studies over the past three decades with conflicting results. (Kirschner 1977, Key et al. 2002). However, recent pooled data from the Endogenous Hormones and Breast Cancer Collaborative Group (EHBCG, Key et al. 2002) provide strong evidence that plasma hormone levels predict breast cancer risk. Nine centers participated in the original published study. Hormone levels were divided into quintiles based upon the ranges observed in each of the nine laboratories. As confirmation, very recent results were reported from the large EPIC study.

Figure 1 Top panel illustrates the appearance of mammograms in each of the six density categories. Bottom panels illustrate the relative risk of breast cancer associated with each category as reported by Boyd et al. (1995). Copyright 2005, The Endocrine Society.
based upon 677 patients developing breast cancer and 1309 controls after a blood sample collection (Kaaks et al. 2005). Methods used in the EHBCC study included assessment of hormone levels in a single sample taken from postmenopausal women who were then followed. The median time to diagnosis of breast cancer ranged from 2.0 to 12.1 years in the studies included in this analysis. Sex hormone levels were determined in 1765 women who did not develop breast cancer and in 663 who developed this disease. Conditional logistic regression was used to categorize groups into quintiles whose relative risk of breast cancer was determined. The major hormonal components of risk are shown on Fig. 2A and B.

Based upon the collaborative study, the relative risks (RRs) (and 95% CI) for developing breast cancer in women in the top quintile of each hormone level compared with the bottom quintile are listed below in rank order.

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Fifth</th>
<th>RR &amp; 95% CI</th>
<th>( \chi^2 ) for trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free estradiol</td>
<td>2.58</td>
<td>1.76–3.78</td>
<td>22.26</td>
</tr>
<tr>
<td>Non-SHBG-E2</td>
<td>2.39</td>
<td>1.62–3.54</td>
<td>30.82</td>
</tr>
<tr>
<td>Testosterone</td>
<td>2.22</td>
<td>1.59–3.10</td>
<td>24.39</td>
</tr>
<tr>
<td>Estrone</td>
<td>2.19</td>
<td>1.48–3.22</td>
<td>18.43</td>
</tr>
<tr>
<td>Androstenedione</td>
<td>2.15</td>
<td>1.44–3.21</td>
<td>11.26</td>
</tr>
<tr>
<td>DHEA</td>
<td>2.04</td>
<td>1.21–3.45</td>
<td>4.19</td>
</tr>
<tr>
<td>Total estradiol</td>
<td>2.00</td>
<td>1.47–2.71</td>
<td>0.66</td>
</tr>
<tr>
<td>Estrone sulfate</td>
<td>2.00</td>
<td>1.26–3.16</td>
<td></td>
</tr>
<tr>
<td>DHEAS</td>
<td>1.75</td>
<td>1.26–2.43</td>
<td></td>
</tr>
<tr>
<td>SHBG</td>
<td>0.66</td>
<td>0.43–1.00</td>
<td></td>
</tr>
</tbody>
</table>

Total plasma estradiol correlated (i.e. correlation coefficient or R value) substantially with the other hormones measured with free E2 0.96, non-sex hormone-binding globulin (non-SHBG) E2 0.87, E1 0.59, and E1S 0.60. Correlations of estradiol with androgens were significant but weaker with testosterone 0.37, androstenedione 0.35, dehydroepiandrosterone sulfate (DHEA-S) 0.29, and DHEA 0.2. Levels of androgens (as the substrate for the aromatase enzyme) and estrogens (the product of aromatase) appeared to provide independent information according to available statistical analyses. For example, when estradiol was not adjusted for androgens, the RR associated with a doubling of hormone concentration was 1.31 (95% CI = 1.17–1.48) and when adjusted for testosterone 1.18 (95% CI = 1.04–1.34). When testosterone was unadjusted, the RR associated with a doubling of hormone concentration was 1.42 (95% CI = 1.25–1.61) and when adjusted for estradiol 1.32 (95% CI = 1.15–1.51) These data, and those in the EPIC Study (Kaaks et al. 2005), suggest that measurement of androgens in addition to estrogens will provide independent information about risk.

