The implementation of pharmacogenetics: evidence and preferences

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Danielle Johnson

April 2022

Acknowledgements

Dedicated to my grandmother, Mary Smith.

I would like to take this opportunity to thank all those that have supported, taught, and helped me through this PhD process. Firstly, my funding comes from a PhD studentship funded by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2-R19). I want to extend my huge gratitude to my supervisors: Professor Andrea Jorgensen (University of Liverpool), Professor Dyfrig Hughes (Bangor University), and Professor Sir Munir Pirmohamed.

I want to particularly thank my primary supervisor, Professor Jorgensen, for her teaching, support, and encouragement over this PhD process. I truly could not wish for a better supervisor.

Additionally, I want to recognise Dr Emily Holmes (Bangor University) for the generation of a fractional factorial DCE design and help with DCE analysis. Dr Catrin Plumpton also aided with the analysis of my DCE.

I would like to thank those at Bilendi for their support distributing the DCE. At the University of Liverpool, I would like to thank Rob Sherman and Martin Bellwood for setting up the DCE randomisation system and for liaising with Bilendi to set up the online survey.

I would also like to take the opportunity to thank all those of the Department of Health Data Science, along with the University of Liverpool Computer Services Department, for all their support.

This work could not have been completed without the participation and insights of survey, focus group, and pilot study participants, and for this I thank them all.

I would like to personally thank all 4 of my parents for their love and support. Further thanks go to my PhD-having friends: Jade, Kayleigh, Emma, Chris, Will, Aimee, Richard, Nicky, and Javier for advice and much-needed encouragement. I also want to thank Amber, Emily, Martin, Pooja, and Tim, for their love and support, and of course, Sandra. Also all the dogs and cats in my life – too many to name here.

Finally, I want to thank Alex – without whose constant unwavering love and support I could not have completed this work. Thank you for everything, during the dark times and the good times, you are a constant source of strength.

Abstract

Pharmacogenetics has huge potential to transform the field of medicine and deliver personalised treatments to patients. However, its wider use is limited by many factors, particularly a lack of suitable evidence of efficacy or safety for regulatory approval and clinical use. The evidence required can be difficult to ascertain, presenting three main problems.

The first issue is that regulatory guidance for the evidence required is complex and varies greatly between different authorities and contexts. Guidance from the UK Medicines and Healthcare products Regulatory Authority (MHRA) and the US Food and Drug Administration (FDA) was reviewed along with criteria formulated by other industry and academic groups. It was found that there is a clear need for a unified set of standards for evidence gathering in pharmacogenetics. This was strengthened by an analysis of the evidence used by five different randomised controlled trials to justify the inclusion of their pharmacogenetic biomarker. Large variation in the quality and type of this evidence was found. These findings were used to make recommendations for future evidence gathering for trials, regulators, and journals.

Additionally, the evidence required for clinical implementation has traditionally been the prospective randomised controlled trial. Gathering information from two novel systematic reviews and meta-analyses of carbamazepine-induced Stevens-Johnson syndrome, it was shown how these sources of observational evidence can produce effect estimates and measures of clinical validity of greater precision than that of a prospective trial.

Finally, the level of evidence for a pharmacogenetic test that would be acceptable to the general public is not known. A discrete choice experiment (DCE) was designed to quantify these views. The first step was a systematic review of existing DCEs in this area, to extract useful information from these to inform the work. An extensive programme of qualitative work with healthcare professionals, patients, and the general public then further informed the design of this novel DCE. Participants were randomised to complete one of eight DCEs in different disease areas, with either a ‘high’ evidence scenario or a ‘low’ evidence scenario described. Launched in May 2021, over 2,000 responses were collected and the results were analysed in preference-weighted utility models. Although there was no difference in utility between ‘high’ and ‘low’ evidence tests, several important insights were generated

(particularly in regard to data sharing and privacy) that will potentially have large impacts on policy in this area.

Table of Contents

Chapter 1: Introduction

1.1 Personalised medicine

What cures one person may do nothing for another. A drug that causes unpleasant side-effects in one person may cause no issues for another. The drug that vastly improves one person's quality of life might harm or even cause death in another person. This variability in responses has been known since at least the 1950s (1). Medicine has attempted broad measures to predict and modulate these effects. Moderating drug dosages based on kidney and liver function is common practice in many drug prescriptions (2-4). However, this approach is something of a blunt instrument: a lower dose may reduce the risk of harmful side-effects, but it may also drastically reduce drug efficacy. This approach also does not cover dose-independent side-effects, which are much less predictable. What is needed is a way to tailor drugs to each individual's unique biology.

This is termed personalised, precision, or stratified medicine, often defined as "the right drug for the right person at the right time" (5-7) . These new approaches to clinical care describe the tailoring of medical treatment to the individual characteristics of each patient (8). It does not usually refer to unique treatments designed for individuals, but the classification of patients into subpopulations for treatment (8).

This can include individualised treatment approaches based on genotype, known as pharmacogenetics or pharmacogenomics (9-13) where genetic variations are used to predict the best drug, or the optimal dose of a drug to prescribe (13). This is in contrast to traditional, empirical medicine, where the same drug, at similar doses, are given to all patients with the same condition (12) (Figure 1.1). Stratified medicine, somewhere between the two, uses characteristics (such as genotypes or molecular profiles) to stratify patients into groups that inform treatment. However, while stratified and personalised medicine are technically distinct terms, it is important to note that they are often used interchangeably in practice (14).

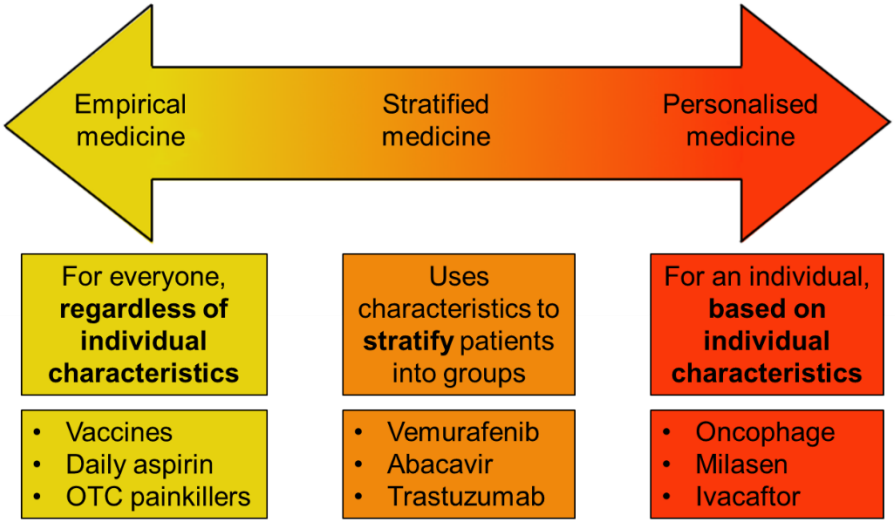


Figure 1.1 -The scale of medicine specificity moves from empirical to personalised medicine, with stratified medicine as a mid-point (12). Examples of drugs that are prescribed in each way are listed in bullet points. OTC = over the counter

Pharmacogenetics and pharmacogenomics both describe the process of using genotype data for risk assessment and guiding treatment, which may then impact on a patient's prognosis (4, 9, 15-17). A distinction has been drawn between the terms. Pharmacogenetics may be understood as the focus on a single or small number of gene(s), while the scope of pharmacogenomics includes many genes or a whole genome (18). These terms have also been distinguished by Møldrup (2001), who argued that pharmacogenetics refers to "the study of drugs based on known allele-specific genetic variations", while pharmacogenomics is the "identification and elucidation of genetic variations which will impact on the efficacy of drugs or offer different targets" (19). While acknowledging that the terms may be used interchangeably, Williams-Jones and Corrigan (2003) defined pharmacogenomics as a "broad-based pharmaceutical industry-led initiative", in contrast to the "narrower spectrum of inherited differences in drug metabolism and disposition linked to individual genetic variations" of pharmacogenetics (20). The definitions used by the Food and Drug Administration (FDA) of the United States hinge on the type of genetic material used - pharmacogenetics is considered to use deoxyribonucleic acid (DNA) as a biomarker, while pharmacogenomics uses both DNA and ribonucleic acid (RNA) (21). DNA resides in cell nuclei and contains the instructions for protein synthesis. RNA is similarly structured to DNA but is used for transcription of synthesis instructions from DNA to ribosomes and ribozymes (22, 23). The European Medicines Agency (EMA) defines pharmacogenetics as the study of interindividual variations in DNA sequence related to drug response, and pharmacogenomics as the broader

science of genes relevant to disease susceptibility and drug response (24). Interindividual variation

Regardless of the term used, this field has enormous potential to revolutionise medicine. This promise has already been realised in several disease areas. In HIV, genetic testing has near-eliminated severe adverse drug reactions (ADRs) to the drug abacavir (25-27). This reaction, the abacavir hypersensitivity reaction (AHS) causes fever, rash, gastrointestinal, and respiratory symptoms, and occurred in approximately 3.7% of patients (28, 29).

In cardiovascular medicine, personalised dosing for warfarin can improve drug efficacy, reduce the risk of ADRs and reduce the burden of multiple blood tests on patients (30-32). In addition, a personalised approach offers potential improvements in cost-effectiveness for funders of healthcare. For example, with fewer ADRs, fewer medical interventions are required for the patient population (33-35). An improvement in patient quality of life is also important when discussing cost-effectiveness (see below) (36).

The pharmaceutical industry has also seen a major paradigm shift away from ‘blockbuster’ drugs (those generating more than $1 billion in revenue per year (37)) to ‘patient-centrism’ (38), leading to renewed interest in personalised medicine. Often, blockbuster drugs treat conditions that are common, such as sildenafil for the treatment of erectile dysfunction (39). However, the newer and more personalised breed of blockbusters include the targeted cancer immunotherapy pembrolizumab (40).

Regulatory authorities have also had to adapt to this shift. As of June 2021, there were 409 drugs with listed pharmacogenetic information on their drug labels on the Pharmacogenomics Knowledgebase (PharmGKB) (Figure 1.2) (41, 42). These range from the strongest recommendations (Testing is required before using the drug) to markers that are used in an exploratory or informative basis. The FDA has the highest number of these approvals. This can be compared to 5 years prior, when 208 drug labels were listed, with less information available for each (43).

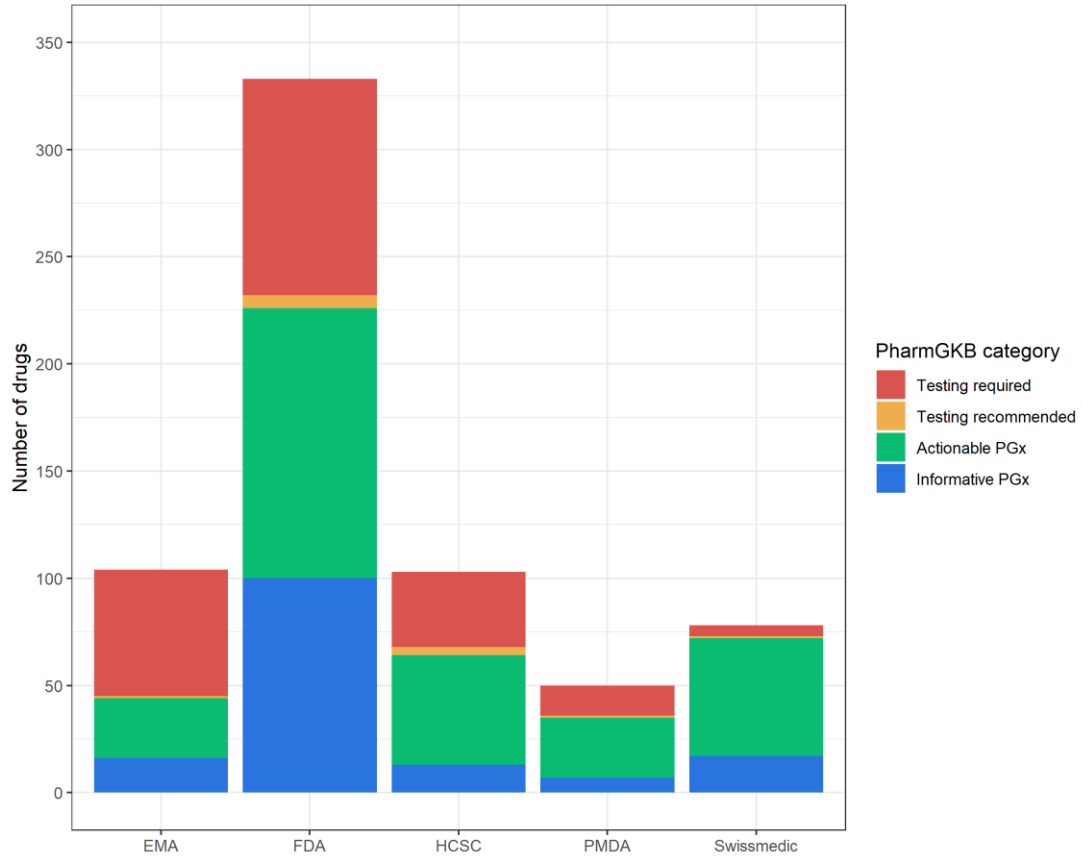


Figure 1.2 - PharmGKB categories of pharmacogenetic testing, by regulatory agency. Full descriptions of each category are available here: https://www.pharmgkb.org/page/drugLabelLegend#pgx-level . EMA = European Medicines Agency. FDA = Food and Drug Administration (USA). HCSC = Health Canada (Santé Canada). PMDA = Pharmaceuticals and Medical Devices Agency (Japan). Swissmedic = Swiss Agency of Therapeutic Products. PGx = pharmacogenetics.

The 'moment of critical impetus' for pharmacogenetics was the 1998 FDA approval of the breast cancer drug trastuzumab (Herceptin) and its companion diagnostic test HercepTest (44, 45). Trastuzumab treats HER2+ breast cancer, a marker overexpressed in 20-30% of breast cancers (46-48) (discussed further in 1.1.3.1 Cost and cost-effectiveness below).

Since then, the number of similar approvals has soared. The Clinical Pharmacogenetics Implementation Consortium (CPIC), which publishes freely available gene-drug guidelines, has completed 25 gene-drug guidelines (as of October 2020), pooling information from several national regulatory agencies and the literature (49)

In 2019, a sample of UK primary care patients aged 50 and over found that the majority (58%) had been prescribed at least one drug with a pharmacogenetic dosing guideline over a 1-year period, rising to 80% over a 20-year period (50).

There was a continual increase in the use of these drugs, and the authors predicted that this would continue to rise in the future. A 2021 study also found that around 20% of all new prescriptions in a community pharmacy over one year had an actionable drug-gene interaction (51). Recent work in Denmark with younger participants also highlights the increasing use of pharmacogenetic drugs, particularly in the field of mental health (52).

