The clinical effectiveness and cost-effectiveness of genotyping for *CYP2D6* for the management of women with breast cancer treated with tamoxifen: a systematic review

N Fleeman,[1]* C Martin Saborido,[2] K Payne,[3] A Boland,[1] R Dickson,[1] Y Dundar,[1] A Fernández Santander,[4] S Howell,[5] W Newman,[6] J Oyee[1] and T Walley[7]

[1]Liverpool Reviews and Implementation Group (LRiG), University of Liverpool, Liverpool, UK
[2]School of Nursing and Physiotherapy, Universidad Pontificia Comillas, Madrid, Spain
[3]Health Sciences – Methodology, University of Manchester, Manchester, UK
[4]Department of Biomedical Sciences, Universidad Europea de Madrid, Madrid, Spain
[5]The Christie NHS Foundation Trust, Manchester, UK
[6]Genetic Medicine, University of Manchester, Manchester, UK
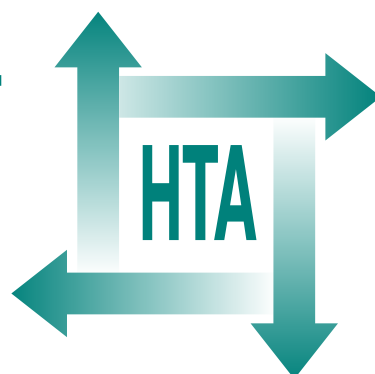[7]Health Services Research, University of Liverpool, Liverpool, UK

*Corresponding author

Executive summary

Executive summary

Background

Breast cancer is the most common cancer affecting women in the UK. Tamoxifen (TAM) is considered the standard of care for premenopausal women with oestrogen receptor positive (ER+) breast cancer and for postmenopausal women with ER+ early breast cancer considered to be at low risk of disease recurrence.

A link between drug metabolism and drug response has been widely discussed in the literature, and a significant proportion of this literature is focused on the cytochrome P450 (CYP450) enzyme system, which has been identified as a major metabolic pathway for many drugs and a source of interindividual variability in patient response. In particular, TAM is metabolised to its active metabolites *N*-desmethyl TAM and 4-hydroxytamoxifen by a number of CYP450 enzymes, including CYP2D6, CYP3A4, CYP2C9, CYP2C19, and CYP2B6. *N*-desmethyl TAM is further metabolised to endoxifen by CYP2D6. Endoxifen, which is also formed via the action of CYP2D6 is 30- to 100-fold more potent than TAM in suppressing oestrogen-dependent cell proliferation, and is considered an entity responsible for significant pharmacologic effects of TAM.

Wide variability in the response of individuals to drugs at the same doses may occur as a result of interindividual differences which may be inherited (pharmacogenetics). Genes are instructions that produce enzymes. The CYP2D6 enzyme is highly polymorphic: there are more than 60 different alleles of the *CYP2D6* gene which may be deficient or overactive in enzyme activity. It is the alleles that determine an individual's genotype and there is believed to be an association between genotype and the expected drug effects (i.e. the phenotype). For patients with normal enzyme activity [extensive metabolisers (EMs)], usual doses of a drug should result in expected drug concentrations and normal therapeutic response. Patients with deficient alleles [poor metabolisers (PMs) or intermediate metabolisers (IMs)] are likely to have lower exposure to endoxifen and may have compromised clinical effects, whereas patients with multiple alleles [ultra-rapid metabolisers (UMs)] will have increased metabolism.

CYP2D6 activity may be affected not only by an individual's genotype but also by co-administration of drugs that inhibit the metabolic activity of CYP2D6. For example, patients treated with TAM are commonly also prescribed selective serotonin reuptake inhibitors to treat adverse events (AEs) such as hot flushes, but it has been reported that fluoxetine or paroxetine effectively changes the phenotype from EM to PM in some individuals. Co-administration of such substances therefore needs to be taken into consideration.

Objectives

Clinical validity

In patients treated with TAM:

- Do women with breast cancer, identified as EMs for *CYP2D6*, have similar or different clinical outcomes to those identified as PMs, IMs or UMs?

- Is there a relationship between *CYP2D6* status and endoxifen concentrations?
- Are endoxifen concentrations related to clinical outcomes?

Clinical utility

- Do women with breast cancer who are identified as EMs for *CYP2D6* have similar or different clinical outcomes with TAM compared with aromatase inhibitors?

Cost-effectiveness

- What is the relative cost-effectiveness of *CYP2D6* testing as a management option for women with breast cancer?