Problems with hormone assays

The recently published collaborative hormone study (EHBCG; Key et al. 2002) illustrated the high degree of variability of median hormone measurements among postmenopausal women when comparing one laboratory with another (Table 2). In addition, a recent review of
<table>
<thead>
<tr>
<th>Study, country</th>
<th>Study subject</th>
<th>No.</th>
<th>Estradiol (pmol/l)</th>
<th>Free estradiol (pmol/l)</th>
<th>Non-SHBG estradiol (pmol/l)</th>
<th>Estrone sulfate (pmol/l)</th>
<th>Androstenedione (pmol/l)</th>
<th>DHEA (mmol/l)</th>
<th>DHEAS (mmol/l)</th>
<th>Testosterone (pmol/l)</th>
<th>SHBG (nmol/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Columbia, MO, United States</td>
<td>Case 71</td>
<td>55.1</td>
<td>0.70</td>
<td>27.3</td>
<td>129.4</td>
<td>509</td>
<td>3.46</td>
<td>7.04</td>
<td>2776</td>
<td>0.76</td>
<td>49.4</td>
</tr>
<tr>
<td></td>
<td>Control 133</td>
<td>51.4</td>
<td>0.62</td>
<td>22.1</td>
<td>129.4</td>
<td>526</td>
<td>3.11</td>
<td>5.58</td>
<td>2204</td>
<td>0.59</td>
<td>53.4</td>
</tr>
<tr>
<td>Guernsey, UK</td>
<td>Case 61</td>
<td>45.5</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>54.0</td>
</tr>
<tr>
<td></td>
<td>Control 178</td>
<td>35.0</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>63.0</td>
</tr>
<tr>
<td>Nurses’ health study, USA</td>
<td>Case 155</td>
<td>29.4</td>
<td>0.44</td>
<td>7.2</td>
<td>114.6</td>
<td>662</td>
<td>2.16</td>
<td>7.28</td>
<td>2367</td>
<td>0.80</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Control 310</td>
<td>25.7</td>
<td>0.40</td>
<td>5.9</td>
<td>103.6</td>
<td>549</td>
<td>1.99</td>
<td>7.11</td>
<td>2136</td>
<td>0.76</td>
<td>n/a</td>
</tr>
<tr>
<td>NYU WHS, USA</td>
<td>Case 129</td>
<td>134.0</td>
<td>1.94</td>
<td>79.5</td>
<td>52.7</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>1.14</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Control 247</td>
<td>101.0</td>
<td>(0.84–2.15)</td>
<td>41.0</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>1.01</td>
<td>n/a</td>
</tr>
<tr>
<td>ORDET, Italy</td>
<td>Case 67</td>
<td>21.9</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>39.8</td>
</tr>
<tr>
<td></td>
<td>Control 264</td>
<td>21.7</td>
<td>(15.6–30.0)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>41.5</td>
</tr>
<tr>
<td>Rancho Bernardo, USA</td>
<td>Case 31</td>
<td>36.7</td>
<td>(25.7–58.7)</td>
<td>n/a</td>
<td>n/a</td>
<td>114.6</td>
<td>1.82</td>
<td>7.28</td>
<td>2367</td>
<td>0.80</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Control 287</td>
<td>40.4</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>107.3</td>
<td>1.94</td>
<td>7.28</td>
<td>2367</td>
<td>0.80</td>
<td>n/a</td>
</tr>
<tr>
<td>RERF, Japan</td>
<td>Case 23</td>
<td>63.1</td>
<td>(49.4–88.1)</td>
<td>28.2</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>1.01</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Control 45</td>
<td>64.5</td>
<td>(41.9–95.0)</td>
<td>24.7</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>51.4</td>
</tr>
<tr>
<td>SOF, USA</td>
<td>Case 97</td>
<td>29.4</td>
<td>(29.4–62.4)</td>
<td>4.8</td>
<td>88.8</td>
<td>832</td>
<td>1.54</td>
<td>7.03</td>
<td>2260</td>
<td>0.85</td>
<td>58.7</td>
</tr>
<tr>
<td></td>
<td>Control 243</td>
<td>22.0</td>
<td>(18.4–40.4)</td>
<td>3.8</td>
<td>74.0</td>
<td>475</td>
<td>1.26</td>
<td>7.03</td>
<td>2260</td>
<td>0.85</td>
<td>58.7</td>
</tr>
<tr>
<td>Washington, Country, USA</td>
<td>Case 29</td>
<td>62.4</td>
<td>(40.4–77.1)</td>
<td>n/a</td>
<td>144.2</td>
<td>n/a</td>
<td>3.11</td>
<td>7.03</td>
<td>2260</td>
<td>0.85</td>
<td>58.7</td>
</tr>
<tr>
<td></td>
<td>Control 58</td>
<td>58.7</td>
<td>(29.4–77.1)</td>
<td>0.81</td>
<td>140.5</td>
<td>n/a</td>
<td>2.74</td>
<td>6.25</td>
<td>1815</td>
<td>0.85</td>
<td>58.7</td>
</tr>
</tbody>
</table>

