CYP2D6 Genotyping and the Use of Tamoxifen in Breast Cancer

Donald Berry

Correspondence to: Donald Berry, PhD, Department of Biostatistics, University of Texas MD Anderson Cancer Center, 1400 Pressler St, 4-5062 Pickens Academic Tower, Houston, TX 77030-1402 (e-mail: dberry@mdanderson.org).

Cytochrome P450 2D6 (CYP2D6) is involved in the metabolism of many drugs, including tamoxifen (1), but the potential role of CYP2D6 genotype assessment in determining whether breast cancer patients with estrogen receptor (ER)-positive tumors should receive tamoxifen is controversial. This controversy has been played out in the Journal (2–10).

The focus of empirical studies regarding this issue has been postmenopausal disease in the adjuvant setting. In these studies, researchers have assumed that CYP2D6 genotype is unrelated to breast cancer outcome, and hence they focus on patients who have received tamoxifen, whether in a clinical trial or in clinical practice. Because tamoxifen is highly effective in delaying recurrence of ER-positive disease (11), if its effect is substantially reduced in poor metabolizers, then this would be detectable in a study with sufficiently large sample size.

In summarizing the available literature, a June 2010 Technology Assessment Report to the Agency for Healthcare Research and Quality of the US Department of Health and Human Services concluded that “there were no consistent associations between CYP2D6 polymorphisms and outcomes in tamoxifen treated breast cancer across 16 studies included in [this] systematic review” (12). This conclusion continues to be appropriate; there still does not exist credible empirical evidence showing a benefit for CYP2D6 genotyping. Indeed, two large trials published in the Journal in 2012 showed no association (2,3). The editorialists concluded, “[T]he fact that these two studies confirm each other suggests that this matter has likely been laid to rest” (4).

Some researchers objected that these two studies were fundamentally flawed for at least two reasons (5–7,13). One reason was that both studies used tumor tissue to assess CYP2D6 polymorphisms. According to Pharoah et al., “A major limitation of both new studies is the use of tumor DNA to determine germline genotype” (6). The other common complaint was the lack of Hardy–Weinberg equilibrium (HWE). For example, “A standard χ² test shows highly statistically significant deviation of these frequencies [in the Regan study (2)] from those expected under HWE (χ² = 416, P = 10⁻⁹²). Deviation of this magnitude would suggest a severe measurement error” (6).

Regarding the first point, tumors may cause deletions at CYP2D6 that result in loss of heterozygosity. Therefore, as the argument goes, to properly assess the ability of the human body to metabolize tamoxifen requires normal tissue. But it is also possible that a tumor that modifies CYP2D6 may indicate a differential response to tamoxifen, perhaps even a better response than expected otherwise. So tumor may be the most relevant of all tissues.

Not all readers will be familiar with the concept of HWE. I offer a brief explanation. Consider a gene having two alleles, A and B. Their frequencies in the population are a and b, where a + b = 1. The possible genotypes are AA, AB, and BB. Assuming
randomness in mating as regards this gene, the respective frequencies in the population will be \(a^2, 2ab, b^2\), as shown in the tree diagram in Figure 1.

The numerical relationship \(a^2, 2ab, b^2\) in Figure 1 defines HWE in this case. Under HWE, the value of any one of \(a^2, 2ab, b^2\) uniquely determines the other two. Suppose \(A\) and \(B\) are equally prevalent in the population in question—that is, \(a = b = 1/2\). The HWE genotypic frequencies are then \(1/4, 1/2,\) and \(1/4\). If \(a = 2/3\) then the frequencies are \(4/9, 4/9,\) and \(1/9\). If \(a = 0.2\), then they are \(0.04, 0.32,\) and \(0.64\).

Of course, there is variability when sampling from a population, whether the population is in HWE or not. Suppose a sample of 100 individuals shows frequencies of 28, 54, and 18 for \(AA, AB,\) and \(BB\), respectively. The number of \(A\) alleles in the sample is \(2 \times 28 + 54 = 110\) with proportion \(110/200 = 0.55\). The expected genotypic frequencies are \(0.55^2, 2(0.55)(0.45),\) and \(0.45^2\), which equal 0.3025, 0.4950, and 0.2025. These theoretical frequencies are not the same as the observed frequencies of 0.28, 0.54, and 0.18, but they are sufficiently close that HWE cannot be ruled out.

On the other hand, suppose the observed frequencies are 38, 24, and 38. There are 100 As and 100 Bs in this sample, so the best estimate of \(a\) is \(1/2\). The expected frequencies under HWE are 25, 50, and 25, which are very different from the observed. There are many possible explanations for such an observation. Perhaps homozygotes \(AA\) and \(BB\) are more likely to mate with “their own kind.” Perhaps heterozygotes \(AB\) are sickly and tend to die early. Perhaps \(A\) and \(B\) are difficult to distinguish from each other in the assessment technique, so \(AB\) is frequently read as homozygotic, either \(AA\) or \(BB\).