Methods

Two systematic reviews related to genotyping for *CYP2D6* in the management of women with breast cancer were conducted. The first reviewed the clinical effectiveness, while the second considered economic evaluations related to *CYP2D6* testing.

Several search strategies of bibliographic databases were undertaken of various databases including MEDLINE, EMBASE, The Cochrane Library (Cochrane Database of Systematic Reviews and Cochrane Controlled Trials Register), Web of Science (for the Science Citation Index and Conference Proceedings Citation Index) and the Centre for Reviews and Dissemination databases (Database of Abstracts of Reviews of Effects, NHS Economic Evaluation Database, Health Technology Assessment), the Human Genome Epidemiology Network Published Literature database, Proceedings of the American Society of Clinical Oncology, the San Antonio Breast Cancer Symposium and the European Society for Medical Oncology. Current research was identified from database citations through searching the National Research Register, the Current Controlled Trials register, the Medical Research Council Clinical Trials Register and the US National Institutes of Health website (ClinicalTrials.gov). Relevant reviews were hand searched in order to identify any further studies. Searches were completed by 21 July 2009. However, further studies that became known to the authors via relevant conferences or e-mail alerts from an automatically updated search of the Scopus database were also included as the review progressed, up to, and including, 17 March 2010.

Data were extracted into structured tables and narratively discussed in the relevant sections of the report. In the absence of clinical utility studies and owing to heterogeneity of the alleles genotyped, phenotypes derived, patients included and outcomes measured, meta-analyses of the clinical validity data could not be performed; exploratory analysis of clinical sensitivity and specificity was therefore conducted to supplement the narrative. Data extracted from the clinical and economic reviews were intended to inform the future development of an economic model.

Inclusion criteria

For the clinical review, any study design except single-case studies was included. The patient population was women with ER+ breast cancer treated with TAM and genotyped for *CYP2D6*. Relevant outcome measures included efficacy end points, AEs and measures of endoxifen concentrations. For the economics literature review, economic evaluations that considered both the costs and benefits of *CYP2D6* genotyping and strategies comparing aromatase inhibitors with TAM were included.

Results

Clinical evaluation

Number and quality of studies

The literature search yielded 1186 citations, of which 39 were included in the review. These citations reported on 34 separate studies, but it was apparent that many of the studies reported on the same cohort of patients although with a few subtle differences, such as using only a specific subgroup of patients, considering different genotypes, taking into account concomitant medication that inhibits CYP2D6 or analysing different outcomes. Thus, in total, 25 cohorts were included in the review.

While the majority of the studies included in these cohorts were published as full papers in peer-reviewed journals, six cohorts were reported only as findings in conference proceedings. The majority of cohorts ($n = 18$) were explicit about both the source population from which the study population was derived and the definition of the study population itself. While 5 out of 12 cohorts with missing genotype data failed to state why there were missing data, all but four of the cohorts (which were published only as abstracts) presented the number of patients contributing to each analysis.

Cohort characteristics

The size of the cohorts varied, with the smallest containing 12 subjects and the largest containing 2880 (which also included patients from three published studies). However, the majority ($n = 19$) of cohorts included between 60 and 300 patients. The seven cohorts that measured endoxifen plasma concentrations were conducted prospectively, with all other studies being analysed retrospectively, using archived samples.

Cohorts included patients from the USA and/or Europe ($n = 18$) or the Asian countries of China, Japan and South Korea ($n = 6$) or from all continents ($n = 1$). In all but four studies, the TAM dose was either stated to be 20 mg/day or believed to be this in the absence of these data being provided. The majority of included patients were postmenopausal with early ER+ breast cancer. Adequate data on adjuvant chemotherapy and CYP2D6 inhibitor use were often missing (in 14 and 13 cohorts, respectively). There was wide variety in a number of other patient characteristics, such as tumour size and nodal status, across the studies.

Fifteen cohorts measured efficacy, six cohorts reported on AEs and seven cohorts measured endoxifen concentrations in relation to CYP2D6 status.

Derivation and classification of phenotypes

An important finding from our review was that there is no consensus about how *CYP2D6* phenotypes should be derived from their genotypes and how they should thus be compared, which has made the conduct of this review particularly problematic. Thus, for the purpose of this review, the following 'standardised comparisons' were used to analyse the efficacy data:

- PM versus EM
- IM versus EM
- PM + IM versus EM
- PM versus EM + IM
- Asian patients genotyped *10* allele (i.e. a common allele found in these populations)
- other.