SHBG, sex hormone-binding globulin; DHEA, dehydroepiandrosterone; DHEAS, dehydroepiandrosterone sulfate; NYUWHs, New York University Women’s Health Study; ORDET, Study of hormones and diet in the etiology of breast tumors; RERF, Radiation effects research foundation; SOF, Study of osteoporotic fractures; n/a, hormone measurement not available in this study.

*These numbers correspond to the number of case patients and control subjects for whom estradiol measurements were obtained in each study, the numbers of case patients and control subjects for whom measurements for the other hormones were obtained varied slightly for the different hormones.
Existent literature on plasma estradiol measurements documented a five-fold range in median estradiol levels between studies as measured by radioimmunoassay methods (Fig. 3; Wang et al. 2005). The differences among assays are likely due to lack of specificity and sensitivity at low levels of estradiol, particularly when direct, non-extraction methods are used.

Comparisons among assays
Prior to the St Gallen meeting, a project to compare several assay methods was initiated by members of the BCPCG. Seven different radioimmunoassays and one GC/MS/MS assay (Lee et al. 2006) were used to measure estradiol in the same samples from 40 normal postmenopausal women. As shown in this review, median values with RIA ranged from 1.7 to 13 pg/ml whereas the GC/tandem mass spec assay yielded a median level of 3.8 pg/ml. To provide an independent, biologically based assessment of assay validity, all values were correlated with body weight. The GC mass spec assay performed best with a correlation coefficient of 0.67. Correlation coefficients using either body mass index or body weight did not differ. The GC/MS/MS assay also had the greatest sensitivity at 0.6 pg/ml with an inter- and intra-assay coefficient of correlation of 8% at a level 0.9 pg/ml. The GC/MS/MS assay correlated with the best RIA (i.e. Royal Marsden-correlation with body weight of 0.58) with an R-squared correlation of 0.87 (Fig. 4).

Consensus regarding methodology
These data favor use of the GC/MS/MS assay with its greater sensitivity than RIA and better correlation with biologic parameters such as body weight. A further advantage of the GC/MS/MS assay is that five steroids, E2, E1, E1-S, testosterone and androstenedione can be measured in a single, 1 ml sample. Other recent studies demonstrated that the GC/MS/MS technique is more sensitive and specific for measurement of testosterone as well (Wang et al. 2004). The above considerations favored GC/MS/MS for measurement of multiple steroids whereas a highly validated, plasma extraction-based RIA was considered adequate for measurement of estradiol. A decision about necessity for use of GC/MS/MS techniques could only be made after multivariate analysis demonstrates the need to measure steroids other than estradiol (which can be adequately measured with high-sensitivity radioimmunoassay methods involving extraction and purification). For use in a risk prediction model, the assays for plasma hormone levels will require standardization so that absolute ranges for each quintile can be established.

