Meta-Analysis of Breast Cancer Outcomes in Adjuvant Trials of Aromatase Inhibitors Versus Tamoxifen

Mitch Dowsett, Jack Cuzick, Jim Ingle, Alan Coates, John Forbes, Judith Bliss, Marc Buyse, Michael Baum, Aman Buzdar, Marco Colleoni, Charles Coombes, Claire Snowdon, Michael Grant, Raimund Jakesz, Manfred Kaufmann, Francesco Boccadoro, Jon Godwin, Christina Davies, and Richard Peto

ABSTRACT

Purpose
To conduct meta-analyses of randomized trials of aromatase inhibitors (AIs) compared with tamoxifen either as initial monotherapy (cohort 1) or after 2 to 3 years of tamoxifen (cohort 2).

Materials and Methods
Data submitted to the Early Breast Cancer Trialists’ Collaborative Group were used in separate meta-analyses of two cohorts. Primary analyses involve postmenopausal women with tumors reported to be estrogen receptor positive. Log-rank P values are two-sided.

Results
Cohort 1 comprised 9,856 patients with a mean of 5.8 years of follow-up. At 5 years, AI therapy was associated with an absolute 2.9% (SE = 0.7%) decrease in recurrence (9.6% for AI v 12.6% for tamoxifen; 2P < .00001) and a nonsignificant absolute 1.1% (SE = 0.5%) decrease in breast cancer mortality (4.8% for AI v 5.9% for tamoxifen; 2P = .1). Cohort 2 comprised 9,015 patients with a mean of 3.9 years of follow-up. At 3 years from treatment divergence (ie, approximately 5 years after starting hormonal treatment), AI therapy was associated with an absolute 3.1% (SE = 0.6%) decrease in recurrence (5.0% for AI v 8.1% for tamoxifen since divergence; 2P < .00001) and an absolute 0.7% (SE = 0.3%) decrease in breast cancer mortality (1.7% for AI v 2.4% for tamoxifen since divergence; 2P = .02). There was no convincing heterogeneity in the proportional recurrence reduction with respect to age, nodal status, tumor grade, or progesterone receptor status and no indication of an increase in nonbreast deaths with AIs in either cohort.

Conclusion
AIs produce significantly lower recurrence rates compared with tamoxifen, either as initial monotherapy or after 2 to 3 years of tamoxifen. Additional follow-up will provide clearer information on long-term survival.

INTRODUCTION

Tamoxifen is a selective estrogen receptor (ER) modulator that acts (along with its metabolites) largely or wholly by competitive binding to the receptor protein. In the 75% to 80% of patients with early breast cancer who have ER-positive disease, treatment with 5 years of tamoxifen immediately and substantially reduces local, contralateral, and distant recurrence rates and reduces 15-year breast cancer mortality, with little effect on the aggregate of all other causes of death. However, there is little or no effect on breast cancer outcomes in ER-negative disease. The early results from the first clinical trials of aromatase inhibitors showed only a small absolute gain in 5-year survival, but the additional mortality reductions during years 5 to 9 and 10 to 14 were each about as great as the mortality reduction during years 0 to 4 (the 5 years while tamoxifen was still being taken). Hence, the absolute reduction in breast cancer mortality produced by 5 years of tamoxifen was almost three times as great 15 years after diagnosis as it had been only 5 years after diagnosis.

The development of well-tolerated inhibitors of aromatase, the enzyme that synthesizes estrogens from androgens, has provided an alternative means of depriving breast tumors of stimulation by endogenous estrogens in postmenopausal women whose ovaries no longer produce significant amounts of the hormone and in premenopausal women in whom ovarian function has been suppressed or the ovaries have been removed. In recent years, several large randomized trials in early breast cancer with 5
years of tamoxifen as the control arm have reported on the use of one of three third-generation aromatase inhibitors (AIs)—anastrozole, letrozole, or exemestane; other large third-generation AI trials have yet to be reported.\(^5\) Although the trials considered in these meta-analyses have generally found reduced recurrence rates compared with 5 years of tamoxifen, none on its own has yet clearly demonstrated a reduction in breast cancer mortality, and none possesses adequate statistical power to do so or to yield reliable subgroup analyses of the effects on recurrence. Therefore, the trialists have collaborated through the Oxford Early Breast Cancer Trialists’ Collaborative Group (EBCTCG) overview process to pool data from all currently available trials of AI versus tamoxifen, using the methods of previous EBCTCG meta-analyses. When results from individual trial publications are amalgamated in a systematic review, no allowance can be made for the different duration of follow-up in each study or for the differing definitions of some end points, and subgroup analyses may be reported inconsistently. For example, the primary end point of most trials in this meta-analysis was disease-free survival but variably included or excluded occurrence of a new nonbreast cancer. Individual patient meta-analysis as carried out here allows each of these issues to be addressed.