An early study of \(CYP2D6\) that involved 190 ER-positive breast cancer patients showed a relationship similar to this example (14). The tissue being genotyped was tumor tissue. The frequencies of genotypes depending on the \(*4\) allele vs wild-type are shown in Table 1. I have added the last column in the table based on the assumption of HWE and based on the observed proportion of \(*4\) alleles: \(66/380 = 0.174\). There were more than twice as many \(*4\) homozygotes (poor metabolizers) in this study than expected under HWE.

Possible explanations for the apparent Hardy–Weinberg disequilibrium suggested by Table 1 include those mentioned in the example above. In addition, it could be a statistical fluke. Or women who are \(*4\) homozygotes may be more likely to develop ER-positive breast cancer. Perhaps heterozygotes are protected from breast cancer. But the most likely explanation is the one discussed above—that the tumor may induce loss of heterozygosity.

Interestingly, the lack of HWE in the Regan study (2) was essentially the same as that of the Goetz study (14), but the Regan study was 20 times as large; compare Table 2 with Table 1. As many correspondents have suggested, the lack of HWE in the Regan study (2) is most unlikely to be a statistical fluke. Most interesting is that the Goetz study (14) showed a correlation between \(CYP2D6\) genotype and cancer recurrence—one that was statistically significant but weak. Because both studies had the same HWE status, the Regan study was ideal in addressing confirmation of the results of the Goetz study, with the bonus of having much greater statistical power. There were some differences in the two studies of course, but the Regan study was resoundingly clear in failing to corroborate the Goetz observation.

The Rae et al. article in this issue of the Journal is still another volley in this controversy (15). The authors compared \(CYP2D6\) genotype between tumor and normal tissue (whole blood and also, separately, unaffected lymph nodes) in 123 breast cancer patients. Concordance was very high. Additionally comforting is that this new study showed consistency in HWE regardless of tissue source for the more common \(CYP2D6\) allelic variants, although the sample size was not large and the statistical power was modest. This latest article shows that agreement is possible across tissue sources in one laboratory and in a moderate-sized study. But it does not shed light on the lack of HWE observed in earlier studies.

Does HWE matter? Assuredly yes, from a scientific point of view, but not from a practical point of view. A wide variety of techniques have been used for genotyping \(CYP2D6\), even in the context of the published studies (12). Standards and circumstances in clinical practice will be no better than they were in the published studies. In clinical practice, physicians and patients are faced with the vagaries of different laboratories and different methodologies.

![Figure 1. Tree diagram showing Hardy–Weinberg equilibrium in the simple case of two alleles A and B with population frequencies a and b with a + b = 1.](https://academic.oup.com/jnci/article-abstract/105/17/1267/910100/06-March-2019)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency (proportion)</th>
<th>Hardy–Weinberg equilibrium proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type/wild-type</td>
<td>137 (72.1%)</td>
<td>68.2%</td>
</tr>
<tr>
<td>Wild-type/*4</td>
<td>40 (21.1%)</td>
<td>28.7%</td>
</tr>
<tr>
<td>*4/*4</td>
<td>13 (6.8%)</td>
<td>3.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Proportion</th>
<th>Hardy–Weinberg equilibrium proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type/wild-type</td>
<td>70.9%</td>
<td>65.8%</td>
</tr>
<tr>
<td>Wild-type/*4</td>
<td>20.5%</td>
<td>30.6%</td>
</tr>
<tr>
<td>*4/*4</td>
<td>8.6%</td>
<td>3.6%</td>
</tr>
</tbody>
</table>
Relevant studies are those that mirror clinical practice. If that means some studies are not in HWE, then those are relevant for making treatment decisions. Indeed, they may be the most relevant. My answer to the question “Should CYP2D6 genotyping guide the use of tamoxifen in breast cancer” is negative. CYP2D6 genotyping should not be used to guide assignment of tamoxifen treatment, not even in clinics that have great confidence that their genotyping methods are in HWE. On the other hand, researchers should continue to try to understand relationships between CYP2D6 and other genes in the way they may interact with the effects of tamoxifen and other drugs used in treating breast cancer.

References

Note
D. Berry led the initial statistical analysis of The International Tamoxifen Pharmacogenomics Consortium of the Pharmacogenetics and Pharmacogenomics Database—as yet unpublished—and dropped out of the consortium because I disagreed with the shift toward basing conclusions on ad hoc subset analyses.

Affiliation of author: Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX.