Prioritizing Pharmacogenetic Research: A Value of Information Analysis of CYP2D6 Testing to Guide Breast Cancer Treatment

Beth Woods, MSc1,*, David Veenstra, PhD, PharmD,2 Neil Hawkins, PhD, MSc1,3
1Oxford Outcomes Ltd., Oxford, United Kingdom; 2University of Washington, Seattle, WA, USA; 3University of York, York, United Kingdom

ABSTRACT

Objectives: To demonstrate how value of information (VOI) analysis can be used to establish research priorities regarding the use of pharmacogenetic tests using CYP2D6 testing to select adjuvant hormonal therapy in early stage breast cancer as a case study. Methods: The following four treatment pathways are compared in a Markov model: tamoxifen treatment; CYP2D6 test and treat homozygous and heterozygous wild type patients (wt/wt; wt/∗4) with tamoxifen and ∗4/∗4 patients with anastrozole (HetTam); CYP2D6 test and treat homozygous wild type patients with tamoxifen and others with anastrozole (HomTam); and anastrozole treatment. Pharmacogenetic testing efficacy is estimated by synthesizing randomized controlled trial data comparing tamoxifen to anastrozole with observational data linking CYP2D6 genotype to tamoxifen outcomes. Results: In order of increasing effectiveness the comparators are tamoxifen, HetTam, HomTam, anastrozole. Health outcomes for test and treatment strategies are highly uncertain. Differences in comparator costs depend on assumptions made regarding anastrozole patent expiry. The expected value of a decision taken with perfect information is £69 to £106 million (pound sterling) for the United Kingdom depending on patent expiry assumptions and the acceptable cost-effectiveness threshold. The most valuable research (VOI £53–£82 million) elucidates the relationship between CYP2D6 genotype and tamoxifen effectiveness. It is uncertain whether values of other research designs would exceed their costs. Conclusions: Retrospective analysis of one of the large adjuvant aromatase inhibitor trials is warranted to better understand any association between CYP2D6 genotype and tamoxifen outcomes. VOI approaches may be helpful for prioritising evidence needs and structuring coverage with evidence development agreements for pharmacogenetics. Keywords: CYP2D6, decision modeling, pharmacogenetics, tamoxifen, value of information.

Copyright © 2011, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

A paucity of clinically relevant supporting evidence for pharmacogenetic tests has been identified as a limiting factor for inclusion of pharmacogenetic information in drug labels [1] and the uptake of pharmacogenetic testing [2]. This lack of clinically relevant evidence is attributable to the regulatory environment for diagnostics, test pricing dynamics, and the difficulties of designing trials of biomarker-based treatment strategies.

The divergence between regulator and payer evidence requirements for pharmaceuticals is well documented [3]. For pharmacogenetic tests and other diagnostics this divergence is exacerbated because regulatory approval does not require direct evidence of clinical benefit from a randomized controlled trial (RCT). Instead, regulators of diagnostics in the United States and the European Union focus on test analytic validity (accuracy and consistency of genotype detection) and to a lesser extent clinical validity (accuracy of phenotype prediction) [4,5]. In addition, reimbursement systems for diagnostics are currently not value based [1,6,7]. For example, in the United States new tests are reimbursed based on their technical similarity (in terms of effort and complexity) to existing tests [1]. Thus, evidence of improved patient outcomes is not reflected in a higher reimbursed price. This lack of regulatory and reimbursement incentives to produce clinically relevant supporting data is compounded by the difficulties of designing adequately powered trials for biomarker-based treatment strategies [8].

The result is a limited evidence base with which to assess clinical utility (the net balance of health risks and benefits associated with using a test in routine practice) and economic value (the balance of net health outcomes against cost) that form the basis for efficient allocation of health care resources. To facilitate timely uptake of pharmacogenetic tests, a number of bodies have recommended coverage with evidence development agreements [9,10]. These agreements make reimbursement conditional upon further data collection, which is then used to re-appraise the decision [11]. However, clear guidance regarding the methods that should be used to determine when additional research is warranted and what form this research should take is lacking.

In this context, decision analysis allows us to explicitly and quantitatively address key payer questions regarding clinical util-
ity [12], economic value [7], and the uncertainty around these. Using the example of CYP2D6 testing to guide adjuvant breast cancer treatment, we report how decision modeling along with formal value of information (VOI) methods can also be used to directly evaluate the potential value of future research.

Our example evaluates pharmacogenetic testing as a strategy for treatment selection in postmenopausal women with early estrogen-receptor–positive breast cancer who have undergone surgery. This example is pertinent because of the potential health impact, uncertainty surrounding recommendations for pharmacogenetic testing, and ongoing research efforts to establish the genotype–phenotype association. The analysis takes a United Kingdom decision-maker perspective, though the disease model developed could be used as the basis for evaluations in other jurisdictions.

For this patient group, hormonal therapy is recommended to reduce risk of recurrence. Previously the hormonal therapy of choice was 5 years of tamoxifen; recently the aromatase inhibitors (anastrozole, exemestane, and letrozole) have been recommended in the United Kingdom due to their increased efficacy and acceptable cost-effectiveness [13]. Emerging clinical evidence suggests that a further treatment pathway, whereby hormonal therapy selection is determined by patient genotype, could be considered.