Menopausal hormone therapy
Existing data
A large body of data exists regarding risk of breast cancer from use of MHT. Recent, randomized clinical trials comparing placebo with MHT have somewhat clarified the increased risks of breast cancer. The Women’s Health Initiative (WHI) randomized trial of combined estrogen plus progestin (Rossouw et al. 2002, Chlebowski et al. 2003) and recent observational studies (Collaborative Group on Hormonal Factors in Breast Cancer 1997) provide compelling evidence that current MHT with estrogen plus progestin increases breast cancer risk and that the risk increases with duration of use. The influence of estrogen alone on breast cancer risk in postmenopausal, hysterectomized women is less clear (Allen et al. 2002, Beral & Million Women 2003, Kerlikowske et al. 2003, Anderson et al. 2004) The WHI randomized trial evaluating conjugated equine estrogen as well as some recent observational studies report no increase (Allen et al. 2002, Li et al. 2003, Anderson et al. 2004), or perhaps a decrease (Kerlikowske et al. 2003, Chen et al. 2006b, Stefanick et al. 2006), of breast cancers in women using estrogen alone for <9 years. However, other observational studies suggest that especially long estrogen use of >20 years is associated with some breast cancer increase (Colditz et al. 1995, Chen et al. 2006b). Recent studies indicate that most women do not continue MHT for more than 20 years. Accordingly, no excess risk can be currently assigned.

Figure 3 Median estradiol levels in normal postmenopausal women as reported from a review of the literature (Wang et al. 2005).
to use of estrogen alone for <9 years based upon the available data.

**Consensus opinion**

The group concluded that measurement of plasma hormones would represent a high priority for determining breast cancer risk. No data from randomized, prospective trials are available to ascertain if plasma measurements would provide predictive data in women receiving MHT. The Nurse’s health study, which is observational in type did suggest that hormonal measurements could still be useful for prediction in MHT users (*Tworoger et al. 2005*). The consensus was not strong regarding this issue, but the group concluded that a higher priority be given to measurements of plasma hormone levels in women not taking MHT. A minority favored the use of history of MHT in addition to measurements of plasma hormones.
Bone density and fracture history

Existing data

The rationale for considering bone density as a risk factor for breast cancer is that increased bone density may be a surrogate marker for enhanced estrogen exposure over a woman’s lifetime (Zhang et al. 1996, 1997, O’Brien & Caballero 1997, Lucas et al. 1998, Tavani et al. 1998, Gantry et al. 1999, Nguyen et al. 2000, Buist et al. 2001, Palmer et al. 2001, Zmuda et al. 2001, van der Klift et al. 2003, Kerlikowske et al. 2005). Placing women into quartiles of bone density predicts an increasing relative risk of breast cancer with each quartile and with a threefold increased risk from the lowest to the highest quartile in one study (Cauley et al. 1996; Fig. 5A). Height loss since age 20 could represent a surrogate marker for reduced estrogen exposure (Fig. 5B; Newcomb et al. 2001). Women losing more than 2.5 cm in height since age 20 had a RR of developing breast cancer of 0.79. The fracture history is another putative surrogate marker for reduced estrogen exposure (Fig. 5C; Newcomb et al. 2001). Women with a history of fracture within 5 years had a RR of developing breast cancer of 0.6. The majority of studies to date have found that women with fractures have a lower risk of breast cancer (Newcomb et al. 2001, Silverman et al. 2004). Further studies are needed to confirm and solidify these findings and to assess the accuracy of recall history of height loss or prior fractures.

Genetic factors

Family and twin studies indicate that genetic susceptibility is an important determinant of breast cancer risk. However, few of the genes responsible have been identified and the susceptibility alleles in the known genes are relatively rare (Fig. 6; Thompson & Easton 2004, Thompson et al. 2005; The CHEK2 (check point kinase 2) Breast Cancer Case–Control Consortium 2004). Mutations in the BRCA1 and BRCA2 genes...
confer substantial risks and genetic testing for these genes is widespread in women with a family history, but testing for them is too expensive to consider for studies in an unselected postmenopausal population. Variants in CHEK2 and ataxia telangiectasia (ATM) are associated with approximately twofold risks of breast cancer. Testing for CHEK2*1100delC would be straightforward; however, it is questionable whether a test for this variant alone would be worthwhile in this context. No common genetic variants have yet (Thompson & Easton 2004, Bonadona et al. 2005) been definitively associated with breast cancer. However, with the improvement in genotyping technologies, this may alter rapidly in the next few years.

Family history of breast cancer is a well-established risk factor for breast cancer, although it is only a weak surrogate for genetic susceptibility. Family history is included in several risk models including the Gail and Claus models (McTiernan et al. 2001). A newer, more comprehensive model developed by Tyrer et al. (2004) includes paternal as well as maternal second degree relatives with a history of breast and ovarian cancer and the ages of onset of these cancers. This model has been less extensively validated than the Gail and Claus models but performed substantially better in one prospective study (Amir et al. 2003).

Prioritization of breast cancer risk factors

As a method of determining which additional factors might be included in the validation phase of a risk prediction model, the BCPCG prioritized various factors with an NIH style priority score (1.0 = highest; 5.0 = lowest) basing these determinations on a critical review of the existing literature and issues of study design (Hankinson et al. 2004, Harvie et al. 2005; Supplementary Table 1, which can be viewed online at http://erc.endocrinology-journals.org/supplemental/). (Kelsey et al. 1993, Key et al. 2001, 2002, van den Brandt et al. 2000, Rossouw et al. 2002, Okasha et al. 2003, Harvie et al. 2003, Howell et al. 2006) This prioritization reflects the expert opinion of the members of the BCPCG and is not based on level I evidence.

High priority

Priority score 1.0
Gail Model parameters
Quantitative Breast Density
Free estradiol plasma concentrations
Parity (yes/no, number)
Age at menopause

Priority score 1.5
BMI
Weight gain ages 20–50

Priority score 2.0
Total estrogen plasma concentrations (E1 + E2 + E1S)
SHBG plasma concentrations
Testosterone plasma concentrations
Paternal history of breast cancer
Second degree relatives with breast cancer
Age of onset of breast cancer in family member
Ovarian cancer and age of onset
Familial history of breast cancer with identification whether unilateral or bilateral

Priority score 2.5
Waist–hip ratio
History of height loss since age 20
History of fracture in the past 5 years
History of MHT * (no uniform consensus on this factor)

Low priority

Priority score > 2.5
Alcohol use
Years of breast feeding
BR Ca 1 and 2 status
(only 5% of population; expensive)
Bone density
Prolactin plasma concentrations
IGF plasma concentrations
Linear HT-E use alone
Exercise
Therapeutic abortion
CHEK 2
MHT E + P or E alone ever/never
Early for gestational dates
High fat diet
Oral contraceptive use
Antioxidant use
Estradiol metabolite plasma concentrations

There were several reasons to define certain factors as low priority. For some there is limited, if any, evidence of an important association with breast cancer (e.g., therapeutic abortion). For others (e.g., physical activity and alcohol intake), the associations, although relatively well-confirmed, are modest and it was deemed unlikely that a sufficient number of studies had information collected in a sufficiently detailed and comparable manner to be useable in steps 1 and 2. Similarly, few studies have assayed circulating prolactin levels to date and few have data available on bone density. For the design of a future study, the current use of MHT as a risk factor would only be assessed in a subset to determine if plasma hormone levels and MHT would provide independent risk information. This decision was based upon the fact that in available studies of plasma estrogens and androgens in relation to breast cancer, none of the women were current users of postmenopausal hormones. For risk factors where evidence of an association is quite consistent, we will consider collecting and including this information if a new prospective study is developed.

Methodology for data collection
Application of any risk prediction model depends upon the methods utilized for data collection and analysis. The BCPCG evaluated various methods and recommended those described below for use.