The trials have been grouped into two cohorts, and these cohorts have been analyzed separately. Cohort 1 consists of the trials comparing 5 years of an AI with 5 years of tamoxifen, both starting soon after surgery,\(^6,7\) and cohort 2 consists of the trials in which tamoxifen therapy is switched after 2 to 3 years to 2 to 3 years of an AI compared with 2 to 3 more years of tamoxifen (making a total of 5 years of hormonal therapy in both groups).\(^8,10\) Some of the trials included a small proportion of patients with ER-negative or ER untested tumors. These have both been excluded from the present analyses because of the lack of any material effect of 5 years of tamoxifen in ER-negative disease.\(^9\) Trials comparing treatment with an AI after 5 years of tamoxifen versus no continuation of any hormonal therapy are not covered by this overview. The database for this analysis was locked before April 26, 2010 from jco.ascopubs.org and provided by Bibliothek der MedUniWien (51061) on April 26, 2010 from 149.148.140.98.

Trial identification and data handling procedures have been described previously.\(^2\) Information was sought from all trials that had started by the year 2000 in which two arms differed for some period of time only in one having an AI and the other having tamoxifen. The only two such trials that are known to be unavailable are, first, the Austrian Breast and Colorectal Cancer Study Group (ABCSG) XII trial, which started in 1999 and randomly assigned 1,803 premenopausal women, whose ovarian function was being suppressed by goserelin, to 3 years of AI (anastrozole) or 3 years of tamoxifen and, second, the two switching arms from the Breast International Group (BIG) 01-98/ Internation Breast Cancer Study Group (IBCSG) 18-98 trial, which started in 1998. This trial randomly assigned 6,193 women four ways (between 5 years of AI [letrozole], 5 years of tamoxifen, switching from tamoxifen to AI, and switching from AI to tamoxifen); 3,095 of women were randomly assigned to the two subdivisions into years 0 to 1 (first 2 years of treatment), 2 to 4 (last 3 years of treatment), and 5+ (after treatment). For cohort 2, follow-up is subdivided into years 0 to 2 (the 3 years of randomly assigned treatment) and 3+ (after treatment) after the treatments were to diverge; these divisions equate to years 2 to 4 and 5+, respectively, from the start of adjuvant hormonal treatment because of the initial 2 years of treatment of all patients in this cohort with tamoxifen. For both cohorts, Kaplan-Meier graphs are plotted to 8 years from the start of adjuvant hormonal treatment. The lock date for the data analysis was September 30, 2006 for all trials.

### MATERIALS AND METHODS

The statistical methods and the format of presentation of the results are as described in the most recent EBCTCG report,\(^2\) with minor amendments. One amendment is to the format of the forest plots in which squares and horizontal lines are used to describe event rate ratios (AI v tamoxifen) in various subgroups. As before, a solid vertical line is used to denote no effect (ie, an event rate ratio of 1.0), and a broken vertical line is used to denote the average treatment effect in a meta-analysis of all the available evidence. In the present report, the broken vertical line is made particularly bold in the figures that present multiple subgroup analyses. This is to help emphasize tests of heterogeneity between subgroups, thereby reducing (but not eliminating) the relevance of tests of the null hypothesis within particular subgroups.\(^11,12\) Another amendment is to present, among other things, analyses of time to first distant recurrence, ignoring any prior locoregional or contralateral recurrences, as well as analyses of various types of recurrence as a first event. (Recurrence was defined as the first reappearance of breast cancer at any site and thus included second primary breast cancer or locoregional or distant recurrence of the original cancer.) As before,\(^2\) log-rank analyses are presented for any death, death without recurrence, and, by log-rank subtraction, death with recurrence. Because the incidence of recurrence was monitored carefully in these trials, the final amendment is that deaths from an unknown cause without any recorded recurrence are taken as non-breast cancer deaths.

Analyses are presented entirely separately for the trials of approximately 5 years of AI versus the same duration of tamoxifen (cohort 1) and for the switching trials of some years of tamoxifen and then some years of AI versus the same duration of tamoxifen (cohort 2). In the switching trials, the analyses are of time since the treatments diverged, which is approximately 2 to 3 years after originally starting adjuvant tamoxifen shortly after diagnosis. All but one of the switching trials randomly assigned patients after the first few years of tamoxifen (to switch or continue). The remaining trial (ABCSG VIII trial)\(^8\) randomly assigned patients at the start of adjuvant tamoxifen between 2 years of tamoxifen and then 3 years of AI and, as control, 5 years of tamoxifen. The analyses of this one trial are of time since the treatments would have diverged, and any patients who experienced recurrence, death, or loss to follow-up within the first 2 years are thus excluded from numerators, denominators, and all analyses.

Keeping to the convention of previous EBCTCG analyses, the first year of adjuvant treatment is referred to as year 1, to 2 years after the treatments were to diverge; these divisions equate to years 2 to 4 and 5+, respectively, from the start of adjuvant hormonal treatment because of the initial 2 years of treatment of all patients in this cohort with tamoxifen. For both cohorts, Kaplan-Meier graphs are plotted to 8 years from the start of adjuvant hormonal treatment. The lock date for the data analysis was September 30, 2006 for all trials.