1.1.1 What is a biomarker?

The 2001 Biomarkers Definitions Working Group defined a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic response to a therapeutic intervention (53)". This definition is widely accepted. Pharmacogenetics, the focus of this thesis, concerns the use of genetic variants as biomarkers relating to drug response (15, 54-56). While these are the most widely studied biomarkers, others include proteins, imaging data, and epigenetic changes (44, 57).

Genetic mutations (used as biomarkers) can be germline or somatic. Germline mutations occur in eggs or sperm and are passed on to the next generation. Individuals are born with germline mutations. Somatic mutations occur in single cells in any tissue, and are not passed onto the next generation (58).

Biomarkers can be sub-categorised by their function (Figure 1.3) (44, 59-61). These often-overlapping functions enable pharmacogenetics to impact many fields. These include reducing ADRs, improving drug efficacy, diagnostics, economics/cost-effectiveness, and drug development (17). The FDA divides biomarkers into diagnostic, where they identify patients with a particular disease or disease subset; prognostic, where they indicate future clinical course of a disease irrespective of treatment; predictive, where they identify patients likely to respond in a particular way to a treatment; and response, where they indicate a biological response after an intervention (60).

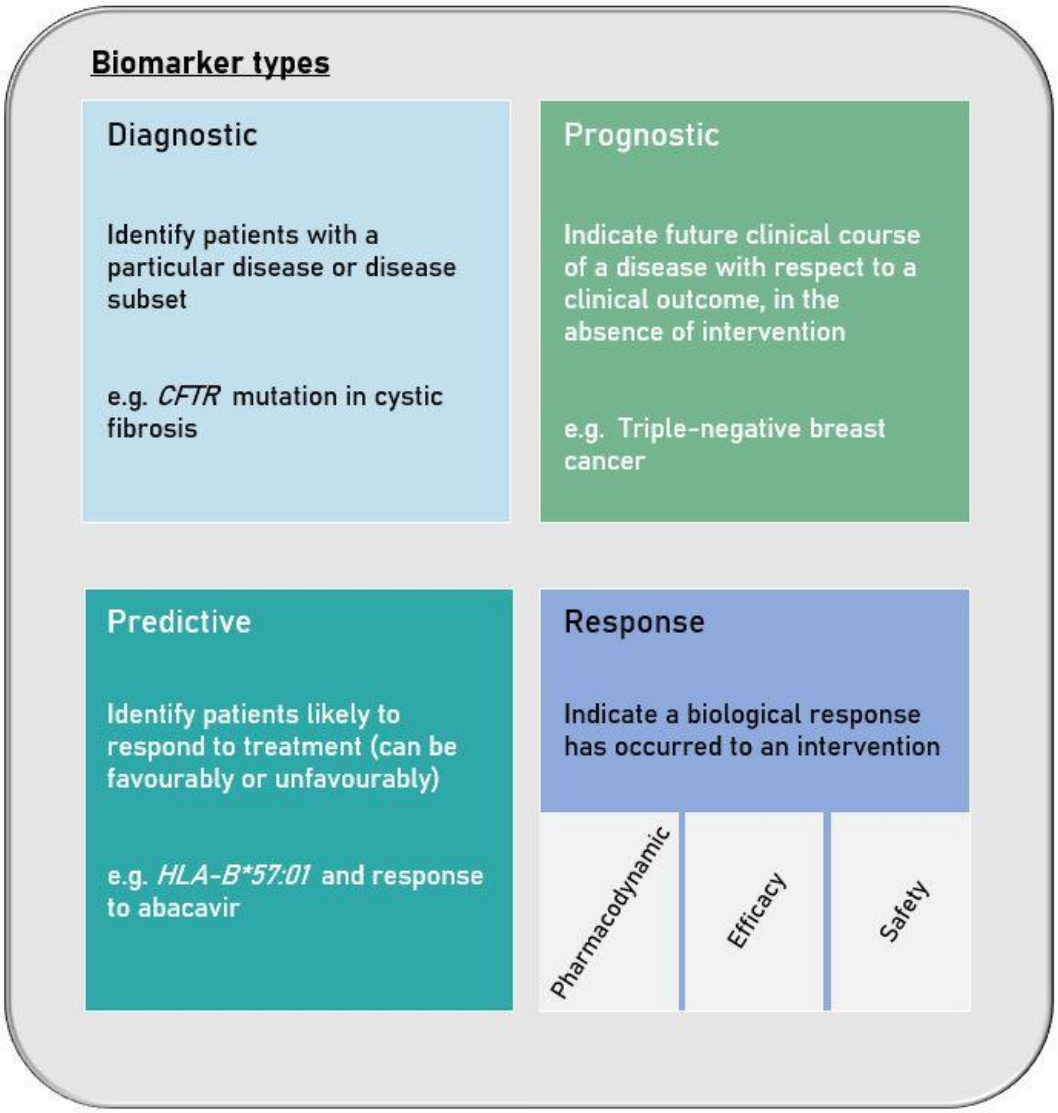


Figure 1.3 - biomarker categories as defined by a meeting of stakeholders in industry, regulatory bodies, academia (60). CFTR – cystic fibrosis transmembrane conductance regulator. HLA-B – human leukocyte antigen B.

Biomarkers must be also distinguished from surrogate endpoints, which are defined as a subset of biomarkers (62-64). Surrogate endpoints are laboratory or physical signs used in trials in place of clinical efficacy measures (62, 64, 65).

The use of biomarkers improves patient outcomes and prognoses by reducing their risk of ADRs, and improving drug efficacy (through more efficient dosing, or choice of drug). I now present two case studies of these applications.

1.1.2 Benefits of pharmacogenetics

1.1.2.1 Reducing the risk of ADRs

The World Health Organisation (WHO) defines an ADR as “a response to a medicine which is noxious and unintended, and which occurs at doses normally used in man” (66). ADRs are the cause of 6.5% of UK hospital admissions, and 14.7% of extended hospital stays, costing the NHS an estimated £466 million a year (67-69). A 2021 meta-analysis estimated a prevalence of ADRs of 8.32% (95% CI 7.82 – 8.83%) among 1,568,164 patients across 12 countries (70).

Many clinical trials have already taken place with the aim of using genetic testing to reduce ADRs (71-76). Arguably the most successful example of the integration of pharmacogenetics into clinical practice is the case of the antiretroviral drug abacavir, used for HIV treatment (34, 77, 78). Abacavir (Ziagen) was discovered in 1997 (79) and approved for use by the FDA in 1998 (78, 80, 81). It is used in combination with other drugs to treat HIV infection (77, 80).

During Phase II trials of abacavir, it was observed that 3-4% of patients suffered a specific adverse reaction to the drug (82, 83). A 2001 review reported the incidence of this reaction among clinical trial participants to be 4.3% (82). The reaction, AHS, is characterised by symptoms including: fever, chills, rash, nausea, vomiting, and fatigue, and can be life-threatening (83, 84). AHS can be diagnosed clinically and confirmed with a skin patch test (76). Initially, it was difficult to predict which patients would suffer from AHS in response to abacavir treatment.

Susceptibility to AHS is strongly associated with the presence of the *HLA-B*57:01* allele (variation of the *HLA-B* gene). This link was first reported in 2002 (85-87). This association (odds ratio [OR] 859.1 [95% CI 189.2 – 3901.4], $p < 0.001$ in one meta-analysis (88)) is “one of the strongest associations ever described between a genetic marker and a disease” (89). The prevalence of *HLA-B*57:01* is highest in Caucasian patients, and lowest in Black patients (90). However, the specificity of the screening test is still comparable between these groups (76).

With the link discovered, clinical trials could then take place to investigate the effectiveness of testing patients for the presence of the *HLA-B*57:01* allele prior to commencing abacavir. The large PREDICT-1 randomised controlled trial (RCT) randomised patients to receive abacavir with or without genetic testing for *HLA-B*57:01* (85). The incidence of AHS was significantly reduced in the tested group, with a calculated OR of 0.03 (95% CI 0.00 – 0.18, $p < 0.001$). The negative predictive

value of screening was found to be 100% - meaning that a patient who tests negative will definitely not develop AHS.

Other investigations include the SHAPE and ARIES studies. SHAPE was a retrospective case-control study that matched white and black AHS patients with racially similar patients who tolerated abacavir (76). White patients were more likely to develop symptoms earlier in the course of treatment. Symptoms also differed between groups, with fever and gastrointestinal symptoms the most reported symptoms in white and black patients, respectively. *HLA-B*57:01* was significantly associated with AHS in both white (OR 1945 [95% CI 110 – 34352]) and black (OR 900 [95% CI 38 – 21045] participants. The sensitivity of *HLA-B*57:01* screening was 100% in both groups, indicating the value of testing patients regardless of racial background.

The ARIES trial was an RCT comparing two common HIV regimens, including abacavir (91). Within this trial, Young, *et al.* recruited 517 *HLA-B*57:01* negative patients to be evaluated for AHS over the 30-week study period (92). Less than 1% of these patients developed AHS symptoms, providing further evidence for the utility of *HLA-B*57:01* as a marker of AHS.

Testing was recommended as a cost-effective measure in the UK in 2004 (93). Today, British National Formulary (BNF) guidelines require testing before commencing abacavir in every patient where *HLA-B*57:01* status is unknown (25). Screening was recommended by the FDA in 2008 (80, 94), with the cost-effectiveness shown in a later study (95). Today, both the UK and US labels for abacavir include a boxed warning on the importance of *HLA-B*57:01* screening prior to abacavir use (81, 96). Screening is also recommended by the CPIC (26).

1.1.2.2 Improving drug efficacy

Matching patients with drugs most likely to provide benefit is a key offering of pharmacogenetics. A drug's clinical efficacy is a measure of clinical disease improvement after administration of the drug (97). Many drugs have highly variable rates of efficacy (7) (Table 1.1).

Condition	Efficacy rate (%)
Alzheimer's disease	30
Asthma	60
Diabetes	57
Hepatitis C virus	47

Oncology	25
Osteoporosis	48
Rheumatoid arthritis	50
Schizophrenia	60

Table 1.1 – Broad efficacy rates for drugs in different conditions, adapted from Pirmohamed and Lewis (2004) (7), from data in Physicians' Desk Reference (2000)

Drug response is usually a complex process, influenced by polymorphisms (variations in the population) in multiple genes (17). Drug response is also influenced by external factors, including diet, adherence, and individual health status (15). It is also important to note that attributing a phenotype or function (including efficacy or ADRs) to a particular polymorphism does not prove causation (17). Even with these caveats in mind, there are clear examples of benefits to drug efficacy with pharmacogenetics. A 2015 meta-analysis found that personalised therapies were associated with longer overall survival in cancer trials compared to non-personalised therapies (13.7 vs 8.9 months, $p < 0.001$) (98).

A successful example of this is the oncology drug vemurafenib (Zelboraf), a kinase inhibitor used to treat cancers with the *BRAF* V600E mutation (Figure 1.4) (99-103). This mutation is seen in around 50% of melanomas, and at lower frequencies in other cancers (99, 100, 104-109). Vemurafenib was first used in a Phase I trial in 2010. This trial initially recruited 55 patients with any cancer, then launched an extension that recruited only patients with confirmed *BRAF* V600E mutations (110). In this second group, 81% of patients had a response, with two patients experiencing complete responses. A 2012 Phase II trial in 132 *BRAF* V600E metastatic melanoma patients found a median overall survival of 16 months, and a progression-free survival of 6.8 months (101). A subsequent Phase III trial enrolled 675 metastatic melanoma patients with *BRAF* V600E mutations (111). Patients were randomised 1:1 to receive either vemurafenib or dacarbazine (a chemotherapy drug). Those on vemurafenib had significantly better outcomes, with a hazard ratio of 0.37 for death compared to dacarbazine patients (95% CI 0.26 to 0.55, $p < 0.001$) (111). A more recent study in patients with and without this specific mutation had to stop recruitment early in the non-*BRAF* V600E arm due to lack of efficacy in this group, compared to the *BRAF* V600E positive group (112).

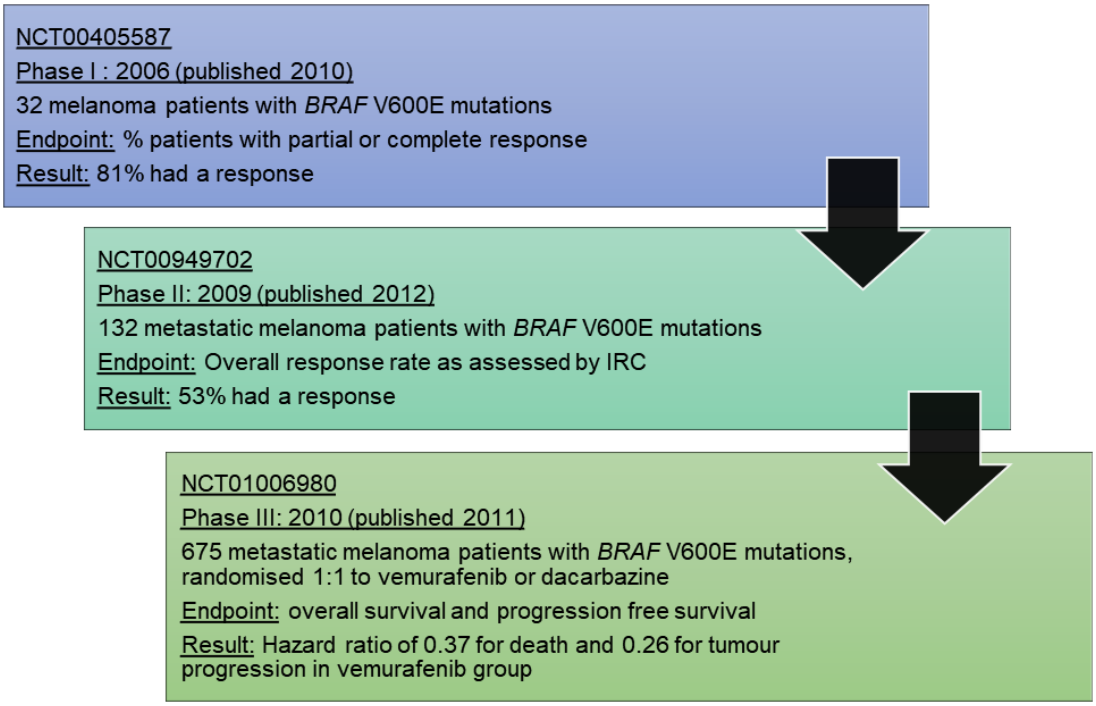


Figure 1.4 - development of vemurafenib through selected Phase I, II and III trials. Phase I (NCT00405587) took place in 2006 (110), Phase II (NCT00949702) in 2009 (101), and Phase III (NCT01006980) in 2010 (111). IRC = independent review committee

Vemurafenib is extremely selective for tumours with the *BRAF* V600E mutation (103). Additionally, vemurafenib has been shown to promote tumour growth in patients with wild-type *BRAF* tumours (102), and has been associated with acute kidney injury (113, 114). Therefore, it is essential that clinicians are able to accurately detect which patients will benefit before prescribing. The companion diagnostic test for the *BRAF* V600E biomarker was developed by Roche in 2005, prior to clinical testing of vemurafenib (99, 115). Development was made more difficult since melanin, found in melanomas, inhibits the DNA polymerase enzyme used in testing (115, 116). The assay has a correct mutation call rate of over 96% and is additionally able to detect other *BRAF* mutations (V600D and V600K) (99, 115).