Bone density
Practical considerations suggest that bone density measurements might be difficult to obtain in all postmenopausal women (as opposed to the nearly routine performance of mammography). As a surrogate for measuring bone density, a history of fracture or of height loss since age 20 would be easily obtained in a questionnaire.

Family history parameters
The more comprehensive family history elements included in the Tyrer et al. (2004) model are considered most appropriate for use in a proposed study. Specifically, these elements include history of breast or ovarian cancer in paternal grandmother or aunts and maternal grandmother and aunts.

Study design steps
The BCPCG outlined several steps that might be used in the future design and validation of a new breast cancer risk prediction model. Step I (the ‘training set’) would utilize clinical material available from either the placebo arms of cohort or prior randomized, controlled trials or ongoing prospective observational studies in which patients had been followed for a period of up to 15 years after data collection. This would allow both univariate and multivariate analyses of risk factors in women observed over a 5–15-year period in the setting of a randomized controlled trial. A nested case–control format would provide the most efficient study design. Although not discussed, Step II would later involve a retrospective ‘validation study’ and Step III, a ‘prospective validation’ study.

Power analysis
Sufficient power for multivariate analysis would require ~1000 breast cancer cases and 3000 controls. This provides adequate power to detect an independent increase in risk of 40% in a factor occurring in 10% of the population in a multivariate model, which also includes other standard factors or an increase of 25% in a factor occurring in 30% of the population, and a factor which independently doubled the risk would only need to be present in 4% of the population to be detectible. This estimate is consistent with the data obtained from the collaborative study on sex hormones and breast cancer, which included 663 women with breast cancer and 1765 controls. This study was sufficiently large for univariate analysis and covariate adjustments. Accordingly, a substantially larger study is required for the multivariate analysis proposed. Since estimates of the degree of dependence and independence of each factor are not currently known, the confidence limits of our power analysis cannot be precisely determined.
Existing data sets

Prior to the BCPCG meeting, a spreadsheet was developed to identify existent randomized, controlled trials or prospective observational studies that enrolled postmenopausal women and had mammograms, blood samples and demographic data available as part of the study design. Additional contacts after the meeting identified the potential feasibility of utilizing these sources in a collaborative arrangement. The total number of potential cases was determined to be 4097 and the number of controls sufficient to match 3 controls per case (Supplementary Table 1). From this pool of cases and controls, it should be feasible to obtain all necessary mammograms, blood samples, and demographic information on 1000 cases and 3000 controls. Breast cancer cases were defined as those women who developed breast cancer over a 5–10 year period of observation after study entry and Controls as those women who did not develop breast cancer.

Potential study design based upon critical assessment of risk factors

The principle components of a study to develop a comprehensive breast cancer risk model, as determined by the BCPCG are outlined in Table 3.

Discussion

The specific goal of developing a more powerful risk prediction model is to identify a subgroup of women with a minimum risk of breast cancer over a 10-year period of ~15%. Selection on this basis would lower the NNT to prevent one breast cancer substantially from 56, as observed in the NSABP P1 trial, to ~13 (Fisher et al. 2005). Increased efficacy of preventative agents would further improve the risk/benefit ratio. For example, the aromatase inhibitors appear to be 50% more effective than tamoxifen in preventing contralateral breast cancer. If these data can be confirmed in the primary prevention setting, then the NNT would fall further to the range of 9. These NNT estimates may be too conservative since the lifetime risk of breast cancer may be reduced substantially more than the reduction over a 10-year period.

Several concerns might be raised about the design of a future study based on the principles outlined in the critical review. The major one is whether quantitative assessment of breast density in local mammography units is practical and will be widely accepted. As breast density provides a powerful means of predicting breast cancer, the BCPCG raised the concept of considering ‘breast hyper-density’ as a condition analogous to hypertension or hyperlipidemia. With increased awareness of ‘breast hyper-density’, acceptance of quantitation of breast density will likely develop as did acceptance of LDL-cholesterol measurements. An additional consideration regarding breast density is that this can be a modifiable factor, increased by MHT and decreased by use of SERMs (Lasco et al. 2006).