### RESULTS

#### Cohort 1: 5 Years of AI Versus 5 Years of Tamoxifen

Two trials (Arimidex, Tamoxifen, Alone or in Combination [ATAC] and BIG 01-98/IBCSG 18-98; Appendix Fig A1, online only) contribute data, including a total of 9,856 women and more than 55,000 woman-years of follow-up (mean, 5.8 years). Mean and median follow-up times were 7.1 and 7.9 years, respectively, for ATAC and 4.3 and 4.0 years, respectively, for BIG 01-98. Most of the recurrences thus far recorded occurred during the first 5 years, while both groups were still on active treatment (Fig 1A). Use of an AI rather than tamoxifen was associated with a highly significant 23% (SE = 5%) proportional reduction in these early recurrences (\(P < .00001;\) Appendix Fig A1A), There seemed to be greater proportional decreases in isolated local recurrence (30%; \(SE = 10\%\); \(2P = .003;\) Appendix Fig A1B) and in contralateral disease (41%; \(SE = 12\%\); \(2P = .0009;\) Appendix Fig A1C) than in distant recurrence either as a first event (16%; \(SE = 6\%\); \(2P = .009;\) Appendix Fig A1D) or at any time (18%; \(SE = 6\%\); \(2P = .002;\) Appendix Fig A1E), although this apparent
heterogeneity of effect between recurrence sites was not statistically significant ($\chi^2 = 5.0, P = .08$).

The proportional reduction in the risk of recurrence was not significantly different during the first 2 years and during the last 3 years of treatment (Fig 2A; in both periods, it is separately significant), and there seemed to be some further reduction in subsequent years, although this further reduction was not, on its own, statistically significant. At 8 years after the start of hormonal therapy, the absolute difference in recurrence risk was 3.9% (SE = 1.0%; 15.3% for AI vs 19.2% for tamoxifen).

Subgroup analyses of the recurrence results with respect to progesterone receptor (PgR) status, age, nodal status, and tumor grade are shown in Figures 2B to 2E. The global test of heterogeneity between all of these subgroups was not statistically significant ($P = .1$), so undue emphasis on any one of them may yield misleading conclusions. There was no apparent heterogeneity with respect to age, nodal status, or grade between the proportional reductions in recurrence, but there was with respect to PgR status, which was known in 8,745 patients (89%). The proportional reduction in recurrence (AI vs tamoxifen) was 40% (SE = 9%) in the 22% of patients categorized as ER positive/
PgR poor and only 17% (SE = 6%) in patients categorized as ER positive/PgR positive. Although the apparent heterogeneity between these two proportional risk reductions is formally statistically significant (2P = .02), in view of the nonsignificant global test of heterogeneity, the total number of subgroups considered with respect to this and other factors, and the PgR-associated effect being significant in one of the two trials (ATAC) but not the other (BIG 01-98/IBCSG 18-98), the apparent heterogeneity with respect to PgR status could be

![Fig 2. Subgroup analyses of recurrences by (A) time period; (B) progesterone receptor (PgR) status; (C) entry age; (D) nodal status; and (E) tumor grade, for estrogen receptor–positive patients in trials of approximately 5 years of aromatase inhibitor (AI) versus tamoxifen. Dashed vertical line corresponds to zero heterogeneity of effect. NS, not significant; unspec., unspecified; O, observed; E, expected.](image)