The cytochrome P450 (CYP450) metabolic enzyme CYP2D6 has a major role in tamoxifen metabolism [14] and is known to vary among individuals. The hypothesis is that plasma levels of endoxifen (the most abundant active tamoxifen metabolite) and clinical outcomes among patients administered tamoxifen are related to patients’ CYP2D6 genotype [15]. The availability of a commercial test kit for the CYP2D6 genotype (AmpliChip CYP450, Roche Diagnostics, Basel, Switzerland) and laboratory developed testing services mean that use of pharmacogenetic testing to select hormonal therapy is a viable treatment pathway.

The evidence base for the genotype–phenotype association, however, is contested. A recent systematic review identified 15 studies in the adjuvant setting [16] and concluded that there was no consistent association between the CYP2D6 polymorphism and tamoxifen outcomes, with studies differing in both the direction and formal statistical significance of their results. The authors identify small sample sizes, poor analytical methods, and heterogeneity in the definitions of CYP2D6 genotype derived metabolizer categories as limiting factors in evaluating the association.

The uncertainty around the genotype–phenotype association and absence of evidence regarding clinical utility has generally resulted in recommendations against uptake of the test [14,17,18]. This has included a decision by the US Food and Drug Administration (FDA) not to include CYP2D6 testing guidance in the tamoxifen label, despite advisory committee support [19]. In Europe, the Pharmacovigilence Working Party has recommended that labeling information should highlight the “possible reduction in response to tamoxifen in poor CYP2D6 metabolisers” [20]. It remains to be seen whether this will be implemented by the member states, including the United Kingdom where CYP2D6 testing is currently not used widely outside of research [21].

We evaluate the use of the AmpliChip CYP2D6 test assuming that following testing, treatment choice is based solely on the presence or absence of the *4 allele, as this allele has been most commonly studied as a determinant of tamoxifen outcomes [16]. As patient genotype is determined by two alleles, three different genotypes are possible using a *4-only based classification: wild-type (wt)/wt; wt/*4 and *4/*4, where wt refers to alleles other than *4. AmpliChip queries 26 CYP2D6 alleles of approximately 100 known polymorphisms [22]. The *4-defined treatment pathways represent only a fraction of those possible with AmpliChip. Laboratory tests (e.g. Taqman, pyrosequencing, Roche Diagnostics, Basel, Switzerland) could be used as an alternative to AmpliChip. These tests have different allele coverage and a different set of associated feasible treatment pathways.

Our analysis does not aim to provide a definitive recommendation regarding use of CYP2D6 testing to determine treatment choice. Instead, by using decision modeling, we provide estimates based on currently available evidence of the comparative effectiveness and cost-effectiveness of pharmacogenetic testing, the level of uncertainty around these estimates, and whether and what type of additional research should be conducted to support future decision making.

Methods

The decision model takes a UK National Health Service (NHS) perspective. The health outcomes of interest are 5-year recurrence-free survival (RFS) and lifetime quality-adjusted life years (QALYs). The results of the study are relevant to white patients due to the racial composition of the main clinical studies [23,24]. All patients are assumed to be recurrence free and 65 years old on entry into the model, which is in line with the key clinical trials. Costs are in 2007 United Kingdom pound sterling (£).

Synthesizing RCT and observational evidence to estimate comparative efficacy

Evidence from an RCT comparing anastrozole to tamoxifen (the ATAC trial [23]) is synthesized with observational data comparing tamoxifen efficacy across the CYP2D6 genotypes wt/wt, wt/*4, and *4/*4 (Goetz and colleagues [24]) to estimate the relative efficacy of the pharmacogenetic testing and drug treatment pathways.

The model focuses on the aromatase inhibitor anastrozole due to the richness of data available from the ATAC trial [23]. ATAC is a large (n = 9241) double-blind RCT. A recent systematic review [25] identified ATAC as the only RCT of anastrozole in the primary adjuvant setting and the largest trial of an aromatase inhibitor in the adjuvant setting.

The Goetz study [24] analyzes data from 171 patients enrolled in the tamoxifen arm of the phase III North Central Cancer Treatment Group RCT of tamoxifen compared to tamoxifen plus fluorouracil in the adjuvant setting (NCCTG 89-30-52) [26]. The Goetz study re-analyses 12 years of outcome data by CYP2D6 metabolizer categories. The metabolizer categories are defined by patient CYP2D6 genotype and exposure to selective serotonin-reuptake inhibitors (SSRIs) – a class of drugs known to inhibit the CYP2D6 enzyme and thought to affect tamoxifen outcomes. A recent systematic review [16] identified three studies that examined the association between *4 allele defined genotype categories and breast cancer recurrence for patients receiving the 5 year tamoxifen protocol [24,27,28]. Of these, Goetz [24] included the largest number of patients and was the only study that used clinical trial data and accounted for concomitant SSRI use.

The estimated RFS hazard ratios for patients receiving tamoxifen in the Goetz study [24] were 1.4 (95% confidence interval [CI] 0.68–3.05) for the comparison of wt/*4 with wt/wt genotypes and 3.2 (95% CI 1.37–7.55) for *4/*4 compared with wt/wt genotypes, based on sample sizes of 115, 40, and 16 for the wt/wt, wt/*4, and *4/*4 groups. The wt/wt, wt/*4, and *4/*4 labels used here refer to the extensive, intermediate, and poor metabolizer groups reported by Goetz [24]. Because these metabolizer classes are defined by both genotype and SSRI exposure, use of these data makes two assumptions. First, the relative efficacy of tamoxifen across metabolizer classes is representative of the relative efficacy of tamoxifen across genotype classes in the absence of SSRI administration; second, if pharmacogenetic testing was adopted, the SSRIs would not be administered. The first assumption is difficult to assess because the impact of SSRI co-administration on tamoxifen outcomes is not yet well quantified [29]. The second assumption is likely to be valid because, if pharmacogenetic testing is adopted, this will reflect confidence in the relationship between
CYP2D6 inhibition and tamoxifen outcomes that also underpins the concerns regarding SSRI co-administration.