Vemurafenib was also included in the SHIVA trial that randomised 195 patients with any cancer type to molecularly targeted treatments, or to standard of care (physician's choice) (see Chapter 3) (117). Vemurafenib has also been tested in trials in combination with other targeted agents, including cobimetinib (118, 119), ipilimumab (120), and irinotecan/cetuximab (107).

The FDA approved vemurafenib for use in metastatic melanoma on the condition that the *BRAF* V600E mutation is detected by an FDA-approved test (102). There are currently two of these tests available for use (121). In the UK, vemurafenib was

approved for use in 2012 for the treatment of unresectable or metastatic *BRAF* V600E positive melanoma (122).

1.1.2.3 Drug development

These kinds of trials, essential for the regulatory and clinical approval of a biomarker, can be very expensive. Adams & Brantner (2009) calculated that the average expenditure per drug on human clinical trials is $27-74 million a year (123). Drugs can later be withdrawn, or their use restricted based on variable efficacy or ADR risk. While pharmacogenetic trials can be expensive, pharmacogenetics can also be used by pharmaceutical companies to 'de-risk' drug development by enhancing target identification, clinical testing, and drug safety (124-126). Pharmacogenetics has utility at every phase of drug development and testing (Figure 1.5).

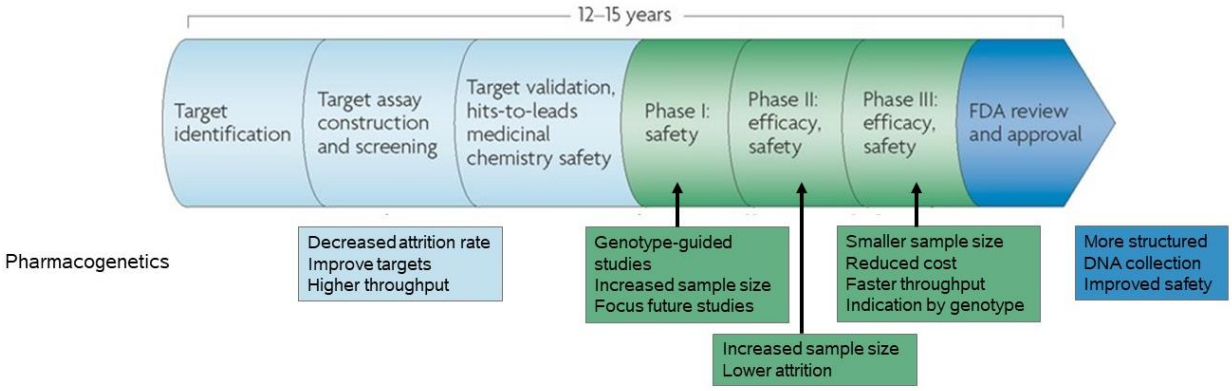


Figure 1.5 - Pharmacogenetics during the drug development timeline. Produced from information in Roses (2008) (126) and Pirmohamed and Lewis (2004) (7)

A 2019 report on precision medicine moving into clinical applications stated the need for the drug development pipeline to be accelerated using pharmacogenetics, including the importance of deep phenotypic characterisation (the precise and comprehensive analysis of phenotypes) (127). This is important for accurately classifying patients into subpopulations for personalised and precision medicine. The authors also discussed the need for including diverse populations in genomics research (127).

For example, a meta-analysis of genome-wide association studies focussing on HbA1c (used to monitor diabetes) found that the rs1050828 mutation in the *G6PD* gene was associated with significantly lower HbA1c levels in patients of African American ancestry (128). As raised HbA1c is used as a marker for diabetes, the authors calculated that 650,000 adults in the US may have diabetes but will be missed by HbA1c screening since their levels are kept low by their mutation (128). This study shows the need for investigators to consider diverse populations in genome-wide association studies. It is also valuable to developers of drugs to treat diabetes. If *G6PD* screening became widely available, a large additional market for diabetes drugs would open up.

A further example is the process of salvaging drugs, or 'drug rescue' (7, 129-131). This occurred with abacavir, the HIV drug that can cause hypersensitivity reactions (see above). This drug, effective for HIV treatment, may have been withdrawn from market due to its associated risks if not for the discovery of the link between hypersensitivity and *HLA-B*57:01* (85, 129). Potential patients can now be tested before prescription, and abacavir use limited to those negative for *HLA-B*57:01*.

Despite the advantages of including pharmacogenetics in drug pipelines, uptake has been slower than expected (130). Barriers to implementation in this setting include limited knowledge of genotypes, limited technologies, and financial hurdles (130). However, a 2016 survey of industry perspectives on biomarker qualification found that regulatory complexity was a greater challenge to qualification than scientific or technical complexity in the case of most biomarkers (132). These difficulties have also been acknowledged by the FDA (133).

> *"a recent survey looked at the drug pipeline portfolios of about 20 major companies and showed that a very small minority included what could be called stratified or personalized medicine. When you ask people in industry about this, some feel it is because there's no clear regulatory pathway or guidelines." - The director of the FDA Office of Clinical Pharmacology, Issam Zineh, 2016 (133).*

These regulatory issues will be explored further in Chapter 2.

1.1.3 Issues in pharmacogenetics

While there are clear benefits to using pharmacogenetics, there are also issues around cost, implementation, and ethics to consider. Finally, it is morally, ethically, and clinically essential to consider the preferences of patients, healthcare professionals, and the general public regarding the wider use of pharmacogenetics.

1.1.3.1 Cost and cost-effectiveness

New technologies are generally associated with extra costs (134, 135), and genetic biomarkers are no exception (35, 135). Both benefits and costs need to be considered when evaluating a new intervention or biomarker for reimbursement or adoption by a health service provider/payer (136). A technology that is more beneficial and less costly than an alternative would usually be accepted (depending on cost, disease prevalence, practicalities, etc). Conversely, an intervention that costs more and is less beneficial than the alternative should always be rejected. The difficulty lies in deciding for interventions that sit between these extremes. An increase in cost may be acceptable where there is a certain level of increased benefit. Setting this level can be complex.

There are four possible outcomes of a cost-effectiveness analysis, two of which require payers to make a judgement (Figure 1.6) (137).

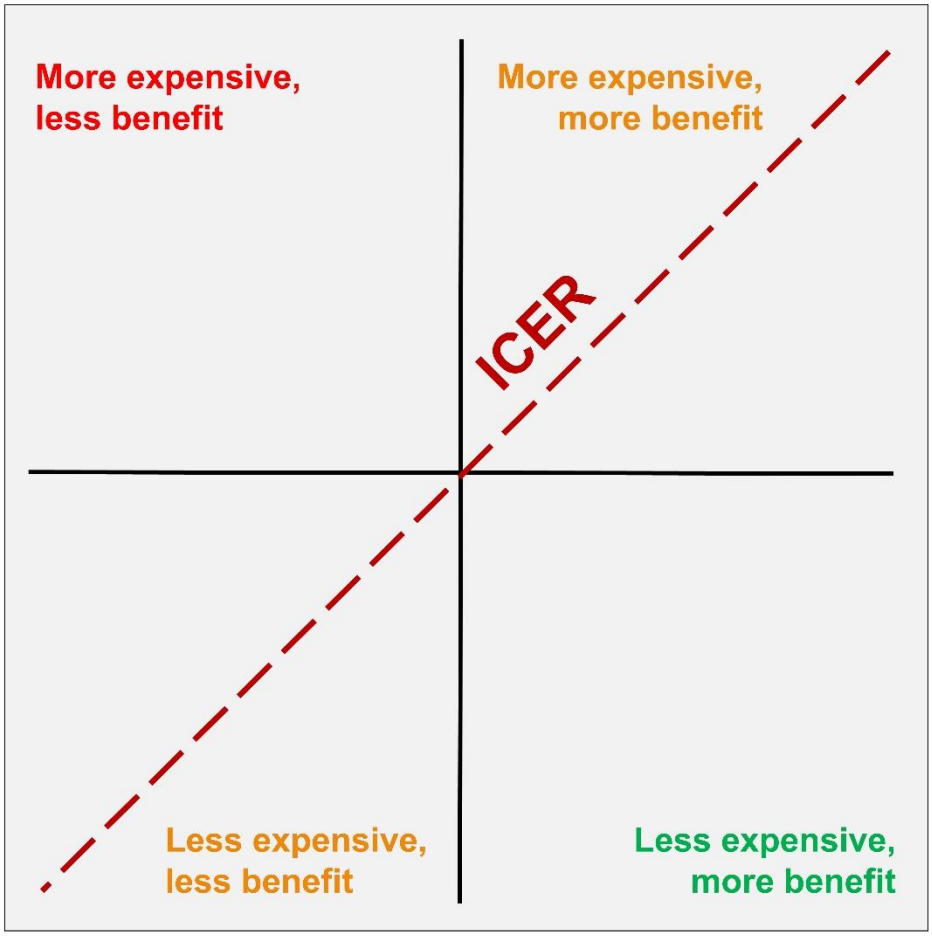


Figure 1.6 – The four outcomes of a cost-effectiveness analysis. Based on Figure 7.1 of Walley, et al. (2004) (136). The top left quadrant denotes an intervention that should definitely be rejected since it is more expensive and has less benefit than the comparator intervention. Conversely, the bottom right quadrant denotes an intervention that should normally be accepted since it is less expensive and delivers greater benefits. The top right and bottom left sectors require a judgement to be made using the incremental cost-effectiveness ratio (ICER).

The difference between the cost of an intervention and an alternative programme is known as the incremental cost. This can be divided by the incremental benefit (the difference between the benefit of the intervention against an alternative). The incremental cost per unit of benefit gained is known as the incremental cost-effectiveness ratio (ICER) (Equation 1.1) (136, 138).

$$Incremental\ cost\ effectiveness = \frac{(cost\ of\ drug\ A - cost\ of\ drug\ B)}{(benefits\ of\ drug\ A - benefits\ of\ drug\ B)}$$

Equation 1.1 – Incremental cost effectiveness ratio for hypothetical drugs A and B.

Quality-adjusted life years (QALYs) are one measure commonly used to evaluate the benefit of an intervention. For an intervention to be accepted, the incremental costs

per QALY gained should not exceed a pre-determined value assigned to one QALY (137). In the UK, this value is normally £20,000 to £30,000 (139). In the USA, this value has been quoted as up to $150,000 per QALY (140).

However, precision medicine has 'unique economics' in regard to innovation, pricing, diagnostics, and access (141). For example, if a company can market a drug to a subgroup of patients (identified by a biomarker) in which it will be most effective, they can demand higher prices to reflect this higher drug efficacy (141). Trastuzumab (Herceptin) is an example of this, becoming a blockbuster drug despite only being efficacious in a subset of breast cancer patients (12, 142).

Trastuzumab is a monoclonal antibody targeting HER2, a tyrosine kinase that mediates cell growth, differentiation, and survival (143, 144). HER2 is overexpressed in 20-30% of breast cancers (48) and is associated with shorter overall survival (145). Trastuzumab was first developed in 1991 by Genentech (146), and has since become one of the top 10 bestselling drugs in the world (44, 147). It is only indicated in $HER2^+$ breast cancer (47). A 2011 meta-analysis found a significant benefit to overall survival of trastuzumab in early-stage breast cancer (OR = 0.78, 95% CI 0.69- 0.88, $p < 0.001$) (148). The UK's National Institute of Health and Care Excellence (NICE) approved trastuzumab for use in $HER2^+$ breast cancer in 2006, citing ICERs of £4461 to £32,701 per QALY gained (149).

Conversely, an analysis of *HLA-B*15:02* screening prior to carbamazepine prescription in Hong Kong found that while screening in an 'ideal situation' would be cost-effective ($11090 per QALY gained), the actual situation in which screening was taking place led to it becoming a very expensive screening programme ($85697 per QALY gained) (150). The study found that this was mainly due to decreased use of carbamazepine, as clinicians switched to drugs that did not have a screening mandate, but which were more expensive and had little additional benefits. Notably, these drugs also placed patients at risk of serious ADRs, albeit at lower rates than carbamazepine. This led to no overall reduction in the rates of serious ADRs in patients with epilepsy treated with anticonvulsants, and higher than expected costs per QALY. Poor adherence to the guidance to use genetic testing prior to prescribing carbamazepine was also a factor. This case shows the importance of including a full, realistic economic evaluation in the assessment of biomarkers.

Other principles from economics, namely stated preference and discrete choice experiments (DCEs), can be used in health research (151, 152). This is explored more fully in Chapter 5.

1.1.3.2 Implementation

Pharmacogenetics is not as widely used as it has the potential to be. In 2008, Hong Kong implemented a policy of screening patients for *HLA-B*15:02* before commencing carbamazepine, to prevent the serious ADR Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN). Researchers in Hong Kong investigated the effect this implementation had on rates of screening and of carbamazepine-induced SJS/TEN (153). While rates of carbamazepine-induced SJS/TEN fell significantly ($p = 0.027$), the overall rate of SJS/TEN remained steady. This occurred because physicians were instead prescribing other antiepileptic drugs that did not have a genetic testing mandate. Cases of SJS/TEN caused by these other drugs then rose. Prescriptions of carbamazepine decreased by 81% once the policy was implemented.

A similar case took place in Taiwan, where *HLA-B*15:02* screening was mandated from 2010 onwards (154). Over a 10-year study period (2005-2014), the number of carbamazepine-related ADRs fell by 87.1%. However, the number of new users of carbamazepine fell by 82.6%. The use of other antiepileptic drugs increased after screening was mandated, but the authors did not collect data on any subsequent increase in SJS/TEN related to these. The study demonstrated that there are also medical educational barriers – more specialised physicians working in larger medical centres were more likely to utilise pharmacogenetic screening (154).

These cases show that human behaviour must be accounted for when implementing pharmacogenetics. I have explored this issue further in my DCE, in Chapters 6 and 7. However, this is far from the only reason why pharmacogenetic testing is lagging behind its potential. There are significant regulatory hurdles (1, 68, 155), technological barriers, and practical issues (particularly in low- and middle-income countries) (1, 68, 156-159). One of the largest issues is the lack of well-controlled trial evidence to justify the implementation of pharmacogenetics (156, 160). Regulatory bodies normally require RCT evidence as a minimum to accept a new test or intervention (161). However, a certain standard of evidence needs to be met before an RCT can be performed. This is to ensure the safety of trial participants and efficient use of resources. The issue is that this standard is ill-defined and very different standards have been met by different approved biomarkers (155). I have explored these issues further in Chapter 3.