### Table: Subgroup Analyses of Recurrences

<table>
<thead>
<tr>
<th>Category</th>
<th>Events/Women Allocated</th>
<th>Allocated tamoxifen</th>
<th>All events</th>
<th>Log-rank O–E</th>
<th>Variance ratio of annual event rates</th>
<th>Ratio of annual event rates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(A) Time period (trend χ²₁, 2P &gt; .1, NS)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years 0–1</td>
<td>163/4,954 (3.3%)</td>
<td>234/4,902 (4.8%)</td>
<td>–38.4</td>
<td>96.6</td>
<td>0.67 (SE 0.08)</td>
<td></td>
</tr>
<tr>
<td>Years 2–4</td>
<td>261/4,670 (5.6%)</td>
<td>307/4,557 (6.7%)</td>
<td>–29.5</td>
<td>137.9</td>
<td>0.81 (SE 0.08)</td>
<td></td>
</tr>
<tr>
<td>Years 5+</td>
<td>160/2,939 (5.4%)</td>
<td>180/2,803 (6.4%)</td>
<td>–15.7</td>
<td>83.0</td>
<td>0.83 (SE 0.10)</td>
<td></td>
</tr>
<tr>
<td><strong>(B) PgR status (χ²₁, 5.6, 2P = .02)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PgR poor</td>
<td>118/958 (12.3%)</td>
<td>182/937 (19.4%)</td>
<td>–36.9</td>
<td>72.5</td>
<td>0.60 (SE 0.09)</td>
<td></td>
</tr>
<tr>
<td>PgR+</td>
<td>381/3,452 (11.0%)</td>
<td>443/3,398 (13.0%)</td>
<td>–37.3</td>
<td>201.8</td>
<td>0.83 (SE 0.06)</td>
<td></td>
</tr>
<tr>
<td>PgR unknown</td>
<td>85/544 (15.6%)</td>
<td>96/567 (16.9%)</td>
<td>–4.4</td>
<td>43.6</td>
<td>0.90 (SE 0.14)</td>
<td></td>
</tr>
<tr>
<td><strong>(C) Entry age (trend χ²₁, 0.4, 2P &gt; .1, NS)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 50</td>
<td>25/225 (11.1%)</td>
<td>33/202 (16.3%)</td>
<td>–5.9</td>
<td>13.9</td>
<td>0.65 (SE 0.22)</td>
<td></td>
</tr>
<tr>
<td>50–59</td>
<td>175/1,680 (10.4%)</td>
<td>228/1,764 (13.6%)</td>
<td>–30.0</td>
<td>98.3</td>
<td>0.74 (SE 0.09)</td>
<td></td>
</tr>
<tr>
<td>60–69</td>
<td>222/1,898 (11.3%)</td>
<td>266/1,904 (14.0%)</td>
<td>–26.4</td>
<td>118.8</td>
<td>0.80 (SE 0.08)</td>
<td></td>
</tr>
<tr>
<td>70+</td>
<td>162/1,151 (14.1%)</td>
<td>194/1,122 (17.3%)</td>
<td>–21.1</td>
<td>86.1</td>
<td>0.78 (SE 0.10)</td>
<td></td>
</tr>
<tr>
<td><strong>(D) Nodal status (N0/N− v all N+ χ²₁, 0.6, 2P &gt; .1, NS)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0/N−</td>
<td>118/1,937 (6.1%)</td>
<td>171/2,021 (8.5%)</td>
<td>–23.9</td>
<td>71.2</td>
<td>0.72 (SE 0.10)</td>
<td></td>
</tr>
<tr>
<td>N1–3</td>
<td>83/734 (11.3%)</td>
<td>91/702 (13.0%)</td>
<td>–6.9</td>
<td>42.3</td>
<td>0.85 (SE 0.14)</td>
<td></td>
</tr>
<tr>
<td>N4+</td>
<td>324/1,866 (17.4%)</td>
<td>372/1,779 (20.9%)</td>
<td>–36.8</td>
<td>167.4</td>
<td>0.80 (SE 0.07)</td>
<td></td>
</tr>
<tr>
<td>N+ unspec.</td>
<td>45/254 (17.7%)</td>
<td>65/244 (26.6%)</td>
<td>–12.1</td>
<td>26.5</td>
<td>0.63 (SE 0.16)</td>
<td></td>
</tr>
<tr>
<td>N unknown</td>
<td>14/163 (8.6%)</td>
<td>22/156 (14.1%)</td>
<td>–4.5</td>
<td>8.9</td>
<td>0.60 (SE 0.26)</td>
<td></td>
</tr>
<tr>
<td><strong>(E) Tumor grade (trend χ²₁, 0.3, 2P &gt; .1, NS)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well differentiated</td>
<td>63/1,186 (5.3%)</td>
<td>93/1,217 (7.6%)</td>
<td>–14.6</td>
<td>38.4</td>
<td>0.68 (SE 0.13)</td>
<td></td>
</tr>
<tr>
<td>Moderately</td>
<td>265/2,417 (11.2%)</td>
<td>331/2,388 (12.9%)</td>
<td>–40.4</td>
<td>145.2</td>
<td>0.76 (SE 0.07)</td>
<td></td>
</tr>
<tr>
<td>Poorly</td>
<td>181/854 (21.2%)</td>
<td>209/823 (25.4%)</td>
<td>–23.8</td>
<td>92.1</td>
<td>0.77 (SE 0.09)</td>
<td></td>
</tr>
<tr>
<td>Grade unknown</td>
<td>75/497 (15.1%)</td>
<td>88/474 (18.6%)</td>
<td>–9.8</td>
<td>39.3</td>
<td>0.78 (SE 0.14)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>584/4,954 (11.8%)</td>
<td>721/4,902 (14.7%)</td>
<td>–83.6</td>
<td>317.5</td>
<td>0.769 (SE 0.049)</td>
<td>2P &lt; .00001</td>
</tr>
</tbody>
</table>

99% or 95% confidence intervals
largely or wholly a chance finding. If so, the overall proportional reduction in recurrence rate (AI vs tamoxifen) may provide the best guide to the treatment effect in each of the subgroups in Figures 2B to 2E.

Thus far, there has been only a small and nonsignificant proportional reduction (11%; SE = 7%) in breast cancer mortality in cohort 1 (Figs 1B and 3A). The absolute difference between AI and tamoxifen in breast cancer mortality was 1.1% (SE = 0.5%) at 5 years. Although it seemed to be smaller (only 0.5%; SE = 0.8%) at 8 years (Fig 1B), the mean duration of follow-up is only 5 years, so this could change with the accumulation of further data. Of the 1,261 deaths from any cause, 559 (44%) were of women who had not experienced a recurrence and had died of an unrelated cause. There was no apparent difference between AI and tamoxifen in these non-breast cancer deaths (Figs 1C and 3B). Hence, the difference in overall mortality was also nonsignificant (Figs 1D and 3C).