It should be noted that, for the purposes of these comparisons, UMs are likely to be classified as EMs. This is because not all genotyping methods are able to detect UMs, and where cohorts have used methods that did, UMs appear to be classified with EMs.

Differences in cohort characteristics by genotype or phenotype

As well as differences in cohort characteristics, such as tumour size, across studies, it was evident that there were also differences within individual studies by genotype or phenotype. While eight cohorts provided these data in their publications, five of these and three others adjusted for such variables in their analyses.

Efficacy by genotype or phenotype

Not all clinical end points measured by the cohorts were clearly defined. Where end points were defined, it was apparent that different definitions were commonly used, for example DFS. Crucially, not all cohorts genotyped for the same alleles. Thus, comparisons across studies should be treated with a degree of caution.

Poor metaboliser versus extensive metaboliser

From two cohorts, no evidence of a difference in overall survival (OS) between PMs and EMs was reported. However, there was evidence of improved outcomes in terms of relapse/recurrence (disease-free survival, recurrence-free survival or time to recurrence) in the three cohorts that compared these outcomes.

Intermediate metaboliser versus extensive metaboliser

There was no evidence of a difference in OS or relapse/recurrence between IMs and EMs from the only cohort that compared outcomes for these two phenotypes.

Poor metaboliser plus intermediate metaboliser versus extensive metaboliser

In the four cohorts that explored OS between these groups of patients, there was no evidence of a difference between PMs + IMs and EMs. However, five out of eight cohorts reported significantly improved outcomes for relapse/recurrence in EMs. Interestingly, in one of these cohorts, reported only as an abstract, the significant differences were found only when using the AmpliChip® (Roche Molecular Systems) to genotype for an extensive number of alleles and not when four common alleles were tested for.

Poor metaboliser versus extensive metaboliser plus intermediate metaboliser

There was no evidence of a difference in OS or of relapse/recurrence between PMs and EMs + IMs from any of the three cohorts that compared these outcomes in these groups of patients.

Asian patients genotyped for the *10 allele

No cohorts reported convincing evidence of differences by genotype for OS (one cohort), breast cancer mortality (two cohorts) or relapse/recurrence (four cohorts).

Other

Summarising the data from the three cohorts that reported outcomes by phenotypes that do not fit the 'standard comparisons' explored above is problematic owing to the different genotype/phenotype/functional classifications used. However, in each of the cohorts there was some suggestive evidence that EMs have better relapse/recurrence outcomes than patients with other phenotypes.

Adverse events by genotype or phenotype

Three cohorts reported that EMs and IMs were more likely than PMs to experience hot flushes. One cohort also suggested that EMs were more likely to develop severe or very severe hot flushes, and also reported that, of those patients who discontinued treatment because of TAM side effects, just under half did so as a result of hot flushes. None of these patients was found to be a PM. In fact, this cohort reported that EMs were at greatest risk of discontinuing treatment as a result of TAM side effects.

Endoxifen concentrations by genotype or phenotype

Seven cohorts examined endoxifen concentrations in relation to *CYP2D6*; five included patients from the USA or Europe and two included patients from Asia. All seven cohorts reported lower endoxifen concentrations in PMs or those with the **10/*10* genotypes than in those with the *wt/wt* genotype (EM); pronounced decreases in mean endoxifen plasma concentrations were also evident in patients taking potent CYP2D6 inhibitors in two of these cohorts. Two cohorts of Caucasian patients reported conflicting findings with regard to concentrations for IMs, one reporting these to be closer to EMs and the other reporting them to be closer to PMs. Finally, one of the cohorts that also included patients taking an aromatase inhibitor [anastrozole (ANA)] reported that ANA concentrations were not affected by the combination with TAM but endoxifen levels were lower. Furthermore, the differences for endoxifen were no longer significant after excluding PMs.

Exploratory analysis

Because of the lack of convincing data for clinical validity, comparing EMs with other genotypes, an exploratory analysis for sensitivity and specificity was undertaken based on the limited number of studies ($n = 9$) that presented these data. Data suggested that the sensitivity of testing simply for the **4* allele in the adjuvant setting was 15% for OS and between 21% and 37% for relapse/recurrence. Specificity was calculated to be between 15% and 73% for OS and between 52% and 86% for relapse/recurrence. Utilising data from the only cohort to test simply for **10* suggested a sensitivity of 50% and specificity of 95% for recurrence/relapse. When a more comprehensive genotyping strategy was used, a sensitivity of 18% and specificity of 83% were calculated for OS, and, from two cohorts, sensitivity of between 18% and 30% and specificity between 86% and 88% for relapse/recurrence. It should be noted, however, that the exact same alleles were not genotyped in each of these two cohorts.