Various risk factors for breast cancer may be interdependent and only those providing independent information need to be assessed. A recent example illustrates this point. Measurement of estrogen levels in an NSABP subgroup selected for high risk of breast cancer did not confirm an increased risk of breast cancer (Beattie et al. 2006), although in a second assessment the association with hormone was quite robust across risk groups (Eliassen et al. 2006). This suggested that women with other risk factors such as age of menarche, age of menopause, date of first birth, MHT, and prior breast biopsy may be the same as those with high estrogen and androgen levels. While this is a potential problem, the proposed study is sufficiently powered to determine, by univariate and multivariate analysis, which factors are dependent upon one another and which are not.

The opinion of the BCPCG was that the use of MHT as a risk factor should not be included in a future study, although there was no uniform agreement on this issue. This opinion was based upon the fact that endogenous plasma estrogen and androgen levels have been shown to predict breast cancer risk with substantial power. The WHI (Chen et al. 2006b, Stefanick et al. 2006) and Nurses Health Studies did not demonstrate an increased risk of breast cancer with estrogen alone when used for 5–9 years. Use of a combination of estrogen plus a progestin in the WHI study was associated with a relative risk of 1.26 but its use would be expected to somewhat confound the predictive nature of basal hormone measurements (Tworoger et al. 2005). Additionally, as indicated above, the majority of studies measured plasma estradiol and testosterone in postmenopausal women who were not current exogenous hormone users at that time. The minority opinion of the group was that a separate cohort would be needed to resolve the role of MHT with and without a progestin in high-risk women. In this subgroup, it would be determined whether or not the use of MHT and plasma hormone levels provide independent predictive information.

A major question is whether a risk prediction model, developed according to the principles outlined in this manuscript, could be practically applied in a patient population. The routine use of mammograms is the most open to question. However, the application of
An additional practical concern is the issue of insurability of women found to be at high risk of breast cancer. This issue has arisen with use of the Framingham risk prediction method for heart disease and for measurement of the BrCa1 and 2 genes in women. Much discussion has arisen around these issues and legislation may ultimately be developed to protect those found to be at high risk.

In conclusion, the members of the BCPCG are convinced that analysis of existing data offers the distinct possibility of identifying risk factors that could be used to develop better breast cancer risk prediction models. Improved models would be of value to all women for management and would be a meaningful advance for the breast cancer prevention efforts which should be of the highest level of priority in the foreseeable future. Models incorporating mammographic density in addition to factors in the Gail model have now been developed (Anderson et al. 2004, Barlow et al. 2006). These provide evidence of the increased predictive power of models utilizing mammography density. Incorporation of factors such as plasma hormone levels and prior fracture history might then improve existing models to an even greater extent.

Acknowledgements
The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

References
Beattie MS, Costantino JP, Cummings SR, Wickerham DL, Vogel VG, Dowsett M, Folkerd EJ, Willett WC, Wolmark N & Hankinson SE 2006 Endogenous sex hormones,


Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, Benichou J & Gail MH 2006a Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *Journal of the National Cancer Institute* 98 1215–1226.


Lam PB, Vacek PM, Geller BM & Muss HB 2000 The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in Vermont. *Cancer* **89** 369–375.


*Journal of Clinical Endocrinology and Metabolism* 91:3791–3797.


*Cancer Epidemiology, Biomarkers & Prevention* 10:333–338.


*Maturitas* 36:27–34.


Palmer JR, Rao RS, Adams-Campbell LL & Rosenberg L 2001 Height and breast cancer risk: results from the Black Women’s Health Study (United States). 

*Cancer Causes & Control* 12:343–348.


*Cancer* 67:2833–2838.


*JAMA* 295:1647–1657.


The CHEK2 Breast Cancer Case-Control Consortium 2004 CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9065 controls from 10 studies. 

*American Journal of Human Genetics* 74:1175–1182.


*Cancer Epidemiology, Biomarkers & Prevention* 11:565–571.

Thompson D & Easton D 2004 The genetic epidemiology of breast cancer genes. 