Given this genotype classification, the following four treatment pathways are relevant (Fig. 1):

- Tamoxifen treatment;
- CYP2D6 HetTam: CYP2D6 test and treat both homozygous and heterozygous wild type patients (wt/wt and wt/*4) with tamoxifen and *4/*4 patients with anastrozole;
- CYP2D6 HomTam: CYP2D6 test and treat homozygous wild type patients with tamoxifen and others with anastrozole;
- Anastrozole treatment.

The genotype composition of the test population is estimated from a Swedish adjuvant breast cancer patient cohort, which included 68% wt/wt patients, 27% wt/*4 patients, and 5% *4/*4 patients (n = 677) [28].

Estimates of comparative efficacy for anastrozole and tamoxifen are available directly from the ATAC trial [23] in the form of parametric Weibull RFS curves. Estimation of RFS curves for the pharmacogenetic testing comparators requires genotype-specific RFS estimates for patients receiving tamoxifen. These are estimated by decomposing the ATAC genotypically unselected tamoxifen RFS curve into genotype-specific curves. This requires the assumption that the genotype composition and variation in RFS across genotypes in ATAC can be approximated by the Goetz data and that the proportional hazards assumption holds across genotypes. The higher recurrence rates among *4 carriers will result in a change in the genotype composition of the recurrence-free population over time. This is accounted for in the model by re-estimating the genotype composition of the recurrence-free population in each time period using the genotype composition at the beginning of the previous time period and the genotype-specific rates of recurrence during the previous time period. These methods are similar to those used by Punglia et al. [30].

The clinical parameters used to estimate the comparative efficacy of the pharmacogenetic testing and drug only treatment pathways are presented in Table 1.

**Extrapolating from surrogate to decision endpoints**

A model is required to extrapolate from the different RFS rates (the surrogate) to expected lifetime costs and QALYs (the decision endpoints). The disease process is modeled using a Markov approach. The disease states, allowed transitions and parameterization of the model, are similar to those used in previous models of adjuvant hormonal therapy [31–33] and are depicted in Figure 2. The Markov model simulates costs and QALYs for four cohorts: anastrozole treated, tamoxifen treated, wt/wt tamoxifen treated, and wt/*4 tamoxifen treated. Costs and QALYs associated with the pharmacogenetic testing pathways are then calculated by weighting the cohort-specific estimates by the genotype composition of the test population and including test-related costs for all patients.

Due to an absence of evidence, the model assumes that the following parameters do not vary between patients with different CYP2D6 genotypes: aromatase inhibitor efficacy and safety, tamoxifen safety, and tamoxifen recurrence type composition. Tamoxifen may well be associated with more frequent adverse events and an improved recurrence type profile in patients better able to metabolize the drug. The net direction of the impact of this assumption on the model results is therefore uncertain.

In line with previous models, patients follow the RFS curves described above for years 1 to 5, for years 6 to 15 all patients follow
the tamoxifen RFS curve (i.e., no difference in RFS according to drug or genotype is assumed). This may be conservative with respect to the effectiveness and therefore cost-effectiveness of both the CYP2D6 testing and anastrozole pathways. Evidence from the Goetz study [24] suggests that the hazard of recurrence on tamoxifen may differ between metabolizer classes up to year 10. Evidence from ATAC suggests that the hazard of recurrence may differ between tamoxifen and anastrozole up to year 8 [34]. For years 16 and onwards all patients are assumed to experience recurrence in line with incident rates in the general population [35]. The recurrence type composition (contralateral, locoregional, and metastatic) is estimated using data from ATAC [23,36], transition probabilities beyond the initial recurrence are estimated using observational data [37–39]. The clinical parameters used to extrapolate from RFS to the decision endpoints are provided in Table 2.

Utility data was taken from Lidgren et al. [40]. A previous systematic review of utility data in breast cancer [41] identified this as the only study reporting utility values for relevant health states that were derived by administering a generic quality-of-life instrument to patients and valuing their responses using a general population based tariff, in line with UK decision-maker preferences [42]. The estimates are based on EuroQol five-dimensional questionnaire health profiles elicited from 17 to 177 patients (depending on health state) attending breast cancer outpatient appointments at the Karolinska University Hospital, Sweden, in April and May 2005. These data are presented in Table 3.

Costs are incurred with CYP2D6 testing and administration, hormonal therapy acquisition and administration, and time spent in each breast cancer health state [43–46]. The cost of a delivered test result is estimated to be £395 (personal communication, Adrian Smith, Roche Diagnostics, June 13, 2008). Patients who were disease free while on treatment were assumed to attend an annual oncology outpatient visit and mammogram as per existing models [45]. Patients were assumed to require an additional oncologist appointment for communication of CYP2D6 test results and discussion of treatment. Resource use estimates for the post-recurrence health states were estimated based on a sample of 199 early breast cancer patients who relapsed between 1991 and 2004 at the Western General Hospital, Edinburgh, and covered resource use related to surgery, radiotherapy, chemotherapy, hormonal