**Cohort 2: 2 to 3 Years of Tamoxifen for All Women and Then 2 to 3 Years of AI Versus Tamoxifen**

Four trials (German Adjuvant Breast Cancer Group/Arimidex-Nolvadex, Intergroup Exemestane Study/BIG 02-97, Italian Tamoxifen Anastrozole Trial (ITA), and ABCSG VIII) contribute data, including a total of 9,015 women with more than 33,000 woman-years of follow-up after the time of switching (mean, 3.9 years); mean and median follow-up times were 6.1 and 6.0 years. 4.4 and 4.4 years, 5.9 and 6.0 years, and 3.0 and 2.9 years, respectively. The analyses of recurrence and death are calculated from the time the allocated treatments would differ. (The time from diagnosis is indicated in the graphs as being approximately 2 years greater than this.)

There was an overall 29% (SE = 6%) proportional reduction in recurrence, with no significant heterogeneity between the contributing trials (Appendix Fig A2A, online only). The absolute cumulative risk of recurrence was reduced by 3.1% (SE = 0.6%) 3 years after switching (ie, approximately 5 years after diagnosis) and by 3.6% (SE = 1.1%) 6 years after switching (ie, approximately 8 years after diagnosis; Fig 4A). There was a significantly greater risk reduction during the first 3 years after switching (ie, while treatment with AI or tamoxifen was continuing) than during subsequent years (ie, after cessation of treatment; 2P = .003 for heterogeneity; Fig 5A). Again, there seemed to be greater proportional reductions in isolated local recurrence (40%; SE = 13%; 2P = .002; Appendix Fig A2B) and contralateral disease (35%; SE = 17%; 2P = .03; Appendix Fig A2C) than in distant recurrence either as the first event (24%; SE = 7%; 2P = .001; Appendix Fig A2D) or at any time (24%; SE = 7%; 2P = .0007; Appendix Fig A2E), but this apparent heterogeneity of effect between different sites of recurrence was not significant.

Subgroup analyses of the recurrence results with respect to PgR status, age, nodal status, and tumor grade are shown in Figures 5B to 5E. The global test of heterogeneity for these four subgroup analyses was not significant. There was no apparent heterogeneity with respect to any of the four subgroups in the proportional reductions in recurrence. PgR status was known in 8,184 patients (91%). The proportional reduction in recurrence was 37% (SE = 12%) in women with ER-positive/PgR-poor disease and 21% (SE = 8%) in women with ER-positive/PgR-positive disease, but these proportional risk reductions are not significantly different from each other.

There was a significant reduction in breast cancer mortality (proportional reduction in mortality with recurrence = 22%; SE = 9%; 2P = .02; Fig 6A). The absolute difference in breast cancer mortality was 0.7% (SE = 0.6%) approximately 5 years after diagnosis and 1.7% (SE = 0.8%) approximately 8 years after diagnosis (Figs 4B and 6A), but the latter result in particular may change somewhat when more women have been observed to this
point. Of the 638 deaths from any cause, 276 (43%) were of women who had not experienced a recurrence and had died of an unrelated cause. There was a nonsignificant reduction in such deaths (Figs 4C and 6B). The absolute difference in overall mortality was 1.1% (SE 0.5%) approximately 5 years after diagnosis (Figs 4D and 6C). It seemed to grow larger over the next few years and was 2.2% (SE 1.0%; 2P = .004) approximately 8 years after diagnosis, but this too may change with further follow-up in these trials and with the availability of more trials.

**DISCUSSION**

Each cohort in this overview included between 9,000 and 10,000 patients, and in aggregate, the two cohorts included almost 19,000

---

**Fig 4.** Life-table curves of (A) recurrence; (B) breast cancer mortality; (C) death without recurrence; and (D) any death, for estrogen receptor–positive patients in trials of 2 to 3 years of tamoxifen and then 2 to 3 years of aromatase inhibitor (AI) versus tamoxifen. O, observed; E, expected; V, variance.
patients in only six trials. These relatively large numbers per trial reflected the expectation of only a small incremental improvement in recurrence-free survival given the relatively good prognosis of tamoxifen-treated, ER-positive patients. Overview analysis remains of value for the characterization of the time dependence of the effects on recurrence; for the eventual effects on 5, 10, and 15-year survival; for subgroup analyses; and for identification of any uncommon favorable or unfavorable adverse effects (not addressed here).

<table>
<thead>
<tr>
<th>Category</th>
<th>Events/Women</th>
<th>Allocated AI</th>
<th>Allocated tamoxifen</th>
<th>Log-rank O−E</th>
<th>Variance of O−E</th>
<th>Ratio of annual event rates Al : Tamoxifen</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Time since treatments differed* (trend χ² = 8.7, P = .003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years 0–2 (≈2–4)</td>
<td>187/4,508 (4.1%)</td>
<td>303/4,507 (6.7%)</td>
<td>–61.0</td>
<td>118.4</td>
<td>0.60 (SE 0.07)</td>
<td></td>
</tr>
<tr>
<td>Years 3+ (≈5+)</td>
<td>170/2,978 (5.7%)</td>
<td>172/2,854 (6.2%)</td>
<td>–7.9</td>
<td>84.0</td>
<td>0.91 (SE 0.10)</td>
<td></td>
</tr>
<tr>
<td>(B) PgR status (χ² = 1.6, 2P &gt; .1, NS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PgR poor</td>
<td>64/704 (9.1%)</td>
<td>107/764 (14.0%)</td>
<td>–19.1</td>
<td>41.1</td>
<td>0.63 (SE 0.12)</td>
<td></td>
</tr>
<tr>
<td>PgR+</td>
<td>240/3,375 (7.1%)</td>
<td>294/3,341 (8.8%)</td>
<td>–30.7</td>
<td>130.5</td>
<td>0.79 (SE 0.08)</td>
<td></td>
</tr>
<tr>
<td>PgR unknown</td>
<td>53/429 (12.4%)</td>
<td>78/402 (19.4%)</td>
<td>–16.0</td>
<td>31.5</td>
<td>0.60 (SE 0.14)</td>
<td></td>
</tr>
<tr>
<td>(C) Entry age (trend χ² = 0.0, 2P &gt; .1, NS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 50</td>
<td>12/119 (10.1%)</td>
<td>15/130 (11.5%)</td>
<td>–1.2</td>
<td>6.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–59</td>
<td>114/1,423 (8.0%)</td>
<td>141/1,393 (10.1%)</td>
<td>–17.9</td>
<td>61.8</td>
<td>0.75 (SE 0.11)</td>
<td></td>
</tr>
<tr>
<td>60–69</td>
<td>145/1,903 (7.6%)</td>
<td>216/1,869 (11.1%)</td>
<td>–38.7</td>
<td>87.6</td>
<td>0.64 (SE 0.09)</td>
<td></td>
</tr>
<tr>
<td>70+</td>
<td>86/1,063 (8.1%)</td>
<td>107/1,116 (9.6%)</td>
<td>–9.9</td>
<td>46.5</td>
<td>0.81 (SE 0.13)</td>
<td></td>
</tr>
<tr>
<td>(D) Nodal status (N0/N− v all N+ χ² = 1.0, 2P &gt; .1, NS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0/N−</td>
<td>96/2,399 (4.0%)</td>
<td>123/2,414 (5.1%)</td>
<td>–13.9</td>
<td>54.0</td>
<td>0.77 (SE 0.12)</td>
<td></td>
</tr>
<tr>
<td>N1−3</td>
<td>34/565 (6.0%)</td>
<td>55/531 (10.4%)</td>
<td>–12.3</td>
<td>21.4</td>
<td>0.56 (SE 0.16)</td>
<td></td>
</tr>
<tr>
<td>N4+</td>
<td>20/124 (16.1%)</td>
<td>39/144 (27.1%)</td>
<td>–8.0</td>
<td>13.9</td>
<td>0.56 (SE 0.20)</td>
<td></td>
</tr>
<tr>
<td>N+ unspec.</td>
<td>156/912 (17.1%)</td>
<td>207/897 (23.1%)</td>
<td>–30.9</td>
<td>86.4</td>
<td>0.70 (SE 0.09)</td>
<td></td>
</tr>
<tr>
<td>N unknown</td>
<td>51/508 (10.0%)</td>
<td>55/521 (10.6%)</td>
<td>–2.1</td>
<td>26.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E) Tumor grade (trend χ² = 0.1, 2P &gt; .1, NS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well differentiated</td>
<td>46/788 (5.8%)</td>
<td>48/786 (6.2%)</td>
<td>–1.7</td>
<td>22.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately</td>
<td>181/2,676 (6.8%)</td>
<td>248/2,703 (8.2%)</td>
<td>–37.7</td>
<td>103.6</td>
<td>0.69 (SE 0.08)</td>
<td></td>
</tr>
<tr>
<td>Poorly</td>
<td>88/529 (16.6%)</td>
<td>88/492 (17.9%)</td>
<td>–4.6</td>
<td>41.4</td>
<td>0.89 (SE 0.15)</td>
<td></td>
</tr>
<tr>
<td>Grade unknown</td>
<td>42/515 (8.2%)</td>
<td>94/526 (17.9%)</td>
<td>–25.3</td>
<td>32.5</td>
<td>0.46 (SE 0.12)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>357/4,508 (7.9%)</td>
<td>479/4,507 (10.6%)</td>
<td>–68.9</td>
<td>202.4</td>
<td>0.711 (SE 0.060)</td>
<td>2P &lt; .00001</td>
</tr>
</tbody>
</table>

Fig 5. Subgroup analyses of recurrences by (A) time period; (B) progesterone receptor (PgR) status; (C) entry age; (D) nodal status; and (E) tumor grade, for estrogen receptor–positive patients in trials of 2 to 3 years of tamoxifen and then 2 to 3 years of aromatase inhibitor (AI) versus tamoxifen. Dashed vertical line corresponds to zero heterogeneity of effect. (*) Values in parentheses are approximate time since diagnosis. O, observed; E, expected; unspec., unspecified; NS, not significant.
Most of the trials recruited patients known to be hormone receptor (ER or PgR) positive. Patients recorded as having ER-negative/PgR-positive tumors, who were included in that hormone receptor–positive definition, are commonly treated with hormone treatment, as are PgR-positive tumors, who were included in that hormone receptor–positive definition. This means that most of the statistical information excludes patients with a particularly unfavorable outcome does not. Direct comparison would be invalid because cohort 1 includes patients who would experience relapse on tamoxifen in the first 2 to 3 years after surgery and cohort 2 does not. Direct comparison would be invalid because cohort 2 excludes patients with a particularly unfavorable outcome on tamoxifen.

Cohort 1 comprises data from ATAC at 95 months of median follow-up and from BIG 1-98 at 48 months of median follow-up. Accordingly, approximately two thirds of the statistical information for the overview results in cohort 1 is provided by ATAC, and the preponderant influence of ATAC in cohort 1 is particularly marked in longer follow-up (5 to 8 years). Updated results from BIG 1-98 have recently been published elsewhere.

The use of 5 years of an AI immediately after surgery (cohort 1) further reduced recurrence by approximately 23% relative to 5 years of tamoxifen. In cohort 2, when the switch to an AI is made after 2 to 3 years of tamoxifen, there seems to be a particularly marked reduction in risk of recurrence (by 40%) during the following 3 years. There is evidence that disease relapsing after tamoxifen treatment can show substantial differences in molecular phenotype from the disease at presentation. If this occurs in subclinical disease during adjuvant therapy, it is possible that the disease may take on an enhanced sensitivity to AI treatment that could underpin this effect, but there are no substantive data to demonstrate this.

### Table: Deaths/Women

<table>
<thead>
<tr>
<th>Year code and study name</th>
<th>Treatment comparison</th>
<th>Deaths/Women</th>
<th>Allocated AI</th>
<th>Allocated tamoxifen</th>
<th>Log-rank O−E</th>
<th>Variance of O−E</th>
<th>Ratio of annual death rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Mortality with recurrence ($\chi^2 = 2.8, 2P &gt; .1, \text{NS}$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94V GABG/ARNO Germany</td>
<td>Tam 2 yr, Ana v Tam 3 yr</td>
<td>15/447</td>
<td>22/455</td>
<td>−3.4</td>
<td>9.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>96Y1 Austrian BCSG VIII</td>
<td>Tam 2 yr, Ana v Tam 3 yr a</td>
<td>15/1,820</td>
<td>24/1,807</td>
<td>−4.2</td>
<td>9.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98C IES / BIG 02-97</td>
<td>Tam 2–3 yr, Exc v Tam 3–2 yr</td>
<td>117/2,021</td>
<td>133/2,021</td>
<td>−8.6</td>
<td>60.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98R ITA Italy</td>
<td>Tam 2 yr, Ana v Tam to = 5 yr</td>
<td>12/220</td>
<td>24/224</td>
<td>−6.0</td>
<td>8.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>169/4,508</td>
<td>203/4,507</td>
<td>−22.2</td>
<td>87.7</td>
<td>0.78 (SE 0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B) Mortality without recurrence ($\chi^2 = 5.0, 2P &gt; .1, \text{NS}$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94V GABG/ARNO Germany</td>
<td>Tam 2 yr, Ana v Tam 3 yr</td>
<td>5/447</td>
<td>14/455</td>
<td>−4.5</td>
<td>4.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>96Y1 Austrian BCSG VIII</td>
<td>Tam 2 yr, Ana v Tam 3 yr a</td>
<td>56/1,820</td>
<td>56/1,807</td>
<td>−1.4</td>
<td>27.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98C IES / BIG 02-97</td>
<td>Tam 2–3 yr, Exc v Tam 3–2 yr</td>
<td>61/2,021</td>
<td>77/2,021</td>
<td>−9.1</td>
<td>34.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98R ITA Italy</td>
<td>Tam 2 yr, Ana v Tam to = 5 yr</td>
<td>5/220</td>
<td>2/224</td>
<td>1.3</td>
<td>1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>127/4,508</td>
<td>149/4,507</td>
<td>−13.7</td>
<td>68.2</td>
<td>0.82 (SE 0.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(C) Any death ($\chi^2 = 2.6, 2P &gt; .1, \text{NS}$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94V GABG/ARNO Germany</td>
<td>Tam 2 yr, Ana v Tam 3 yr</td>
<td>20/447</td>
<td>36/455</td>
<td>−7.9</td>
<td>13.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>96Y1 Austrian BCSG VIII</td>
<td>Tam 2 yr, Ana v Tam 3 yr a</td>
<td>71/1,820</td>
<td>80/1,807</td>
<td>−5.6</td>
<td>37.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98C IES / BIG 02-97</td>
<td>Tam 2–3 yr, Exc v Tam 3–2 yr</td>
<td>178/2,021</td>
<td>210/2,021</td>
<td>−17.7</td>
<td>94.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98R ITA Italy</td>
<td>Tam 2 yr, Ana v Tam to = 5 yr</td>
<td>17/220</td>
<td>26/224</td>
<td>−4.8</td>
<td>10.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>286/4,508</td>
<td>352/4,507</td>
<td>−36.0</td>
<td>156.0</td>
<td>0.79 (SE 0.07)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig 6. Trial-specific forest plot of (A) mortality with recurrence; (B) mortality without recurrence; and (C) any death, for estrogen receptor–positive patients in trials of 2 to 3 years of tamoxifen (Tam) and then 2 to 3 years of aromatase inhibitor (AI) versus tamoxifen. (*) Random allocation at start of tamoxifen. O, observed; E, expected; NS, not significant; GABG, German Adjuvant Breast Cancer Group; ARNO, Arimidex-Nolvadex; Ana, anastrozole; Austrian BCSG, Austrian Breast and Colorectal Cancer Study Group; IES, Intergroup Exemestane Study; BIG, Breast International Group; ITA, Italian Tamoxifen Analastrozole trial; Exc, exemestane.
When examined in relation to time on treatment versus time after cessation of treatment, there was some evidence for an attenuation of the reduced risk of recurrence that was stronger for cohort 2. However, for both cohorts, the CIs of the hazard ratios for the period after cessation of therapy were wide, underscoring the importance of continued follow-up and the ongoing clinical trials examining extensions to the duration of AI therapy.

The global test for heterogeneity of the subgroups was not significant in either cohort, although in cohort 1, the subset according to PgR status individually was significant. Previous trials have demonstrated that 5 years of tamoxifen is highly, and approximately equally, effective in both ER-positive/PgR-positive and ER-positive/PgR-poor disease. The current data seem to suggest that patients reported to be PgR poor derive greater benefit from an AI. However, the two trials contributing these data differ substantially in their findings according to PgR status. In addition, central analyses of PgR (as opposed to the predominately locally performed receptor analyses in the overview) suggest similar proportional treatment benefit for PgR-positive and PgR-poor populations in both trials. Thus, the apparent heterogeneity of effect with regard to PgR status in cohort 1 may well be a chance finding. Widely acknowledged difficulties in reproducing PgR status may contribute to these variable findings. Cohort 2 showed no suggestion of increased effectiveness of AIs over tamoxifen in the PgR-poor group. The overview indicates that the enhanced efficacy of AIs over tamoxifen is proportionally similar across all cohorts examined.

In the trials studying approximately 5 years of adjuvant tamoxifen versus no tamoxifen in ER-positive disease, the main effect of tamoxifen on breast cancer recurrence was seen in the first 5 years (absolute gain of 12.5% at 5 years and roughly constant thereafter), but the main effect on breast cancer mortality was seen later (absolute gain of 3.3% at 5 years and 9.1% at 15 years). In the present trials of AIs versus tamoxifen, the mean follow-up time thus far available is, in both cohorts, only about 6 years after diagnosis (mean follow-up duration after random assignment was 5.8 years in cohort 1 and 3.9 years in cohort 2), and the absolute further recurrence reduction (AI vs tamoxifen) 5 years after diagnosis was approximately 3% and highly significant in both cohorts (2.9%, SE = 0.7% in cohort 1; and 3.1%, SE = 0.6% in cohort 2). Analogous to the recurrence and mortality patterns in the tamoxifen trials suggests that the absolute additional effects (AI vs tamoxifen) on breast cancer mortality 5 years after diagnosis would be expected to be approximately 1% in both cohorts, whereas the observed differences were 1.1% (SE = 0.5%) in cohort 1 and 0.7% (SE = 0.3%) in cohort 2, but with further follow-up, the breast cancer mortality differences may increase. Limited follow-up beyond 5 years after surgery is so far available from either cohort; at this stage, the data from cohort 2 but not from cohort 1 provide evidence for such a further reduction in breast cancer mortality. As with other end points, the mortality results from cohorts 1 and 2 should not be compared directly.

There was no indication that the AIs were associated with an increase in non–breast cancer deaths. This is reassuring regarding the overall safety of AIs relative to tamoxifen. The overall mortality reductions for cohort 2 but not cohort 1 (AI vs tamoxifen) achieved statistical significance. As recently discussed, the combination of safety analyses (integrated here as non–breast cancer deaths) with efficacy analyses (breast cancer–specific deaths) to a summary measure (overall mortality) can lead to substantial loss of information by mixing important signals with irrelevant noise. Analyses of cause-specific mortality have not been conducted at this time but will be an important component of later overviews.

This overview provides clear evidence that third-generation AIs achieve modest absolute improvements in breast cancer end points with significant reductions in recurrence in both cohorts and in breast cancer–specific mortality in cohort 2. Because the proportional benefits do not vary significantly with factors affecting prognosis, the absolute gain is greater for patients with poorer prognosis. There are also significant differences in adverse effects between tamoxifen and AIs that have not been reviewed here; AI use is associated with fewer endometrial cancers and thromboembolic events than tamoxifen but with more arthralgia and fractures. The decision on whether to initiate treatment with an AI or tamoxifen or whether to switch to an AI after 2 to 3 years of tamoxifen rather than continuing with tamoxifen for 5 years depends on a careful evaluation of these factors in individual patients. This overview has provided greater confidence in the efficacy aspects of this evaluation.
REFERENCES