Economic evaluation

A total of 63 studies were identified from the literature search for evidence relating to the costs and benefits of *CYP2D6* genotyping for the management of women with breast cancer, but none of these papers met the inclusion criteria of being an economic evaluation comparing TAM with any aromatase inhibitors and genotyped for *CYP2D6*. However, two studies identified from the search have been discussed to help inform the development of future economic evaluations.

The lack of convincing data for clinical effectiveness, alongside other important parameter uncertainties, precluded the development of a de novo economic model, although a decision tree and a Markov model structure have been proposed. Crucially, the key points that do not allow us to populate the model are related to the undefined number of alleles to be tested, which alleles to test for, the lack of consensus about which test should be used, the lack of consensus about how to classify phenotypes and the heterogeneity around the results from the evidence found in the clinical review.

Discussion

From a number of individual cohorts, there is some suggestive evidence that genotyping for *CY2D6* may have a role to play in the management of women with ER+ breast cancer treated with TAM. Given six cohorts suggest EMs appear to have better outcomes than either PMs or PMs + IMs in terms of relapse/recurrence, this could translate to EMs being suitable candidates for TAM and PMs (and possibly IMs) being offered aromatase inhibitors instead, assuming the differences in relapse/recurrence outcomes between the two phenotypes are similar in magnitude to the differences found in studies comparing aromatase inhibitors with TAM. However, the suggestive evidence is taken from cohorts which, with two exceptions, are relatively small in number (≤ 500 patients). In addition, three cohorts report contradictory findings (albeit not statistically significant). Thus, the evidence must be treated with caution.

Much of the uncertainty in the clinical evidence is derived from the heterogeneity across the cohorts and around confounding prognostic factors within genotype groups. There are also differences in outcome definitions, alleles tested and the ways in which phenotypes are derived, making comparisons problematic. Additional uncertainties also exist around the role that CYP2D6 enzyme plays in the metabolism of TAM and, in particular, the relationship between endoxifen levels and clinical outcomes; our review failed to identify any studies that addressed this association.

Thus, given the lack of convincing evidence for clinical validity, our review did not identify any clinical utility studies or any full economic evaluations relevant to the UK. Given these deficiencies in the evidence base, we encountered a number of problems in attempting to develop and populate an economic model to address the cost-effectiveness of *CYP2D6* testing. Instead, we have begun the process of identifying the important parameters for which additional data will be needed to populate a model that includes the identification of the alleles to be tested, the available techniques, the sensitivity and specificity of these tests, the true costs of the tests, the provision of care that follows once women have been genotyped and the use of concomitant medication that can change the metabolism of TAM.

It is important to emphasise that the actual cost of pharmacogenetic testing is not known. However, test costs would form only a very small proportion of the overall costs of implementing pharmacogenetic testing into patient care pathways.

Conclusions

It has not been possible for this review to ascertain whether pharmacogenetic testing for *CYP2D6* is clinically effective or cost-effective. Key issues include the fact that it is not clear which alleles should be tested for and how phenotypes should then be derived. Assuming we are able to resolve these issues, there remain the uncertainties of how such testing would be implemented, in and impact on, the future pathways of care for these women.

Future studies will need to determine, as a minimum, the alleles that appear to be related to clinical outcomes and therefore need to be tested for. The link between a genotype and the patient response and ultimate clinical outcomes then needs to be determined in clinical utility studies.

The next uncertainty relates to how the pharmacogenetic testing should be carried out. Currently, there is one approved commercially available testing system and a number of bespoke tests being used, but it is not apparent what type of test would be relevant for a UK population. The final issues relate to the lack of evidence of the effectiveness of testing and mechanisms for integrating such testing into the care pathway for women with breast cancer and whether premenopausal and/or postmenopausal women should be targeted, what would be the likely uptake of pharmacogenetic testing and whether this would be mainly driven by clinicians or by patients.

The remit of this review was narrow and specifically examined the role of CYP2D6. Recent data suggest that the metabolism of TAM is complex and may be related to the effects of more than one genotype. It may be necessary, therefore, for future research to examine other metabolic pathways. In the meantime, further examination of the link between endoxifen levels and clinical outcomes could be of value and could be a mechanism that is easily integrated into existing care pathways.

Funding

Publication

NIHR Health Technology Assessment programme

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.