---

### Table 1 – Clinical parameters used to estimate comparative efficacy.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate ([variance] or [95% CI–])</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weibull parameters</strong> – ATAC recurrence-free survival curve (see also Fig. 3)</td>
<td>Intercept – 9.172 (0.00533)</td>
<td>ATAC trial (Hind [31]; Mansel [32] for data)</td>
</tr>
<tr>
<td></td>
<td>Scale – 0.831 (0.00098)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anastrozole – 0.249 (0.00439)</td>
<td></td>
</tr>
<tr>
<td><strong>Hazard ratios for recurrence across genotypes</strong></td>
<td>wt/*4 vs. wt/wt – 1.4 (0.68–3.05)</td>
<td>Goetz [24]</td>
</tr>
<tr>
<td></td>
<td>*4/*4 vs. wt/wt – 3.2 (1.37–7.55)</td>
<td></td>
</tr>
<tr>
<td><strong>Genotype composition in Goetz cohort</strong></td>
<td>wt/wt – 67% (115/171)</td>
<td>Goetz [24]</td>
</tr>
<tr>
<td></td>
<td>wt/*4 – 23% (40/171)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*4/*4 – 9% (16/171)</td>
<td></td>
</tr>
<tr>
<td><strong>Genotype frequencies in test population</strong></td>
<td>wt/wt – 68% (460/677)</td>
<td>Wegman [28]</td>
</tr>
<tr>
<td></td>
<td>wt/*4 – 27% (183/677)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*4/*4 – 5% (34/677)</td>
<td></td>
</tr>
</tbody>
</table>

ATAC, arimidex, tamoxifen, alone or in combination trial; 95% CI, 95% confidence interval.

* SAS parameterization of Weibull distribution.

---

**Fig. 2 – Modeling from recurrence-free survival to decision endpoints (death from all causes possible from all states, not shown). Yr1, year 1; Yr2+, year 2 and beyond.**
Patients are assumed to receive hormonal therapy for 5 years or until breast cancer recurrence. At the time of the analysis, the patent for anastrozole is expected to expire in February 2011 in the UK [47], all analyses are therefore presented for two scenarios, the first assumes the price of anastrozole will be held at the prevailing level (“patent price” scenario) and the second estimates that the anastrozole price will fall by 82% when generics are introduced (“generic price” scenario), based on the experience in New Zealand [48]. The costs estimates used in the model are provided in Table 3.

A time horizon of 35 years is used in the model to generate lifetime cost and health outcome estimates. A 1 year cycle length is used and a half cycle correction is applied. Costs and health outcomes are discounted at 3.5% per annum in line with the stipulations of UK decision makers [42].

### Table 2 – Clinical parameters used to extrapolate from recurrence free survival to decision endpoints.

<table>
<thead>
<tr>
<th>Parameter type</th>
<th>Estimate</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recurrence probability, tamoxifen years 15+</td>
<td>Age 80–84: 0.38%</td>
<td>UK National statistics [35]</td>
</tr>
<tr>
<td></td>
<td>Age ≥85: 0.43%</td>
<td></td>
</tr>
<tr>
<td>Recurrence composition on tamoxifen</td>
<td>Contralateral = 54/420</td>
<td>AstraZeneca [36]; ATAC [23]</td>
</tr>
<tr>
<td></td>
<td>Locoregional = 101/420</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metastatic = 265/420</td>
<td></td>
</tr>
<tr>
<td>Treatment effect of anastrozole on recurrence</td>
<td>Contralateral RR = 0.48 (SE ln(RR): 0.24)</td>
<td></td>
</tr>
<tr>
<td>(yrs 1-5 only)</td>
<td>Locoregional RR = 0.75 (SE ln(RR): 0.15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metastatic HR = 0.84 (SE ln(HR): 0.09)</td>
<td></td>
</tr>
<tr>
<td>Probability metastatic given contralateral / locoregional recurrence</td>
<td>0.72 (10 year probability) (n = 140)</td>
<td>Kamby and Sengelov [37]</td>
</tr>
<tr>
<td>Median survival if metastatic</td>
<td>17.8 months (n = 346)</td>
<td>Chang [38]</td>
</tr>
<tr>
<td>All-cause mortality</td>
<td>Age dependent</td>
<td>Government Actuary’s Department [39]</td>
</tr>
</tbody>
</table>

ATAC, arimidex, tamoxifen, alone or in combination trial; HR, hazard ratio; RR, relative risk; SE, standard error of the mean.

### Table 3 – Utility (quality of life) and cost parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease-free survival (year 1) – receiving hormonal therapy</td>
<td>0.744 (95% CI 0.573–0.841)</td>
<td>Lidgren [40]</td>
</tr>
<tr>
<td>Disease free survival (year 2+) – receiving hormonal therapy</td>
<td>0.824 (95% CI 0.785–0.857)</td>
<td></td>
</tr>
<tr>
<td>Contralateral/locoregional recurrence (year 1)</td>
<td>0.779 (95% CI 0.700–0.849)</td>
<td></td>
</tr>
<tr>
<td>Contralateral/locoregional recurrence (year 2+)</td>
<td>0.779 (95% CI 0.745–0.811)</td>
<td></td>
</tr>
<tr>
<td>Metastatic recurrence</td>
<td>0.685 (95% CI 0.620–0.735)</td>
<td></td>
</tr>
<tr>
<td>AmpliChip delivered result (cost for laboratory to analyze blood sample including AmpliChip device, labor, controls, and consumables)</td>
<td>£395</td>
<td>Personal communication, Adrian Smith, Roche Diagnostics, June 13, 2008.</td>
</tr>
<tr>
<td>Consultant appointment to discuss pharmacogenetic test results</td>
<td>£83.31</td>
<td>NHS reference costs [43]. Cost codes TCLFUMFF 103; TCLFUSFF 103; TCLFUMFF 800; TCLFUSFF 800 weighted by number of attendances.</td>
</tr>
<tr>
<td>Tamoxifen (generic) 20 mg once daily (20 mg × 30 pack)</td>
<td>£3.07</td>
<td>British National Formulary [44,54]</td>
</tr>
<tr>
<td>Anastrozole (arimidex) 1 mg once daily (1 mg × 28 pack)</td>
<td>£68.56 (£12.34 for generic scenario)</td>
<td>British National Formulary [44,54]</td>
</tr>
<tr>
<td>Annual cost disease-free survival</td>
<td>£111.02</td>
<td>Sum of cost of consultant appointment (see above) and mammogram – NHS reference costs [43], cost code RA28Z. Karron [45] 2005 costs, inflated to 2007 values using UK health care inflation indices [46].</td>
</tr>
<tr>
<td>Contralateral tumor (year 1)</td>
<td>£15,684</td>
<td></td>
</tr>
<tr>
<td>Contralateral tumor (year 2+)</td>
<td>£11,887</td>
<td></td>
</tr>
<tr>
<td>Locoregional tumor (year 1)</td>
<td>£479</td>
<td></td>
</tr>
<tr>
<td>Locoregional tumor (year 2+)</td>
<td>£1,058</td>
<td></td>
</tr>
<tr>
<td>Distant metastases</td>
<td>£5,279</td>
<td></td>
</tr>
<tr>
<td>Terminal care</td>
<td>£4,941</td>
<td></td>
</tr>
</tbody>
</table>

Costs given in pound sterling (£). NHS, National Health Service.
Estimating uncertainty

Uncertainty is quantified in the decision model by assigning distributions to all parameters that are subject to sampling uncertainty. This joint uncertainty in the parameter inputs is then propagated through the model via Monte Carlo simulation. A multivariate normal distribution is assigned to the Weibull parameters describing the ATAC RFS curves; Dirichlet distributions are assigned to genotype prevalence estimates for the test and ATAC populations and to the composition of recurrence type; log-normal distributions are assigned to all hazard ratio and relative risk parameters; beta distributions to the probability of progression from contralateral/locoregional to metastatic disease, the probability of death from metastatic disease, the utility parameters, and the proportions used in resource use calculations; and gamma distributions for aggregate costs, unit costs and resource use estimates.

The parameterization of each distribution can be derived from the data presented in Tables 1, 2, and 3 with the following exceptions. For the multivariate normal distribution the following additional parameters are required: covariance (intercept, scale) = 0.00185; covariance (intercept, coefficient) = −0.00134; covariance (scale, coefficient) = 0.00028 (personal communication, Enrico de Nigris, Double Helix, October 2, 2008). For the cost data, limited information regarding sampling uncertainty is available. For aggregate costs we made the conservative assumption that the standard error is equal to half the mean. Where separate unit cost and resource use estimates are used, standard errors for the former are based on the assumption that the interquartile range reported in UK reference costs [42] represents a 50% CI and for the latter are set equal to half the mean values.

Drug costs, the cost of testing, population-based parameter estimates (all-cause mortality; post–15-year recurrence), and discount rates are not subject to stochastic uncertainty and are not assigned probability distributions.

Estimating the value of research

The expected value of perfect information (EVPI) is estimated. EVPI provides an estimate of the expected value of eliminating all uncertainty relating to the decision. Although an unattainable goal, this gives the upper bound on the expected value of further research. Therefore, any research design with costs in excess of EVPI should not be undertaken [49].

EVPI is estimated for each decision threshold by calculating the net benefit distribution associated with the simulated cost and QALY pairs. For each simulation, the cost of uncertainty is calculated as the difference between the net benefit of the technology that offers the maximum net benefit for that particular simulation and the expected net benefit of the technology that maximizes net benefit on average across all simulations [50]. Averaging this across simulations and multiplying by the number of patients expected to benefit from the information generates the EVPI estimate. The number of patients expected to benefit from the information in the UK is estimated as the discounted sum of a 10-year stream of 22,781 incident cases per annum [31].

The expected value of having perfect information about particular groups of parameters (expected value of partial perfect information [EVPPI]) is also calculated. This requires two simulation loops [51]. The outer loop involves randomly sampling (750 iterations) from the parameter group of interest; the inner loop involves, for each outer loop simulation, randomly sampling from the remaining parameters (1000 iterations). For each outer loop simulation the net benefit of the intervention that offers the maximum expected net benefit across inner loop simulations is calculated. This is then averaged across outer loop simulations to estimate the net benefit associated with perfect information about the parameter group of interest. Subtraction of the expected net benefit associated with the most cost-effective technology based on current information then generates the EVPPI. Again, this is multiplied by the total number of patients expected to benefit from the information.

For the EVPPI calculations parameters were grouped according to the type of study required to obtain further data on them.

- Genotyping studies was split into two components: the first aimed to provide further information on genotype prevalence in the test population, which could be achieved using a cross-sectional study; the second was to provide further information on the genotype prevalence and difference in RFS across genotypes used to decompose the ATAC tamoxifen RFS curve, which would require a longitudinal study (e.g., a retrospective association study using existing tamoxifen RCT data);
- RCT: would provide further information regarding recurrence and adverse event probabilities for anastrozole and tamoxifen (i.e., another ATAC-type trial);
- Costing study: would provide further information regarding costs associated with breast cancer and adverse event states;
- Utility study: would provide further information regarding utility weights associated with breast cancer and adverse event states.

All analyses were implemented in R with the MASS and MCMC packages installed [52] (R Foundation for Statistical Computing, Vienna, Austria).

Results

Comparative effectiveness and cost-effectiveness

The genotype-specific RFS curves derived from synthesising the observational and RCT data are presented in Figure 3. The model predicts that wt/wt patients receiving tamoxifen perform almost as well as patients receiving anastrozole. The wt/*4 and *4/*4 patients receiving tamoxifen exhibit poorer performance than genotype-unselected tamoxifen patients whose performance is heavily influenced by the prevalent wt/wt genotype.

The associated mean 5-year RFS rates predicted by the model for each treatment pathway are 87.3% for tamoxifen, 89.0% for CYP2D6 HetTam, 90.2% for CYP2D6 HomTam, and 90.5% for anastrozole.

These differential recurrence rates drive the differences in QALYs and health state costs across comparators (see Table 4). In the patent price scenario, differences in costs across comparators are driven by the proportion of patients receiving anastrozole in the treatment pathway. The test cost does not drive results for this scenario. The incremental cost per QALY results for this scenario suggest that assuming an acceptable cost-effectiveness threshold of £20 to £30,000/QALY (pound sterling), CYP2D6 HomTam (CYP2D6 testing followed by tamoxifen for wt/wt genotypes and anastrozole for all others) would be the optimal treatment pathway. Although this is not quite the most effective comparator, moving to the most effective comparator anastrozole is unlikely to be considered good value as the marginal 0.012 QALY gain is achieved at a cost of £2,081 with an associated incremental cost-effectiveness ratio of £177,096/QALY. The incremental cost-effectiveness ratio for the comparison of anastrozole with tamoxifen (£20,830/QALY) is within the range of findings of previous UK studies that estimated incremental costs per QALY of £11,428 [45], £17,656 [32], and £31,965 [31]. Figure 4 presents the results of the patent price analysis as a cost-effectiveness frontier. Under the generic price scenario, anastrozole dominates the other comparators, and offers a marginally lower cost and the highest expected health outcomes. In this scenario differences in treatment, test, and health state costs all drive the small differences in total costs.
Uncertainty in efficacy, costs, and cost-effectiveness

Depending on expectations regarding the generic anastrozole price, a decision maker may consider adopting the CYP2D6 HomTam comparator to replace the incumbent treatment (tamoxifen or anastrozole depending on the context). Estimates of the uncertainty around the incremental health outcomes (QALYs shown, 5-year recurrence rates display the same trends) and cost of CYP2D6 HomTam relative to each treatment are presented as simulations on the cost-effectiveness plane in Figure 5. Uncertainty in total costs is relatively low, the main uncertainty is around the likely price of anastrozole following patent expiry. Uncertainty around health outcomes is high and is similar regardless of the comparison as it is driven by uncertainty in CYP2D6 HomTam efficacy and not uncertainty in drug treatment efficacy which is relatively low. However, the nature of the uncertainty depends on the comparator treatment. Compared with tamoxifen, the model predicts a high probability that pharmacogenetic testing offers a health gain. Compared with anastrozole, CYP2D6 HomTam offers close to equal probabilities of losses and gains in health outcomes. Cost-effectiveness acceptability curves, indicating the probability that each intervention is cost-effective across plausible cost-effectiveness thresholds, are presented in Figure 6.

Value of further research

Results of the EVPI and EVPPI analyses are presented in Figure 7. Based on a 10-year stream of 22,781 cases per annum, the EVPI is estimated at £69 to £106 million depending on whether a decision threshold of £20,000 or £30,000/QALY is assumed and whether the expected reduction in anastrozole price is modeled. This value is likely to exceed the costs of feasible research designs, indicating that further research may be of value. The EVPPI results suggest that the parameters that could be informed by a longitudinal study relating CYP2D6 genotype to tamoxifen outcomes (i.e., a study that could estimate the genotype composition of the ATAC trial and the hazard ratios for RFS across genotype subgroups) are associated with a research value of £53 to £82 million, which is likely to exceed the research cost. These parameters could be informed by genotyping one of the large trials comparing the aromatase inhibitors with tamoxifen (for example, the ATAC trial). It is less certain that the value

<table>
<thead>
<tr>
<th>Comparator</th>
<th>Life years (undiscounted)</th>
<th>QALYs (discounted)</th>
<th>Costs (discounted)</th>
<th>ICER (£/QALY) vs. comparator (patent/generic)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Testing</td>
<td>Treatment (patent/generic)</td>
</tr>
<tr>
<td>Tamoxifen</td>
<td>15.63</td>
<td>9.20</td>
<td>—</td>
<td>£170</td>
</tr>
<tr>
<td>CYP2D6 HetTam</td>
<td>15.78</td>
<td>9.28</td>
<td>£479</td>
<td>£370 / £199</td>
</tr>
<tr>
<td>CYP2D6 HomTam</td>
<td>15.90</td>
<td>9.35</td>
<td>£479</td>
<td>£1,442 / £355</td>
</tr>
<tr>
<td>Anastrozole</td>
<td>15.92</td>
<td>9.36</td>
<td>—</td>
<td>£4,136 / £744</td>
</tr>
</tbody>
</table>

Costs given in pound sterling (£).

ICER, incremental cost-effectiveness ratio; QALYs, quality-adjusted life years.
of the other research designs evaluated would exceed their costs.

Interestingly, the EVPI and EVPPI for longitudinal genotyping estimates are higher under the generic anastrozole price scenario than the patent price scenario when a £30,000/QALY acceptable cost-effectiveness threshold is assumed. Under the patent price scenario the loss of value associated with choosing CYP2D6 HomTam in the simulations where anastrozole offers

![Cost-effectiveness frontier for patent price analysis. Costs given in pound sterling (£). QALYs, quality-adjusted life years.](image)

Fig. 4 – Cost-effectiveness frontier for patent price analysis. Costs given in pound sterling (£). QALYs, quality-adjusted life years.

![Uncertainty in efficacy and costs of CYP2D6 testing versus incumbent treatments. Costs given in pound sterling (£).](image)

Fig. 5 – Uncertainty in efficacy and costs of CYP2D6 testing versus incumbent treatments. Costs given in pound sterling (£).
higher net benefit is limited by the high additional cost of choosing the drug treatment pathway. Conversely, under the generic price scenario where anastrozole is the cost-effective treatment, the loss of value associated with choosing anastrozole in the simulations where HomTam offers higher net benefit is higher because there is little difference between strategies with respect to cost. When a £20,000/QALY threshold is assumed, the EVPI and EVPPI for longitudinal genotyping estimates are lower under the generic scenario than the patent scenario. This reflects the reduced importance of uncertainty around health outcomes relative to uncertainty around costs in determining the VOI estimates for the lower threshold.

Discussion
By applying decision modeling and VOI analysis to the example of CYP2D6 testing to guide tamoxifen therapy, we have been able to quantify the potential clinical utility and economic value of testing based on current evidence, the uncertainty around these estimates, and the value of further research. Both the clinical utility and economic value of the test are highly uncertain, due to uncertainties in the association between CYP2D6 genotype and tamoxifen outcomes, and due to uncertainties regarding the price trajectory of anastrozole. The VOI analysis provides quantitative guidance regarding research prioritization and indicates that fur-
ther research of the association between CYP2D6 genotype and tamoxifen RFS is likely to represent a cost-effective use of health care research resources. It is less clear that research on the genotype prevalence, relative drug effectiveness (RCT), utility, and cost parameter groups would be cost-effective because the value of research may not exceed the associated costs. Furthermore, it is not clear whether the values associated with these research designs would be maintained if further data collection was conducted to reduce the uncertainty in the genotype-phenotype association.

The VOI estimates presented are restricted to the value for the United Kingdom. Because further data regarding the association between CYP2D6 genotype and tamoxifen recurrence would likely be considered relevant in a number of jurisdictions, the estimates presented can be considered a lower bound to total (global) EVPPI. Furthermore, as the clinical uncertainties apply across jurisdictions it is likely that the VOI estimates would exceed the cost of research in many jurisdictions.

**Study limitations**

The current study is based on the NCCTG 89-30-52 trial data, which found a statistically significant but highly uncertain association between the presence of the *4 allele and poorer tamoxifen outcomes. Other studies of the *4 allele have produced conflicting results. A recent review [16] identified two studies reporting cohorts that exhibited a statistically significant association [24,53] and three which found statistically insignificant trends for improved outcomes in carriers of the *4 allele [27,28,54]. All of these studies are small observational cohorts and only two controlled for concomitant SSRI administration [24,53]. Use of these studies instead of the NCCTG 89-30-52 data or alongside it using evidence synthesis would produce different cost-effectiveness and VOI estimates. However, we would expect the VOI estimates to remain high as all studies report very uncertain outcomes.

The CYP2D6 testing treatment pathways included in the current analysis represent all those possible using only the presence of the *4 allele as a determinant of treatment choice. The review by Terasawa et al. [16] identified four studies in Caucasian populations, which examined the impact of additional alleles on tamoxifen outcomes. These studies report outcomes in different subgroups defined by one or more of 22 genotypes. These genotypes could be used to define a very large number of possible alternative treatment pathways following CYP2D6 testing. Again, we would expect the VOI estimate to remain high if additional treatment pathways were included as there would be more comparators and the associations between genotype and outcomes in these studies were also very uncertain. Given the uncertainties in the current evidence base it is also possible that other CYP2D6 genotypes or other biomarkers that better predict tamoxifen outcomes will be identified in the future.

The model is restricted to white patients due to the racial composition of the key clinical studies. There is evidence to suggest, however, that CYP2D6 genotype may relate to tamoxifen outcomes in other racial groups [55,56]. Further decision modeling work could be used to estimate the clinical utility of testing, the economic value of testing and the need for further research in these populations.

The model focuses on the aromatase inhibitor anastrozole due to the richness of data available from the ATAC trial. Inclusion of the other aromatase inhibitor licensed in the primary adjuvant setting, letrozole, would be unlikely to dramatically alter the conclusions because it is associated with similar efficacy [25] and cost [31] as anastrozole. Further work would be required to confirm this.

Costs of alternative CYP2D6 testing approaches, for example bespoke laboratory testing, would produce different estimates of the cost-effectiveness of testing and value of further research. It is not expected that this would markedly alter the conclusions of the analysis because test cost is not a key driver of the model and the cost of laboratory testing is thought to be similar to or lower than the cost of AmpliChip [21].

The current analysis explores two possible scenarios regarding the long-term price trajectory for anastrozole. It is quite possible that the timing of patent expiry and generic price for anastrozole will differ from the scenarios modeled and thus impact on the VOI. The scenarios modeled, however, do suggest that the value of further research to estimate the relationship between CYP2D6 genotype and tamoxifen outcomes is likely to remain high.

Some of the assumptions made in the current model may not be tenable, for example the assumption that tamoxifen safety and recurrence composition does not vary with CYP2D6 genotype, whereas RFS does. These assumptions were made in the absence of relevant data. We would not expect these assumptions to have a large impact on our results as overall recurrence is the key driver in the model. A retrospective study examining the relationship between genotype and tamoxifen outcomes could also be used to...
examine the relationship between recurrence type and genotype. Understanding the relationship between genotype and adverse event rates will be challenging due to the relatively low event rates.

Implications

Payers, research commissioners, and regulators. Our analysis indicates that there is likely to be a high societal value to retrospectively genotyping one of the large trials comparing the aromatase inhibitors with tamoxifen. The two trials in the primary adjuvant setting – ATAC [23] and BIG 1-98 [57] – were sponsored by the aromatase inhibitor manufacturers AstraZeneca and Novartis. Therefore, it is not clear that a test manufacturer could comply with a “coverage with evidence development” or “only in research” recommendation for CYP2D6 testing by a payer. Coverage for testing with a requirement for prospective observational data collection is plausible, but may be of limited value as evaluation of the genetic association would not be possible if women with a variant genotype were not given tamoxifen. A plausible study design would be an RCT comparing testing versus no testing strategies, but such a trial would likely require a large sample size and be expensive. Furthermore, even if this research was possible it is not clear that it would produce a net benefit to the test manufacturer because the device is relatively cheap and unlike patented pharmaceuticals, it faces direct competition.

This raises the question of whether regulators or payers should mandate the retrospective genotyping study. It could be argued that this should be a condition for continued marketing approval for the aromatase inhibitors because the association between CYP2D6 genotype and tamoxifen outcomes is important implications for the aromatase inhibitor risk-benefit profile. With the pending patent expiration of anastrozole, the financial incentives for such a mechanism are limited. Thus, public funding to evaluate the validity of the genetic association may be the most realistic scenario.

Interestingly, our analysis indicates that a prospective RCT comparing testing to no testing may not be the most cost-effective research design. These findings both help clarify research priorities but also inform the discussion of evidence thresholds for pharmacogenetic tests from a regulatory, guideline, and reimbursement perspective.

Pharmaceutical industry. The current analysis clearly describes the need to evaluate the full “test and treat” pathway in decision models, rather than evaluating a drug in a genotype-defined subgroup. That is, treatments and outcomes in patients who “test negative” need to be understood. This has important implications for pharmaceutical companies operating in a value-based pricing environment and deciding whether to target a particular genetic subgroup. The genotype targeted indication evaluated as a subgroup may appear to offer the potential for a much higher price than the price feasible once the cost of genotyping all patients is incorporated.

Test developers. Decision modeling and VOI analysis should be used by test developers as a method of demonstrating the value of their tests in terms of clinical and economic benefits and the value of further clinical research. Decision modeling could also be used to inform trial design by identifying which test and treat strategies should be included as comparators and whether a trial is likely to be worthwhile (if the probability that testing is more effective or cost-effective than standard care is very low a trial may not be worth undertaking).

Areas for further research. Heterogeneity in the groupings of genotypes across studies describing the genotype–phenotype association makes evidence synthesis difficult. Even within the studies examining only the *4 allele, some studies reported results separately for the wt/wt; wt/*4, and *4/*4 categories whereas others grouped the wt/*4 and *4/*4 categories together. Although pre-specification of the groups for comparison is important to avoid researchers data dredging for statistically significant differences, reporting should also include separate results for each genotype evaluated. This data could then be readily synthesized by applying network meta-analysis methods currently employed to synthesize trials containing different comparator sets.

Conclusion

Our analysis suggests that retrospective analysis of one of the large adjuvant aromatase inhibitor trials to better understand any association between CYP2D6 genotype and tamoxifen outcomes would represent a valuable use of health care research resources. More generally, VOI approaches are likely to be helpful for prioritizing evidence needs and structuring coverage with evidence development agreements for pharmacogenetics.

Postscript

Subsequent to completion of this work, preliminary results from genotyping the ATAC and BIG 1-98 trials with the purpose of examining the association between CYP2D6 genotype and tamoxifen outcomes were presented at the San Antonio Breast Cancer Symposium in December 2010 [58–60].

CYP2D6 genotype data from ATAC were available for 588 and 615 patients randomized to tamoxifen and anastrozole, respectively (13% of the trial population). Patients were genotyped for the seven most common alleles that were then used to assign a score based on predicted allele activities; potent CYP26 inhibitor usage was controlled for. CYP2D6 genotype data from BIG 1-98 were available for 4628 patients (58% of the trial population). Patients were defined as poor, intermediate, and extensive metabolizer categories based on their genotype. Both studies have reported preliminary estimates of the association between expected CYP2D6 activity and breast cancer recurrence rates; however, neither found evidence of a relationship based on the classifications used.

These preliminary findings suggest that CYP2D6 testing is unlikely to provide clinical utility or economic value in this context. Full publication of these results and an assessment of the robustness of the algorithm used to define metabolizer classes will likely determine whether this is the last word on the CYP2D6-tamoxifen story.

Acknowledgments

Two anonymous reviewers are acknowledged for their valuable comments on an earlier version of the manuscript.

Source of financial support: Funding for developing the initial decision model was provided by Roche Diagnostics. No funding was provided for this manuscript and Roche Diagnostics have not commented on or approved the manuscript. David Veenstra was supported by the Center for Genomics and Healthcare Equality at the University of Washington (NIH Grant P50HG003374; Burke, PI). An earlier version of this work was presented at the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 14th Annual International Meeting, Florida, May 16–20, 2009.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at doi:10.1016/j.jval.2010.10.016, or if hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).