1.1.3.3 Ethical issues

Pharmacogenetics should be understood in terms of both benefits and risks. As discussed, the potential benefits are many: reduced incidence of adverse events, increased treatment efficacy, and the identification of better targets for drug development. However, there are several issues surrounding the use of genetic biomarkers that need to be addressed (15).

Genetic information does carry some unique risks. Testing an individual is not truly individualistic. Relatives who share genetic information are inadvertently also tested, often without consent (162). Confidentiality and privacy are consistently rated as important issues in patient surveys of pharmacogenetics (163-165). Clinicians are also concerned about these issues. In one survey, the majority of clinician participants were 'very worried' that patients would be disadvantaged for future health insurance based on genetic test results (163).

These concerns are linked to the stigma and discrimination that may occur in response to some genetic test results. Genetic discrimination is the unjust or prejudicial treatment of people based on their genetics (166, 167). Schizophrenia is an example where genetic testing is emerging as a technology. However, mental illness carries a large stigma and treatment options are often limited. The discrimination that could result from knowledge of someone's risk of schizophrenia is high – including health insurance raising prices to cover costs of psychiatric care, and banks refusing loans fearing an individual's capacity to repay (168). In another study, more participants were concerned about insurance companies, the government, or their employers knowing results of genetic screening for Alzheimer's disease than results of screenings for cancer risk (169). This may be because Alzheimer's is likely to be incurable and more expensive for insurers (and disagreeable for employers), providing a motive for genetic discrimination.

There is legislation to protect against genetic discrimination. The UK follows European law, including the 1997 Convention on Human Rights and Biomedicine. The 2016 recommendation CM/Rec(2016)8 from the Council of Europe ensures that insurers cannot require people to undergo genetic tests for insurance purposes. The recommendation also specifies standards for the processing and storage of personal genetic data (170). The Association of British Insurers has published a Code on Genetic Testing and Insurance that states that companies within the association cannot, under any circumstances, require or pressure customers to have a predictive or diagnostic genetic test (171).

In the US, the 2008 Genetic Information Nondiscrimination Act (GINA) was enacted to prohibit discrimination by insurers or employers based on any genetic information (166, 172). However, some effects are currently outside of the law. GINA does not apply to the US military, where DNA is routinely collected from service personnel (173) There have been cases where adoption agencies have refused to place children in families where one or more parents are at risk of Huntington's disease (174-176) . Privacy is clearly an important part of genetic testing to be considered.

A further ethical dilemma arises when the testing technology outpaces available treatments. A 2020 study of molecular testing by the NHS in advanced lung cancer found that 83% of patients underwent somatic molecular testing for all three recommended biomarkers (*ALK*, *PD-L1*, and *EGFR*), and 96% of these tests yielded useful results (177). Testing for these biomarkers is mandated since there are approved drugs that target these mutations. However, this is not always the case. Direct-to-consumer (DTC) tests provide consumers with at-home germline genetic testing (178). These variants often include germline mutations linked to currently incurable and largely non-preventable conditions, such as Parkinson's and Alzheimer's disease (179). Without the benefits of genetic counselling, persons receiving this information may be at risk of harm (178, 180). Additionally, these tests are often not fully validated, with false positives and negatives occurring at high rates (181). However, these tests are extremely popular, with over 26 million people having been tested as of January 2019 (182). This shows that there is clearly a demand for genetic testing. The challenge is to ensure that future pharmacogenetic testing using clinically validated and approved tests aligns with the preferences of patients, in order to guarantee maximum uptake rates and clinical utility. I have chosen to investigate this issue using a discrete choice experiment, as demonstrated in Chapters 6 and 7.

While there are many benefits, pharmacogenetics must not appear to have all the answers. Genetics is not the only factor that can influence a drug's efficacy. Age, sex, compliance, and environmental effects may also contribute to an individual's response to a drug (20). It is difficult to unravel the complete picture of drug response.

However, we must simultaneously take care not to be overly cautious. It has been suggested that genetic tests are more harshly judged than tests using other technologies (4, 15, 183). This 'genetic exceptionalism' has arguably slowed progress in pharmacogenetics and genetic biomarker research.

1.1.3.4 Preferences

Incorporating the preferences of stakeholders (including patients, healthcare professionals, and the general public) is essential if pharmacogenetics is to be accepted as a part of clinical practice.

A 2012 study of 1463 US residents found that 70-85% of participants would take free genetic tests predicting their risk of future disease (169). When willingness-to-pay (WTP) was calculated by adding a price tag, mean prices of between $320 - 622 were obtained. Although mainly applicable to the US healthcare system, this study showed that the general public are amenable to genetic testing, even when that might provide upsetting news. Despite this acceptance, medical education in pharmacogenetics is usually lacking. A recent survey of 282 US and Japanese paediatricians found that less than half could correctly answer a pharmacogenetics quiz question. Less than 10% of participants said they were 'very familiar' with pharmacogenetics. Despite this, 82% considered pharmacogenetics a valuable tool in improving drug efficacy (184).

The characteristics of an ideal genetic test are less clear. This makes policy decisions in this area even more challenging. The scenarios of testing are often complex and involve difficult decisions. One method often used to measure preferences for test characteristics is the DCE. This method, from use in economic theory, is ideal for the quantification of complex scenarios (185). For example, Dong, *et al.* (2016) measured the preferences of patients for genetic testing to prevent very rare ADRs (186). Modelling revealed that the majority of patients always wanted a genetic test, regardless of cost or other factors. This could potentially impact health policy, if providers know that genetic testing is broadly very important to the general public, even in the context of a 1-in-a-million chance ADR. There have been several uses of DCEs in pharmacogenetics, and I have explored these further in Chapter 5.

1.2 Conclusion

What is clear is that for pharmacogenetics to move forward, the views of all stakeholders must be collected, analysed, and incorporated into policy. A practical method for evaluating views on complex subjects is the DCE (164, 187, 188). This has been used successfully in many areas of pharmacogenetics and is discussed in detail later in the thesis in Chapters 5, 6, and 7.

The aim of this thesis is to explore the evidence base of pharmacogenetics, with a focus on quantifying the views of the general public in this area. I will first focus on

the regulatory issues that have been shown to limit the wider use of pharmacogenetics (Chapter 2). An expanded version of a published paper analysing the evidence cited by existing RCTs to justify inclusion of genetic biomarkers is then included (Chapter 3). There is then an exploration of the use of simulation in the R coding language to produce evidence for genetic testing in a very rare outcome (Chapter 4).

The final chapters examine the use of DCEs and present my own experiment conducted in 2021 with 2000 members of the general public. I first undertook a systematic review of existing DCEs in pharmacogenetics, focussing on those that specifically examined ADRs (Chapter 5). I applied these ideas to my own DCE design, incorporating extensive qualitative work into the design (Chapter 6). The final chapter presents the results of the DCE and how these fit into the wider literature on pharmacogenetics (Chapter 7).

1.3 References

1. Lam YW. Scientific challenges and implementation barriers to translation of pharmacogenomics in clinical practice. ISRN Pharmacol. 2013;2013:641089.
2. Aymanns C, Keller F, Maus S, Hartmann B, Czock D. Review on pharmacokinetics and pharmacodynamics and the aging kidney. Clin J Am Soc Nephrol. 2010;5(2):314-27.
3. Delco F, Tchambaz L, Schlienger R, Drewe J, Krahenbuhl S. Dose adjustment in patients with liver disease. Drug Saf. 2005;28(6):529-45.
4. Pirmohamed M, Hughes DA. Pharmacogenetic tests: the need for a level playing field. Nature reviews Drug discovery. 2013;12(1):3-4.
5. Collins DC, Sundar R, Lim JSJ, Yap TA. Towards Precision Medicine in the Clinic: From Biomarker Discovery to Novel Therapeutics. Trends Pharmacol Sci. 2017;38(1):25-40.
6. Allison M. Is personalized medicine finally arriving? Nat Biotechnol. 2008;26(5):509-17.
7. Mossialos E, Mrazek M, Walley T. EBOOK: Regulating Pharmaceuticals in Europe: Striving for Efficiency, Equity and Quality. Berkshire, UNITED KINGDOM: McGraw-Hill Education; 2004.
8. National Research Council. Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. 2011.
9. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med. 2009;11(1):3-14.
10. Pirmohamed M. Acceptance of biomarker-based tests for application in clinical practice: criteria and obstacles. Clinical pharmacology and therapeutics. 2010;88(6):862-6.
11. Pirmohamed M, Park BK. Genetic susceptibility to adverse drug reactions. Trends Pharmacol Sci. 2001;22(6):298-305.

12. Trusheim MR, Berndt ER, Douglas FL. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. Nature reviews Drug discovery. 2007;6(4):287-93.
13. Barbarino JM, Whirl-Carrillo M, Altman RB, Klein TE. PharmGKB: A worldwide resource for pharmacogenomic information. Wiley Interdiscip Rev Syst Biol Med. 2018;10(4):e1417.
14. Faulkner E, Holtorf AP, Walton S, Liu CY, Lin H, Biltaj E, et al. Being Precise About Precision Medicine: What Should Value Frameworks Incorporate to Address Precision Medicine? A Report of the Personalized Precision Medicine Special Interest Group. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2020;23(5):529-39.
15. Buchanan A, Califano A, Kahn J, McPherson E, Robertson J, Brody B. Pharmacogenetics: ethical issues and policy options. Kennedy Institute of Ethics journal. 2002;12(1):1-15.
16. Abrahams E, Ginsburg GS, Silver M. The Personalized Medicine Coalition: goals and strategies. Am J Pharmacogenomics. 2005;5(6):345-55.
17. Need AC, Motulsky AG, Goldstein DB. Priorities and standards in pharmacogenetic research. Nat Genet. 2005;37(7):671-81.
18. Goldstein DB, Tate SK, Sisodiya SM. Pharmacogenetics goes genomic. Nature reviews Genetics. 2003;4(12):937-47.
19. Moldrup C. Ethical, social and legal implications of pharmacogenomics: a critical review. Community Genet. 2001;4(4):204-14.
20. Williams-Jones B, Corrigan OP. Rhetoric and hype: where's the 'ethics' in pharmacogenomics? Am J Pharmacogenomics. 2003;3(6):375-83.
21. Food and Drug Administration (US). Guidance for industry: E15 definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories 2008 [Available from: https://www.fda.gov/media/71389/download.
22. Zhang S, Davidson DD, Zhang DY, Parks JA, Cheng L. Principles of clinical molecular biology. Molecular Genetic Pathology: Springer; 2008. p. 3-32.
23. Macfarlane WM. Demystified.... Transcription. Mol Pathol. 2000;53(1):1-7.
24. European Medicines Agency. Terminology in Pharmacogenetics 2002 [Available from: https://www.ema.europa.eu/en/terminology-pharmacogenetics.
25. British National Formulary. ABACAVIR [Available from: https://bnf.nice.org.uk/drug/abacavir.html#preTreatmentScreeningInformation.
26. Martin MA, Hoffman JM, Freimuth RR, Klein TE, Dong BJ, Pirmohamed M, et al. Clinical Pharmacogenetics Implementation Consortium Guidelines for HLA-B Genotype and Abacavir Dosing: 2014 update. Clinical pharmacology and therapeutics. 2014;95(5):499-500.
27. Phillips EJ, Mallal SA. Pharmacogenetics and the potential for the individualization of antiretroviral therapy. Current opinion in infectious diseases. 2008;21(1):16-24.
28. Mounzer K, Hsu R, Fusco JS, Brunet L, Henegar CE, Vannappagari V, et al. HLA-B*57:01 screening and hypersensitivity reaction to abacavir between 1999 and 2016 in the OPERA((R)) observational database: a cohort study. AIDS Res Ther. 2019;16(1):1.
29. Hewitt RG. Abacavir hypersensitivity reaction. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2002;34(8):1137-42.
30. Jorgensen AL, FitzGerald RJ, Oyee J, Pirmohamed M, Williamson PR. Influence of CYP2C9 and VKORC1 on patient response to warfarin: a systematic review and meta-analysis. PLoS One. 2012;7(8):e44064.
31. van Schie RM, Wadelius MI, Kamali F, Daly AK, Manolopoulos VG, de Boer A, et al. Genotype-guided dosing of coumarin derivatives: the European

pharmacogenetics of anticoagulant therapy (EU-PACT) trial design. Pharmacogenomics. 2009;10(10):1687-95.
32. Pirmohamed M, Burnside G, Eriksson N, Jorgensen AL, Toh CH, Nicholson T, et al. A randomized trial of genotype-guided dosing of warfarin. The New England journal of medicine. 2013;369(24):2294-303.
33. Plumpton CO, Pirmohamed M, Hughes DA. Cost-Effectiveness of Panel Tests for Multiple Pharmacogenes Associated With Adverse Drug Reactions: An Evaluation Framework. Clinical pharmacology and therapeutics. 2019;105(6):1429-38.
34. Plumpton CO, Roberts D, Pirmohamed M, Hughes DA. A Systematic Review of Economic Evaluations of Pharmacogenetic Testing for Prevention of Adverse Drug Reactions. Pharmacoeconomics. 2016;34(8):771-93.
35. Conti R, Veenstra DL, Armstrong K, Lesko LJ, Grosse SD. Personalized medicine and genomics: challenges and opportunities in assessing effectiveness, cost-effectiveness, and future research priorities. Medical decision making : an international journal of the Society for Medical Decision Making. 2010;30(3):328-40.
36. Phillips KA, Ann Sakowski J, Trosman J, Douglas MP, Liang SY, Neumann P. The economic value of personalized medicine tests: what we know and what we need to know. Genet Med. 2014;16(3):251-7.
37. Attwood MM, Rask-Andersen M, Schiöth HB. Orphan Drugs and Their Impact on Pharmaceutical Development. Trends in Pharmacological Sciences. 2018;39(6):525-35.
38. Khatry DB. Demonstrating efficacy and effectiveness in clinical studies with recurrent event as primary end point: a chronic obstructive pulmonary disease example. J Comp Eff Res. 2018;7(10):935-45.
39. Salonia A, Rigatti P, Montorsi F. Sildenafil in erectile dysfunction: a critical review. Current Medical Research and Opinion. 2003;19(4):241-62.
40. Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. The New England journal of medicine. 2015;372(21):2018-28.
41. PharmGKB. About Us 2020 [Available from: https://www.pharmgkb.org/about.
42. PharmGKB. Drug Label Annotations 2020 [Available from: https://www.pharmgkb.org/labelAnnotations.
43. PharmGKB IAWM. Drug Labels 2016 [Available from: https://web.archive.org/web/20160622075834/https://www.pharmgkb.org/view/drug-labels.do.
44. Taylor CR. Introduction to Predictive Biomarkers: Definitions and Characteristics. In: Badve S, Kumar GL, editors. Predictive Biomarkers in Oncology. Cham: Springer International Publishing; 2019. p. 3-18.
45. Jorgensen JT. Site-agnostic biomarker-guided oncology drug development. Expert Rev Mol Diagn. 2020;20(6):583-92.
46. Kreutzfeldt J, Rozeboom B, Dey N, De P. The trastuzumab era: current and upcoming targeted HER2+ breast cancer therapies. Am J Cancer Res. 2020;10(4):1045-67.
47. Boekhout AH, Beijnen JH, Schellens JH. Trastuzumab. The oncologist. 2011;16(6):800-10.
48. Owens MA, Horten BC, Da Silva MM. HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. Clin Breast Cancer. 2004;5(1):63-9.
49. Clinical Pharmacogenetics Implementation Consortium. Guidelines 2019 [Available from: https://cpicpgx.org/guidelines/.
50. Kimpton JE, Carey IM, Threapleton CJD, Robinson A, Harris T, Cook DG, et al. Longitudinal exposure of English primary care patients to pharmacogenomic

drugs: An analysis to inform design of pre-emptive pharmacogenomic testing. British journal of clinical pharmacology. 2019;85(12):2734-46.
51. Youssef E, Kirkdale CL, Wright DJ, Guchelaar HJ, Thornley T. Estimating the potential impact of implementing pre-emptive pharmacogenetic testing in primary care across the UK. British journal of clinical pharmacology. 2021;87(7):2907-25.
52. Lunenburg CATC, Ishtiak-Ahmed K, Werge T, Gasse C. Life-time Actionable Pharmacogenetic Drug Use: A Population-based Cohort Study in 86 040 Young People With and Without Mental Disorders in Denmark. Pharmacopsychiatry. 2021;55(02):95-107.
53. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clinical pharmacology and therapeutics. 2001;69(3):89-95.
54. Bousman CA, Dunlop BW. Genotype, phenotype, and medication recommendation agreement among commercial pharmacogenetic-based decision support tools. Pharmacogenomics J. 2018;18(5):613-22.
55. Pirmohamed M. Personalized pharmacogenomics: predicting efficacy and adverse drug reactions. Annu Rev Genomics Hum Genet. 2014;15(1):349-70.
56. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clinical pharmacology and therapeutics. 2012;92(4):414-7.
57. Maranville JC, Di Rienzo A. Combining genetic and nongenetic biomarkers to realize the promise of pharmacogenomics for inflammatory diseases. Pharmacogenomics. 2014;15(15):1931-40.
58. Griffiths AJ, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. Somatic versus germinal mutation. An introduction to genetic analysis. 2000.
59. Food and Drug Administration (US). FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource 2016 [updated 2017 Nov 14. Available from: https://www.ncbi.nlm.nih.gov/books/NBK338448/
60. University of Maryland's Center of Excellence in Regulatory Science and Innovation, Food and Drug Administration (US), Critical Path Institute (C-Path). Evidentiary Considerations for Integration of Biomarkers in Drug Development 2015 [Available from: https://c-path.org/wp-content/uploads/2015/09/Proceedings-EvidentiaryConsiderationsSymposium-FINAL-20160104.pdf.
61. Chau CH, Rixe O, McLeod H, Figg WD. Validation of analytic methods for biomarkers used in drug development. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008;14(19):5967-76.
62. Yu C, Bashaw ED. Regulatory perspective of biomarker bioanalysis during drug development. Bioanalysis. 2019;11(7):607-10.
63. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Statistics in medicine. 1989;8(4):431-40.
64. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. Statistics in medicine. 2012;31(25):2973-84.
65. Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Stat Methods Med Res. 2008;17(3):303-40.
66. World Health Organization. Safety of Medicines - A Guide to Detecting and Reporting Adverse Drug Reactions - Why Health Professionals Need to Take Action 2002 [Available from: http://apps.who.int/medicinedocs/en/d/Jh2992e/2.html.
67. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ. 2004;329(7456):15-9.

68. Turner RM, Newman WG, Bramon E, McNamee CJ, Wong WL, Misbah S, et al. Pharmacogenomics in the UK National Health Service: opportunities and challenges. Pharmacogenomics. 2020;21(17):1237-46.
69. Davies EC, Green CF, Taylor S, Williamson PR, Mottram DR, Pirmohamed M. Adverse drug reactions in hospital in-patients: a prospective analysis of 3695 patient-episodes. PLoS One. 2009;4(2):e4439.
70. Insani WN, Whittlesea C, Alwafi H, Man KKC, Chapman S, Wei L. Prevalence of adverse drug reactions in the primary care setting: A systematic review and meta-analysis. PLoS One. 2021;16(5):e0252161.
71. Thompson AJ, Newman WG, Elliott RA, Roberts SA, Tricker K, Payne K. The cost-effectiveness of a pharmacogenetic test: a trial-based evaluation of TPMT genotyping for azathioprine. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2014;17(1):22-33.
72. Newman WG, Payne K, Tricker K, Roberts SA, Fargher E, Pushpakom S, et al. A pragmatic randomized controlled trial of thiopurine methyltransferase genotyping prior to azathioprine treatment: the TARGET study. Pharmacogenomics. 2011;12(6):815-26.
73. Coenen MJ, de Jong DJ, van Marrewijk CJ, Derijks LJ, Vermeulen SH, Wong DR, et al. Identification of Patients With Variants in TPMT and Dose Reduction Reduces Hematologic Events During Thiopurine Treatment of Inflammatory Bowel Disease. Gastroenterology. 2015;149(4):907-17 e7.
74. Azuma J, Ohno M, Kubota R, Yokota S, Nagai T, Tsuyuguchi K, et al. NAT2 genotype guided regimen reduces isoniazid-induced liver injury and early treatment failure in the 6-month four-drug standard treatment of tuberculosis: a randomized controlled trial for pharmacogenetics-based therapy. Eur J Clin Pharmacol. 2013;69(5):1091-101.
75. Kim KP, Kim HS, Sym SJ, Bae KS, Hong YS, Chang HM, et al. A UGT1A1*28 and *6 genotype-directed phase I dose-escalation trial of irinotecan with fixed-dose capecitabine in Korean patients with metastatic colorectal cancer. Cancer Chemother Pharmacol. 2013;71(6):1609-17.
76. Saag M, Balu R, Phillips E, Brachman P, Martorell C, Burman W, et al. High sensitivity of human leukocyte antigen-b*5701 as a marker for immunologically confirmed abacavir hypersensitivity in white and black patients. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2008;46(7):1111-8.
77. Phillips E, Mallal S. Successful translation of pharmacogenetics into the clinic: the abacavir example. Mol Diagn Ther. 2009;13(1):1-9.
78. Illing PT, Purcell AW, McCluskey J. The role of HLA genes in pharmacogenomics: unravelling HLA associated adverse drug reactions. Immunogenetics. 2017;69(8-9):617-30.
79. Daluge SM, Good SS, Faletto MB, Miller WH, St Clair MH, Boone LR, et al. 1592U89, a novel carbocyclic nucleoside analog with potent, selective anti-human immunodeficiency virus activity. Antimicrob Agents Chemother. 1997;41(5):1082-93.
80. Food and Drug Administration (US). Full Prescribing Information: Ziagen 2008 [Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2008/020977s017,020978s020lbl.pdf
81. Food and Drug Administration (US). Full Prescribing Information: Ziagen 2018 [Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/020977s033s034,020978s036s037lbl.pdf.
82. Hetherington S, McGuirk S, Powell G, Cutrell A, Naderer O, Spreen B, et al. Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. Clin Ther. 2001;23(10):1603-14.

83. Clay PG. The abacavir hypersensitivity reaction: a review. Clin Ther. 2002;24(10):1502-14.
84. Shapiro M, Ward KM, Stern JJ. A near-fatal hypersensitivity reaction to abacavir: case report and literature review. The AIDS reader. 2001;11(4):222-6.
85. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, et al. HLA-B*5701 screening for hypersensitivity to abacavir. The New England journal of medicine. 2008;358(6):568-79.
86. Mallal S, Nolan D, Witt C, Masel G, Martin AM, Moore C, et al. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. Lancet. 2002;359(9308):727-32.
87. Hetherington S, Hughes AR, Mosteller M, Shortino D, Baker KL, Spreen W, et al. Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. Lancet. 2002;359(9312):1121-2.
88. Sousa-Pinto B, Pinto-Ramos J, Correia C, Goncalves-Costa G, Gomes L, Gil-Mata S, et al. Pharmacogenetics of abacavir hypersensitivity: A systematic review and meta-analysis of the association with HLA-B*57:01. J Allergy Clin Immunol. 2015;136(4):1092-4 e3.
89. Phillips E, Mallal S. Drug hypersensitivity in HIV. Curr Opin Allergy Clin Immunol. 2007;7(4):324-30.
90. Orkin C, Wang J, Bergin C, Molina JM, Lazzarin A, Cavassini M, et al. An epidemiologic study to determine the prevalence of the HLA-B*5701 allele among HIV-positive patients in Europe. Pharmacogenetics and genomics. 2010;20(5):307-14.
91. Smith KY, Patel P, Fine D, Bellos N, Sloan L, Lackey P, et al. Randomized, double-blind, placebo-matched, multicenter trial of abacavir/lamivudine or tenofovir/emtricitabine with lopinavir/ritonavir for initial HIV treatment. AIDS (London, England). 2009;23(12):1547-56.
92. Young B, Squires K, Patel P, Dejesus E, Bellos N, Berger D, et al. First large, multicenter, open-label study utilizing HLA-B*5701 screening for abacavir hypersensitivity in North America. AIDS (London, England). 2008;22(13):1673-5.
93. Hughes DA, Vilar FJ, Ward CC, Alfirevic A, Park BK, Pirmohamed M. Cost-effectiveness analysis of HLA B*5701 genotyping in preventing abacavir hypersensitivity. Pharmacogenetics. 2004;14(6):335-42.
94. Ma JD, Lee KC, Kuo GM. HLA-B*5701 testing to predict abacavir hypersensitivity. PLoS Curr. 2010;2:RRN1203.
95. Kauf TL, Farkouh RA, Earnshaw SR, Watson ME, Maroudas P, Chambers MG. Economic efficiency of genetic screening to inform the use of abacavir sulfate in the treatment of HIV. Pharmacoeconomics. 2010;28(11):1025-39.
96. Medicines and Healthcare products Regulatory Agency. Summary of Product Characteristics: Abacavir 2019 [Available from: https://mhraproducts4853.blob.core.windows.net/docs/449e73eeb73d9e299df5ecb747d94ed3bd702036.
97. Waldman SA. Does potency predict clinical efficacy? Illustration through an antihistamine model. Annals of Allergy, Asthma & Immunology. 2002;89(1):7-12.
98. Schwaederle M, Zhao M, Lee JJ, Eggermont AM, Schilsky RL, Mendelsohn J, et al. Impact of Precision Medicine in Diverse Cancers: A Meta-Analysis of Phase II Clinical Trials. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2015;33(32):3817-25.
99. Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K, et al. Vemurafenib: the first drug approved for BRAF-mutant cancer. Nature reviews Drug discovery. 2012;11(11):873-86.
100. Dean L. Vemurafenib Therapy and BRAF and NRAS Genotype. In: Pratt VM, Scott SA, Pirmohamed M, Esquivel B, Kane MS, Kattman BL, et al., editors. Medical Genetics Summaries. Bethesda (MD): National Center for Biotechnology Information (US); 2012.

101. Sosman JA, Kim KB, Schuchter L, Gonzalez R, Pavlick AC, Weber JS, et al. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. The New England journal of medicine. 2012;366(8):707-14.
102. Food and Drug Administration (US). Full Prescribing Information: Zelboraf 2017 [Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/202429Orig1s016LbL.pdf.
103. Yang H, Higgins B, Kolinsky K, Packman K, Go Z, Iyer R, et al. RG7204 (PLX4032), a selective BRAFV600E inhibitor, displays potent antitumor activity in preclinical melanoma models. Cancer Res. 2010;70(13):5518-27.
104. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. Nature. 2002;417(6892):949-54.
105. Long GV, Menzies AM, Nagrial AM, Haydu LE, Hamilton AL, Mann GJ, et al. Prognostic and clinicopathologic associations of oncogenic BRAF in metastatic melanoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(10):1239-46.
106. Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay JY, et al. Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. The New England journal of medicine. 2015;373(8):726-36.
107. Kopetz S, McDonough SL, Morris VK, Lenz H-J, Magliocco AM, Atreya CE, et al. Randomized trial of irinotecan and cetuximab with or without vemurafenib in BRAF-mutant metastatic colorectal cancer (SWOG 1406). Journal of Clinical Oncology. 2017;35(4_suppl):520-.
108. Dunn LA, Sherman EJ, Baxi SS, Tchekmedyian V, Grewal RK, Larson SM, et al. Vemurafenib Redifferentiation of BRAF Mutant, RAI-Refractory Thyroid Cancers. J Clin Endocrinol Metab. 2019;104(5):1417-28.
109. Dietrich S, Glimm H, Andrulis M, von Kalle C, Ho AD, Zenz T. BRAF inhibition in refractory hairy-cell leukemia. The New England journal of medicine. 2012;366(21):2038-40.
110. Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. The New England journal of medicine. 2010;363(9):809-19.
111. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. The New England journal of medicine. 2011;364(26):2507-16.
112. Mazieres J, Cropet C, Montane L, Barlesi F, Souquet PJ, Quantin X, et al. Vemurafenib in non-small-cell lung cancer patients with BRAF(V600) and BRAF(nonV600) mutations. Annals of oncology : official journal of the European Society for Medical Oncology. 2020;31(2):289-94.
113. Teuma C, Perier-Muzet M, Pelletier S, Nouvier M, Amini-Adl M, Dijoud F, et al. New insights into renal toxicity of the B-RAF inhibitor, vemurafenib, in patients with metastatic melanoma. Cancer Chemother Pharmacol. 2016;78(2):419-26.
114. Jhaveri KD, Sakhiya V, Fishbane S. Nephrotoxicity of the BRAF Inhibitors Vemurafenib and Dabrafenib. JAMA Oncol. 2015;1(8):1133-4.
115. Halait H, Demartin K, Shah S, Soviero S, Langland R, Cheng S, et al. Analytical performance of a real-time PCR-based assay for V600 mutations in the BRAF gene, used as the companion diagnostic test for the novel BRAF inhibitor vemurafenib in metastatic melanoma. Diagnostic molecular pathology : the American journal of surgical pathology, part B. 2012;21(1):1-8.
116. Eckhart L, Bach J, Ban J, Tschachler E. Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity. Biochemical and biophysical research communications. 2000;271(3):726-30.
117. Le Tourneau C, Delord JP, Goncalves A, Gavoille C, Dubot C, Isambert N, et al. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-

of-concept, randomised, controlled phase 2 trial. Lancet Oncol. 2015;16(13):1324-34.
118. Larkin J, Ascierto PA, Dreno B, Atkinson V, Liszkay G, Maio M, et al. Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. The New England journal of medicine. 2014;371(20):1867-76.
119. Ascierto PA, McArthur GA, Dréno B, Atkinson V, Liszkay G, Di Giacomo AM, et al. Cobimetinib combined with vemurafenib in advanced BRAFV600-mutant melanoma (coBRIM): updated efficacy results from a randomised, double-blind, phase 3 trial. The Lancet Oncology. 2016;17(9):1248-60.
120. Amin A, Lawson DH, Salama AK, Koon HB, Guthrie T, Jr., Thomas SS, et al. Phase II study of vemurafenib followed by ipilimumab in patients with previously untreated BRAF-mutated metastatic melanoma. J Immunother Cancer. 2016;4(1):44.
121. Food and Drug Administration (US). List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools) 2020 [Available from: https://www.fda.gov/medical-devices/vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-vitro-and-imaging-tools.
122. National Institute for Health and Care Excellence (NICE) U. Vemurafenib for treating locally advanced or metastatic BRAF V600 mutation-positive malignant melanoma 2015 [Available from: https://www.nice.org.uk/guidance/ta269/chapter/1-Guidance.
123. Adams CP, Brantner VV. Spending on new drug development1. Health Econ. 2010;19(2):130-41.
124. Amur S, LaVange L, Zineh I, Buckman-Garner S, Woodcock J. Biomarker Qualification: Toward a Multiple Stakeholder Framework for Biomarker Development, Regulatory Acceptance, and Utilization. Clinical pharmacology and therapeutics. 2015;98(1):34-46.
125. Frank R, Hargreaves R. Clinical biomarkers in drug discovery and development. Nature reviews Drug discovery. 2003;2(7):566-80.
126. Roses AD. Pharmacogenetics in drug discovery and development: a translational perspective. Nature reviews Drug discovery. 2008;7(10):807-17.
127. Zeggini E, Gloyn AL, Barton AC, Wain LV. Translational genomics and precision medicine: Moving from the lab to the clinic. Science (New York, NY). 2019;365(6460):1409-13.
128. Wheeler E, Leong A, Liu CT, Hivert MF, Strawbridge RJ, Podmore C, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med. 2017;14(9):e1002383.
129. Deverka PA, Vernon J, McLeod HL. Economic opportunities and challenges for pharmacogenomics. Annu Rev Pharmacol Toxicol. 2010;50(1):423-37.
130. Burt T, Dhillon S. Pharmacogenomics in early-phase clinical development. Pharmacogenomics. 2013;14(9):1085-97.
131. Zhang W, Roederer MW, Chen WQ, Fan L, Zhou HH. Pharmacogenetics of drugs withdrawn from the market. Pharmacogenomics. 2012;13(2):223-31.
132. Lavezzari G, Womack AW. Industry perspectives on biomarker qualification. Clinical pharmacology and therapeutics. 2016;99(2):208-13.
133. (US) FaDA. Personalized medicine: A biological approach to patient treatment, A conversation with Issam Zineh, Director of CDER's Office of Clinical Pharmacology 2016 [Available from: https://www.fda.gov/Drugs/NewsEvents/ucm325475.htm.
134. Schneider JE, Sidhu MK, Doucet C, Kiss N, Ohsfeldt RL, Chalfin D. Economics of cancer biomarkers. Per Med. 2012;9(8):829-37.
135. Najafzadeh M, Johnston KM, Peacock SJ, Connors JM, Marra MA, Lynd LD, et al. Genomic testing to determine drug response: measuring preferences of the

public and patients using Discrete Choice Experiment (DCE). BMC health services research. 2013;13(1):454.
136. Walley T, Haycox A, Boland A. Pharmacoeconomics: Elsevier Health Sciences; 2004.
137. Brouwer W, van Baal P, van Exel J, Versteegh M. When is it too expensive? Cost-effectiveness thresholds and health care decision-making. Eur J Health Econ. 2019;20(2):175-80.
138. Donaldson C, Currie G, Mitton C. Cost effectiveness analysis in health care: contraindications. BMJ. 2002;325(7369):891-4.
139. National Institute for Health and Care Excellence (NICE) U. NICE Charter 2020 [Available from: https://www.nice.org.uk/about/who-we-are/our-charter.
140. Marseille E, Larson B, Kazi DS, Kahn JG, Rosen S. Thresholds for the cost-effectiveness of interventions: alternative approaches. Bull World Health Organ. 2015;93(2):118-24.
141. Stern AD, Alexander BM, Chandra A. How economics can shape precision medicines. Science (New York, NY). 2017;355(6330):1131-3.
142. Jorgensen JT. From blockbuster medicine to personalized medicine. Per Med. 2008;5(1):55-63.
143. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol. 2001;2(2):127-37.
144. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. The New England journal of medicine. 2005;353(16):1659-72.
145. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science (New York, NY). 1989;244(4905):707-12.
146. Sarup JC, Johnson RM, King KL, Fendly BM, Lipari MT, Napier MA, et al. Characterization of an anti-p185HER2 monoclonal antibody that stimulates receptor function and inhibits tumor cell growth. Growth Regul. 1991;1(2):72-82.
147. Lindsley CW. Predictions and Statistics for the Best-Selling Drugs Globally and in the United States in 2018 and a Look Forward to 2024 Projections. ACS Publications; 2019.
148. Yin W, Jiang Y, Shen Z, Shao Z, Lu J. Trastuzumab in the adjuvant treatment of HER2-positive early breast cancer patients: a meta-analysis of published randomized controlled trials. PLoS One. 2011;6(6):e21030.
149. National Institute for Health and Care Excellence (NICE) U. Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer. NICE London; 2006.
150. Chen Z, Liew D, Kwan P. Real-world cost-effectiveness of pharmacogenetic screening for epilepsy treatment. Neurology. 2016;86(12):1086-94.
151. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics. 2008;26(8):661-77.
152. Louviere JJ, Flynn TN, Carson RT. Discrete Choice Experiments Are Not Conjoint Analysis. Journal of Choice Modelling. 2010;3(3):57-72.
153. Chen Z, Liew D, Kwan P. Effects of a HLA-B*15:02 screening policy on antiepileptic drug use and severe skin reactions. Neurology. 2014;83(22):2077-84.
154. Lin CW, Huang WI, Chao PH, Chen WW, Hsiao FY. Temporal trends and patterns in carbamazepine use, related severe cutaneous adverse reactions, and HLA-B*15:02 screening: A nationwide study. Epilepsia. 2018;59(12):2325-39.
155. Wang B, Canestaro WJ, Choudhry NK. Clinical evidence supporting pharmacogenomic biomarker testing provided in US Food and Drug Administration drug labels. JAMA internal medicine. 2014;174(12):1938-44.

156. Klein ME, Parvez MM, Shin JG. Clinical Implementation of Pharmacogenomics for Personalized Precision Medicine: Barriers and Solutions. J Pharm Sci. 2017;106(9):2368-79.
157. Arwood MJ, Chumnumwat S, Cavallari LH, Nutescu EA, Duarte JD. Implementing Pharmacogenomics at Your Institution: Establishment and Overcoming Implementation Challenges. Clin Transl Sci. 2016;9(5):233-45.
158. Limaye N. Pharmacogenomics, Theranostics and Personalized Medicine - the complexities of clinical trials: challenges in the developing world. Appl Transl Genom. 2013;2:17-21.
159. van der Wouden CH, Cambon-Thomsen A, Cecchin E, Cheung KC, Davila-Fajardo CL, Deneer VH, et al. Implementing Pharmacogenomics in Europe: Design and Implementation Strategy of the Ubiquitous Pharmacogenomics Consortium. Clinical pharmacology and therapeutics. 2017;101(3):341-58.
160. McKinnon RA, Ward MB, Sorich MJ. A critical analysis of barriers to the clinical implementation of pharmacogenomics. Ther Clin Risk Manag. 2007;3(5):751-9.
161. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999-2014. BMJ open. 2016;6(6):e011666.
162. Schaper M, Schicktanz S. Medicine, market and communication: ethical considerations in regard to persuasive communication in direct-to-consumer genetic testing services. BMC Med Ethics. 2018;19(1):56.
163. Rogausch A, Prause D, Schallenberg A, Brockmoller J, Himmel W. Patients' and physicians' perspectives on pharmacogenetic testing. Pharmacogenomics. 2006;7(1):49-59.
164. Griffith GL, Edwards RT, Williams JM, Gray J, Morrison V, Wilkinson C, et al. Patient preferences and National Health Service costs: a cost-consequences analysis of cancer genetic services. Fam Cancer. 2009;8(4):265-75.
165. Zoltick ES, Linderman MD, McGinniss MA, Ramos E, Ball MP, Church GM, et al. Predispositional genome sequencing in healthy adults: design, participant characteristics, and early outcomes of the PeopleSeq Consortium. Genome medicine. 2019;11(1):10.
166. Joly Y, Dupras C, Pinkesz M, Tovino SA, Rothstein MA. Looking Beyond GINA: Policy Approaches to Address Genetic Discrimination. Annu Rev Genomics Hum Genet. 2020;21:491-507.
167. Rothstein MA, Anderlik MR. What is genetic discrimination, and when and how can it be prevented? Genet Med. 2001;3(5):354-8.
168. Brannan C, Foulkes AL, Lazaro-Munoz G. Preventing discrimination based on psychiatric risk biomarkers. Am J Med Genet B Neuropsychiatr Genet. 2019;180(2):159-71.
169. Neumann PJ, Cohen JT, Hammitt JK, Concannon TW, Auerbach HR, Fang C, et al. Willingness-to-pay for predictive tests with no immediate treatment implications: a survey of US residents. Health Econ. 2012;21(3):238-51.
170. Council of Europe. Recommendation CM/Rec(2016)8 of the Committee of Ministers to the member States on the processing of personal health-related data for insurance purposes, including data resulting from genetic tests 2016 [Available from: https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016806b2c5f.
171. Association of British Insurers. Code on Genetic Testing and Insurance 2021 [Available from: https://www.abi.org.uk/data-and-resources/tools-and-resources/genetics/code-on-genetic-testing-and-insurance/.
172. Hudson KL, Holohan MK, Collins FS. Keeping pace with the times--the Genetic Information Nondiscrimination Act of 2008. The New England journal of medicine. 2008;358(25):2661-3.

173. Baruch S, Hudson K. Civilian and military genetics: nondiscrimination policy in a post-GINA world. American journal of human genetics. 2008;83(4):435-44.
174. Novak MJ, Tabrizi SJ. Huntington's disease. BMJ. 2010;340:c3109.
175. Bombard Y, Rozmovits L, Trudeau M, Leighl NB, Deal K, Marshall DA. Access to personalized medicine: factors influencing the use and value of gene expression profiling in breast cancer treatment. Current oncology (Toronto, Ont). 2014;21(3):e426-33.
176. Huntington's Disease Association. Predictive Testing for Huntington's Disease. In: Association HsD, editor. Liverpool Science Park, 131 Mount Pleasant, Liverpool UK: Huntington's Disease Association; 2018.
177. National Lung Cancer Audit. Spotlight report on molecular testing in advanced lung cancer. Royal College of Physicians, Healthcare Quality Improvment Partnership (HQIP),; 2020.
178. Horton R, Crawford G, Freeman L, Fenwick A, Wright CF, Lucassen A. Direct-to-consumer genetic testing. BMJ. 2019;367:l5688.
179. Wynn J, Chung WK. 23andMe Paves the Way for Direct-to-Consumer Genetic Health Risk Tests of Limited Clinical Utility. Annals of internal medicine. 2017;167(2):125-6.
180. Roberts JS, Gornick MC, Carere DA, Uhlmann WR, Ruffin MT, Green RC. Direct-to-Consumer Genetic Testing: User Motivations, Decision Making, and Perceived Utility of Results. Public Health Genomics. 2017;20(1):36-45.
181. Tandy-Connor S, Guiltinan J, Krempely K, LaDuca H, Reineke P, Gutierrez S, et al. False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care. Genet Med. 2018;20(12):1515-21.
182. Regalado A. More than 26 million people have taken an at-home ancestry test 2019 [Available from: https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/.
183. Green MJ, Botkin JR. "Genetic exceptionalism" in medicine: clarifying the differences between genetic and nongenetic tests. Annals of internal medicine. 2003;138(7):571-5.
184. Rahawi S, Naik H, Blake KV, Owusu Obeng A, Wasserman RM, Seki Y, et al. Knowledge and attitudes on pharmacogenetics among pediatricians. J Hum Genet. 2020;65(5):437-44.
185. Food and Drug Administration (US). Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling: Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders 2016 [Available from: https://www.fda.gov/media/92593/download.
186. Dong D, Ozdemir S, Mong Bee Y, Toh SA, Bilger M, Finkelstein E. Measuring High-Risk Patients' Preferences for Pharmacogenetic Testing to Reduce Severe Adverse Drug Reaction: A Discrete Choice Experiment. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2016;19(6):767-75.
187. Ryan M. Using Consumer Preferences in Health Care Decision Making: The Application of Conjoint Analysis1996.
188. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2011;14(4):403-13.

Chapter 2: Biomarker assessment and regulation

"A bad tumor biomarker test is as bad as a bad drug" (1)

2.1 Why do we need biomarker assessment and regulation?

Any medicine that is going to be used in humans needs to go through a rigorous assessment and regulatory process to ensure it is safe and effective (2). The same is true of biomarkers. While biomarkers do not directly endanger life or health via toxicity or adverse drug reactions (ADRs), there are still dangers associated with their use (3). Misdiagnosis, inaccurate risk estimates, and incorrect drug choices are possible results of a poorly regulated and/or validated biomarker process (4-6). A poor choice of biomarker to select patients for a trial risks failing to prove the efficacy of a useful drug (7). Clinically, a poor biomarker may lead to incorrect decision making and subsequent patient harm (1). Regulators generally require clear evidence of patient benefit to approve an intervention (8). This evidentiary standard is ill-defined. This has led to approvals of some biomarkers with very different levels of evidence behind them (9). Some groups (in industry, within regulators, and academic) have attempted to produce frameworks for the evaluation of biomarkers, that identify and assess the evidence for their use. However, this has not led to a unified solution for clinicians, researchers, and drug developers to use when investigating genetic biomarkers and there is limited specific guidance on the qualification and use of biomarkers. For the avoidance of doubt, when referring to 'biomarkers' in the remainder of this chapter it can be assumed that the discussion relates equally to pharmacogenetic markers as it does to other types of biomarkers.

The aim of this chapter is to detail how genetic biomarkers may be assessed, including the use of several different formal frameworks. The level, type, and quality of the evidence required under each assessment is detailed. The guidelines issued by regulatory authorities on what evidence is acceptable for the use of a biomarker clinically and/or in a trial setting are discussed. Finally, the challenges associated with integrating genetic biomarkers into trials are discussed.

In general, the more risk associated with a biomarker, the more regulatory scrutiny it undergoes (Figure 2.1). 'Risk' here refers to the potential impact of the decision made by using the biomarker (10). A lower risk biomarker might be one used to identify those at risk of developing a mild, but inconvenient ADR, in combination

with clinical factors. A higher risk biomarker may be used to define eligibility for life-changing treatment options.

In a trial, biomarkers integral to that trial's conduct also require more evidence than biomarkers used on an exploratory basis (6, 11, 12). The level of scrutiny and evidence required are also dependent on the intended use of the biomarker (10, 12-15). Biomarkers for exploratory usages will have less scrutiny than those integral for clinical decision making (16). For example, the SHIVA trial assigned cancer treatment to metastatic cancer patients based on their tumour biomarkers (17). The biomarkers used in this serious disease setting, potentially affecting patient survival, should undergo greater scrutiny and require more and higher quality evidence before use. In contrast, using *CYP2D6* genotyping to determine opioid dosing (18) does not directly impact on survival. This biomarker can therefore be used with less scrutiny and less evidence.

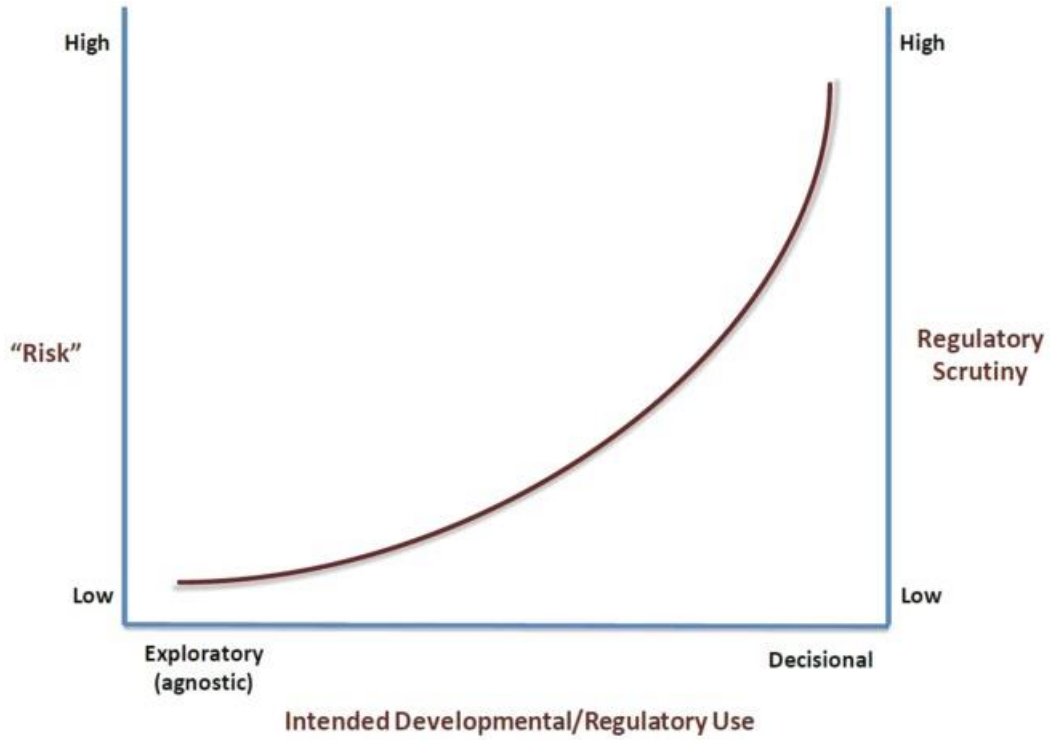


Figure 2.1 - As the risk associated with a biomarker increases, so does the level of regulatory scrutiny levied. This is also influenced by the intended use of the biomarker. "Risk" refers to the impact of the decision to be made based on the biomarker results. Exploratory, or agnostic biomarker use refers here to examining a biomarker in an isolated manner, separate from, e.g., the tumour site. From Amur, et al. *(2015) (10).*

However, the use of biomarkers in drug development or to guide treatment are still relatively new concepts compared to issues normally dealt with by pharmaceutical regulation. Many regulatory authorities are still trying to 'catch up' with the science,

and there is generally little consensus across jurisdictions. The US Food and Drug Administration (FDA) has admitted that next-generation sequencing technologies strain its current regulatory methods (19, 20) and its current approval process for biomarkers has been criticised as not fit-for-purpose (12).

The implementation of biomarkers and pharmacogenetics into clinical practice is limited by a lack of high quality randomised controlled trial (RCT) evidence (21, 22). This issue is also discussed further in Chapter 3. However, it is also difficult for triallists and funders to commence these RCTs without clear regulatory guidance.

The aim of this chapter is to provide an overview of the current frameworks available for the evaluation of biomarkers. Outside of national and international regulatory agencies, many frameworks have been published for the evaluation of biomarkers. Many of these are used by regulators for their own evaluations. I discuss these frameworks in this chapter, along with an overview of the views of UK and US regulators. Finally, I discuss the difficulties of regulating trials for genetic biomarkers and some innovative solutions to these problems.

2.2 How are biomarkers evaluated?

There are several terms relating to biomarkers and their regulation that are formally defined.

Regulation, or licensing, is the process where drugs or biomarkers are assessed by government agencies for safety, efficacy, and quality of production, before being allowed onto the market (23). Regulators require biomarkers to be qualified before they are used clinically. Qualification, in the context of a biomarker to be used for drug development, is “a conclusion, based on a formal regulatory process, that within the stated context of use, a medical product development tool can be relied upon to have a specific interpretation and application in medical product development and regulatory review” (16, 24-26). When considering a biomarker to be used to guide treatment within a clinical context, it is the process by which a biomarker is linked to a clinical outcome or phenotype of interest (16, 27). This qualification can be split into analytical and clinical domains (15, 28). Analytic concerns include sensitivity, specificity, and precision, while clinical concerns include validity, quality assurance, and education of providers (28). Clinical concerns also include the risks and benefits associated with testing (28), and the link between the biomarker and the outcome of interest (29). Prior to a biomarker’s approval by a regulatory authority, analytic validity, clinical validity, and clinical utility

need to be evaluated (Table 2.1). Analytical validity refers to the biomarkers assay's ability to accurately detect the biomarker of interest, clinical validity refers to how well a biomarker test result correlates with the outcome of interest e.g. an adverse drug reaction, whilst clinical utility refers to the benefits a person would derive from a biomarker test. Each is considered in more detail in Table 2.1 below.

Term	Definition	Example
Analytic validity	Ability of a biomarker test to accurately and reliably measure the genotype (or analyte) of interest. Includes analytic sensitivity, specificity, reliability, and robustness (1, 12, 13, 28, 30, 31) Ensures the biomarker performance is fit-for-purpose (29)	Within- and between-laboratory precision
Clinical utility	The risks and benefits associated with a test's introduction to practice – the health outcomes (positive and negative) from the testing (28) The likelihood the test will lead to improved outcome with a given intervention (27)	Choosing an effective treatment in personalised medicine
Clinical validity	Ability of a biomarker test to accurately and reliably predict the disorder or phenotype of interest (27). Includes clinical sensitivity and specificity, positive and negative predictive values, and genetic factors (penetrance, variable expressivity) (28, 32) Ensures the biomarker reflects the outcome of interest (29)	Validation of the test when predicting clinical outcome in all populations where it might be used

Table 2.1 - Analytic validity, clinical utility, and clinical validity definitions used in biomarker qualification. See also Gillis and Innocenti (2014) (33)

Analytic validity involves assessing how biomarker assays perform in laboratory settings (30). The choice of assay for biomarker detection is assessed for sensitivity and specificity, as well as its analytical limits (31). When testing defines the population for a treatment, it is vital that testing is accurate (34). Other important issues to consider at this stage are: sample integrity over storage; training of personnel; and methods for minimising variability at all stages of the assay (31). An example of the importance of testing analytic validity is the failure of a promised proteomics biomarker for the detection of ovarian cancer (35, 36). Initial encouraging results of 100% sensitivity and specificity rates were not reproduceable. This was found to be due to the original researchers not considering the limit of detection of lab equipment, producing artifacts in the dataset (37).

It has historically been difficult to locate evidence on the analytic validity of biomarker tests, due to these results being less likely to be published (38, 39) than other validity and utility measures. This is due to commercialisation and data being held by laboratories and their related proprietary interests (38). This has led to issues with several trials proceeding that were based on weak evidence for the biomarker's analytical validity and its subsequent clinical performance (39). One group of authors identified issues with analytic validity that led to oncology trials being terminated for patient safety, since the underlying biomarker studies were fundamentally flawed (40).

A biomarker has clinical utility if there is evidence of improved measurable clinical outcomes associated with its use (32). Clinical utility is also defined as the health outcomes (both positive and negative) associated with testing (28). For example, an evaluation of clinical utility for a test for a specific cancer genotype which would guide targeted treatment might include assessing the availability of effective targeted treatments for that cancer. If none are available, there may not be sufficient justification for performing the test – the test would have low clinical utility.

Clinical validity denotes a test's ability to detect or predict a clinical status (e.g. presence of disease, treatment outcome etc.). A test has clinical validity when there is a 'strong, well-validated association between having the variant and having a particular disease or predisposition' (19, 32). In the case of pharmacogenetics, this would mean the test's ability to predict a treatment response outcome (e.g. developing an ADR). It is also dependent on the frequency of the allele, and the frequency of the ADR. This is explored further in Chapter 4.

For example, the anti-depressant citalopram is associated with increased side-effects in patients who are *CYP2C19* poor metabolisers (41). The clinical utility of this association was shown by Mrazek, *et al.* (2011) who showed that certain alleles are associated with tolerance of citalopram (42). The clinical validity of testing was confirmed in a study of citalopram in healthy Swedish volunteers where was complete concordance between *CYP2C19* genotypes and metabolism phenotypes (43).

A further critical example would be the importance of validating the test in all relevant populations (such as ethnicities, genders, and age groups). For example, a genetic test that works well in Asian populations may not perform as well in Caucasian populations due to different proportions of allele frequencies (44) and differing patterns of linkage disequilibrium.

Before qualification or clinical use, pharmacogenetic tests need to provide evidence to show they meet minimum criteria in each of these domains. The standards required to prove these criteria are variable and ill-defined (6). To combat this, different groups have produced methods for evaluating biomarkers and the evidentiary standards they should meet. These frameworks are distinct from regulatory approval frameworks but may be used in conjunction with each other.

2.2.1 Existing frameworks for biomarker evaluation

There have been several frameworks produced for the evaluation of biomarkers (45). Below, I have profiled several. ACCE was the earliest and arguably most influential framework (45). From the perspective of industry, I have profiled a 2007 meeting of industry and regulatory figures into the evidence required for evaluation (13). Finally, an overview is presented of the Pharmacogenomics Knowledge Base (PharmGKB) and Clinical Pharmacogenetics Implementation Consortium (CPIC) systems of classification (46, 47).

A full literature review of frameworks for the evaluation of genetic tests was published in 2018 by Pitini, *et al (45).* I have located some more recent frameworks and placed these in context with the older ACCE framework.

2.2.1.1 ACCE

One of the earliest frameworks for biomarker evaluation was the ACCE framework, established from 2000 to 2004 (28, 48). ACCE is an acronym denoting four components – *A*nalytic validity, *C*linical validity, *C*linical utility, and *E*thical, legal and social implications (Figure 2.2). ACCE provides a framework to evaluate tests by using 44 specific questions nested within the four components (28, 49-51). When using the ACCE framework, evidence used to reach conclusions about each question should be specified, including an assessment of the quality of that evidence and possible bias (30).

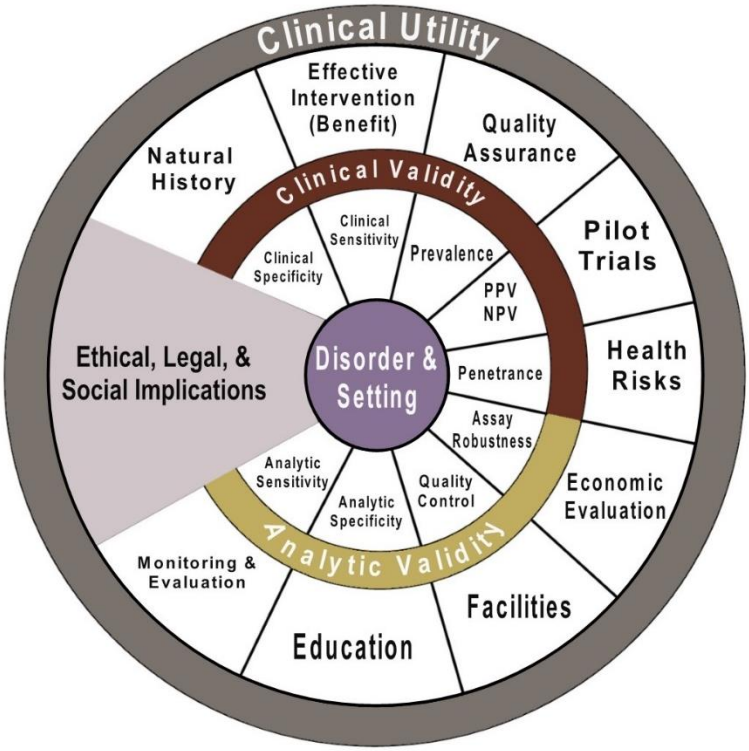


*Figure 2.2 - ACCE framework, used to evaluate tests through questions in each of these domains. ACCE stands for **A**nalytic validity, **C**linical validity, **C**linical utility, and **E**thical, Legal, and Social Implications. Adapted from Haddow and Palomaki (2004) (28).*

ACCE was developed from previous work by Wald and Cuckle (1989) with support from the US Centers for Disease Control and Prevention (CDC) (49, 52). Influential in the development of ACCE was a 2000 US committee formed to advise on the medical, scientific, ethical, legal, and social issues of genetic testing (although ACCE is not itself specific to genetic tests) (53). ACCE was later adapted (adding specific methodological guidance and details of family testing (30)) for use in genetic tests in the UK by the Genetic Testing Network Steering Group (30, 54). ACCE continues to be significant for regulation – a 2018 review found that 13 national evidence evaluation initiatives for genetic tests were based on ACCE (45).

ACCE later led to the 2005 formation of the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative (55). EGAPP was an independent body that issued reports and recommendations on the integration of genetic tests into practice for clinicians and other stakeholders (32). Their last recommendation, in 2015, was regarding the use of Oncotype DX tumour gene expression profiling in breast cancer (56).

While ACCE succeeded at the time, it became increasingly challenging to use as an evidence framework due to the expansion of biomarker complexity across disease areas, biomarker types, and contexts of use (57). The limited evidence available in some areas of genetic testing has also curtailed the use of ACCE (30).

2.2.1.2 Frameworks by industry for biomarker evaluation

Biomarkers are valuable for industry. Drug toxicity is a major reason for drug failure at the clinical trial stage (7, 58). Biomarkers that could predict this would be very valuable, however there are risks associated with their use in drug development.

Biomarker qualification is susceptible to two types of error, type I and type II. In the context of biomarker qualification, these can be described as follows. In a type I error, the biomarker is judged to be useful even though it is not (false positive). In type II, a potentially useful biomarker is not qualified (7). The risks associated with each of these occurrences are perceived differently by different stakeholders such as patients, regulators, and industry (59).

Williams, *et al.* (2006) argued for a framework that evaluates biomarkers both in terms of their risk and cost-effectiveness (24). They argued that biomarkers are too often evaluated through perceived consequences of their failure, and this is not balanced sufficiently with the potential benefits of their implementation. The 'dread' of rare but serious consequences leads to overestimation of their frequency. Equally, the frequency of less serious events is underestimated (Figure 2.3).

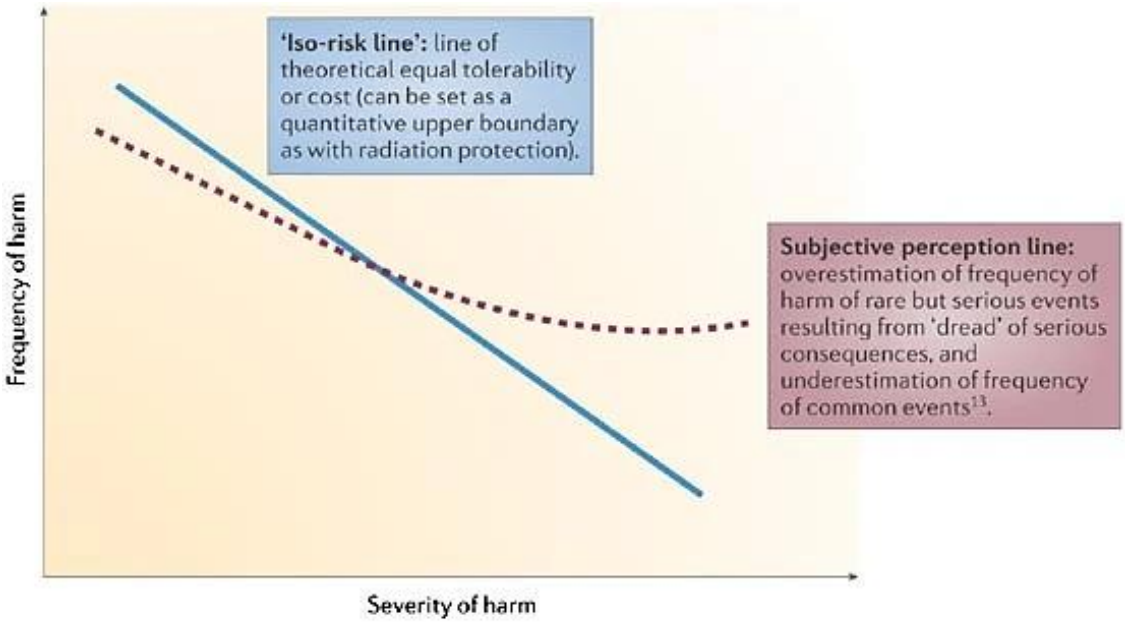


Figure 2.3 - How risks are perceived. The blue line can represent the true situation – events of high severity are rare, whereas events of low severity are more common. The dashed line represents perception – we tend to overestimate the frequency of high severity events and underestimate the frequency of low severity events. From Williams et al. (2006) (24)

Around the same time as this paper, another industry-authored paper provided a fuller perspective on biomarker validation (12). This paper provided some key recommendations for biomarker validation in several areas, but mainly focussed on analytic validity. The paper also recommended that the rigour of validation should be 'fit-for-purpose', meaning that early, exploratory biomarkers should require less rigour than one to be used for critical decision making. This is similar to the previously mentioned FDA perspective (10).

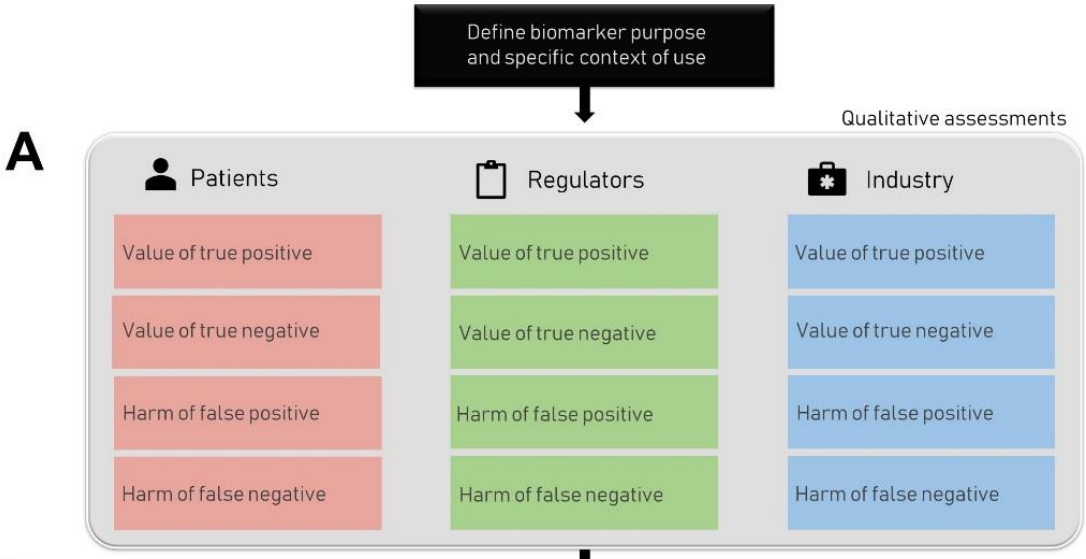
A more recent survey of industry perspectives on biomarker qualification found that regulatory complexity was seen as a bigger barrier to qualification than technical or scientific complexity (60). In this survey, participants were also asked to select which evidentiary standards should support biomarker regulatory qualification. Literature reviews and confirmatory studies were the most common types chosen. It would be interesting to take this work further by comparing and contrasting its conclusions with the views of patients and regulators to see whether those stakeholders desire different evidentiary standards for biomarker qualification. This sort of comparative evaluation was completed by a 2007 committee in the US (13).

2.2.1.2.1 Case study: PhRMA committee 2007

A committee of the Pharmaceutical Research and Manufacturers of America (PhRMA) produced a framework for the evaluation of biomarkers that incorporates patient, regulator, and industry perspectives (13). This process utilises a similar framework as Williams, *et al.* (2006), arguing for full evaluations of the benefits and harms of true and false positives/negatives. Once the purpose and specific context (e.g. for use in a life-threatening disease area) of use of the biomarker is agreed, 12 qualitative assessments are carried out. Here, the views of patients, regulators, and industry are sought to determine an overall value of truth and harm of falsehood for the biomarker (Figure 2.4A). This 'tolerability of risk' defines the next step, an evidence map which aims to outline the type and strength of evidence needed for biomarker qualification.

This evidence map step consists of seven domains (Figure 2.4B): theory on biological plausibility; interaction with pharmacologic target; pharmacologic mechanistic response; linkage to clinical outcome of a disease or toxicity; mathematics replication and confirmation; accuracy and precision (analytic validation); and relative performance. Evidence should be provided for each relevant domain. The map enables the grading of the quality of this evidence from grade D (least relevant evidence) to grade A (most relevant).

This approach provides a labour-intensive but individualised approach for defining evidentiary standards. The authors note that the process was tested in a July 2007 workshop, attended by key stakeholders. Untrained participants were able to complete assessments within one day. Further use of the process could not be located. Moreover, the framework does not include an assessment of the risk of bias. This is an important part of assessing evidence and is discussed later in this chapter.



B

Evidence map

Evidence type	Grade D	Grade D+/C-	Grade C	Grade C+/B-	Grade B	Grade B+/A-	Grade A
Theory on biological plausibility	Observed association only	Theory, indirect evidence from animals	Theory, direct evidence from animals	Theory, indirect evidence in humans	Theory, direct evidence in humans, non-causal pathway possible	As lower grades, but biomarker on causal path	Human evidence based mathematical model of biology showing biomarker on causal pathway
Interaction with pharmacologic target	Biomarker identifies target in *in vitro* binding			Biomarker identifies target in animals	Biomarker identifies target *in vivo* or in human tissue		As lower grade, and with accepted truth standard
Pharmacologic mechanistic response	Evidence *in vitro* that the drug affects the biomarker	Evidence *in vitro* that multiple members of this drug class affects the biomarker	Evidence *in vivo* that this drug affects the biomarker in animals	Evidence *in vivo* that drugs in this class affect the biomarker in animals	Human evidence that this drug affects the biomarker OR animal evidence of specificity	Human evidence across this mechanistic drug class	Human evidence that multiple members of the drug class affect the biomarker and the effect is specific
Linkage to clinical outcome of a disease or toxicity		Biomarker epidemiologically associated with outcome without intervention	Biomarker associated with change in outcome from intervention in another drug class	As lower grade, but in this drug class	As for lower grade but multiple drug classes, BUT inconsistent effect		As for lower grade but consistent and explains majority of disease effect
Mathematics replication, confirmation		Algorithm required to interpret the biomarker and was developed from this dataset		Algorithm developed from different dataset and applied prospectively			Algorithm developed from different dataset, replicated in other sets, and applied prospectively
Analytic validation				Sources of technical variation unknown but steps taken to ensure consistency	Major sources of variation are known and controlled to be less than biological signal, standardisation methods applied		All major sources of technical imprecision known and controlled. Accuracy defined against standards
Relative performance		Does not meet performance of benchmark		Similar to performance of benchmark			Exceeds performance of benchmark or best alternative biomarker

Figure 2.4 – Process proposed by the Pharmaceutical Research and Manufacturers of America (PhRMA) for determining the evidence required by a biomarker for qualification. Once the biomarker purpose and specific context of use is defined, 12 qualitative assessments are carried out to determine

the value and harm of biomarker testing outcomes. The evidence map is then used to create an evidentiary framework for qualification (13).

2.2.1.2.2 Personalised Precision Medicine Special Interest Group

A similar process was recently undertaken by the multi-national Personalised Precision Medicine Special Interest Group (59). This work, within a precision medicine context, aimed to describe the perspectives of patients, clinicians, hospitals, industry, regulators, payers, and policy makers on the evidence required for evaluating the value of a genetic biomarker. The information was taken from published literature and discussions in professional societies from 2010 – 2019.

Common to all groups were concerns about safety, efficacy, and affordable care. Patients were additionally concerned about privacy in relation to genetic data, something that echoes throughout the literature (61-65). Perspectives from clinicians focussed more on evidence. The availability of comparative evidence of utility across different tests was important, as was evidence-based decision support on test use and interpretation. Regulatory perspectives included the importance of evidence for test safety and efficacy, and opportunities for real world evidence from post-market surveillance. From these perspectives, the Group proposed some novel considerations for evidence in precision medicine, including potential for harm associated with a genetic test, payment requirements, and the wider health system effects of precision medicine. This includes the use of new technologies such as artificial intelligence and machine learning and their role in drug development and safety.

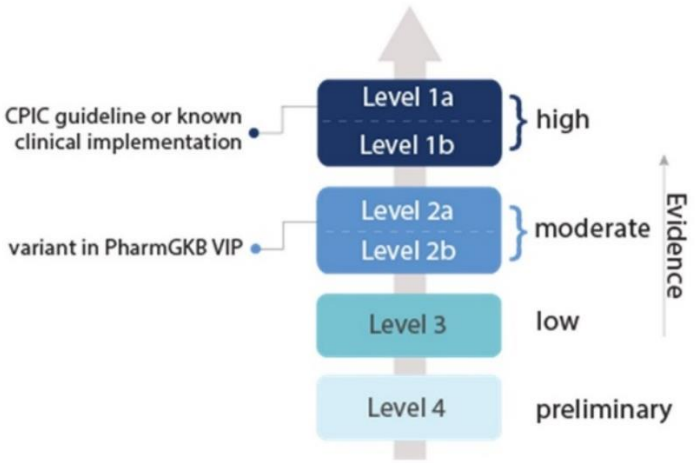
A specific framework for evidentiary standards was not provided in this article, but the authors did provide important considerations that should be taken into account by authors of future frameworks.

2.2.1.3 Summarising the evidence: PharmGKB and CPIC

The Pharmacogenomics Knowledge Base (PharmGKB) (66) and the Clinical Pharmacogenetics Implementation Consortium (CPIC) were set up from groups within the Pharmacogenomics Research Network (PGRN) (47). These groups both aim to summarise the evidence for utility of genetic biomarkers into easy-to-use resources, accessible to clinicians and researchers. They use similar systems of categorisation to define required evidence as the previously mentioned frameworks.

PharmGKB is an online resource that reviews and summarises pharmacogenetic evidence to support clinical annotation of variants, summarise biomarker-drug pathways, and publication of pharmacogenetic guidelines (46, 67-69). PharmGKB has over 4000 clinical annotations to date, and has 34 VIPs ('very important pharmacogenes') as 'Tier 1' genes – these are considered the most urgent and serious of the associations (70) (see Appendix Table 1). These have 'substantial evidence to support their importance in pharmacogenomics'. New genes are regularly added to the database. Each variant-drug interaction is curated by at least two reviewers (70). This process has produced a key resource for researchers and clinicians (46), but is labour-intensive and time-consuming. A new alternative is the automated text-mining process outlined by Lever, *et al.* (2020) (71). This will further improve PharmGKB as a source of pharmacogenetic information by allowing the automated addition of new research. The guidelines produced by PharmGKB are influential and they are widely used by clinicians and triallists (46).

PharmGKB collects and summarises evidence on genetic variant-drug associations (67). The evidence is rated on a six-point scale (Figure 2.5). Each level is well-defined with strict criteria. As of October 2020, the majority of genetic variant-drug combinations are in the Level 3 category (3525, or 77.4%), with 182 (4.0%) in the top evidence categories (Levels 1A and 1B). The quality of studies used as evidence is indirectly assessed using cohort size, effect size, and significance (p-value) as heuristics (72). There is no specific mention of assessing studies' risk of bias.



Level	Description	Number of variant-drug combinations*
Level 1A	A variant-drug combination in a CPIC or medical society-endorsed PGx guideline, or implemented at a PGRN site or in another major health system.	166
Level 1B	A variant-drug combination where the preponderance of evidence shows an association. The association must be replicated in more than 1 cohort with significant p-values, and preferably have a strong effect size.	16
Level 2A	A variant-drug combination that qualifies for level 2B where the variant is within a VIP as defined by PharmGKB. The variants in level 2A are in known pharmacogenes, so functional significance is more likely.	77
Level 2B	A variant-drug combination with moderate evidence of an association. The association must be replicated but there may be some studies that do not show statistical significance, and/or the effect size may be small.	72
Level 3	A variant-drug combination based on a single significant (not yet replicated) study or a variant-drug combination evaluated in multiple studies but lacking clear evidence of an association.	3525
Level 4	Annotation based on a case report, non-significant study, or in vitro, molecular, or functional assay evidence only	701

*Figure 2.5 - PharmGKB levels of evidence assigned to clinical annotations. The table shows the number of variant-drug combinations that are supported by each level of evidence, as of October 2020. *These data obtained from "Variant and Clinical Annotations Data", located at https://www.pharmgkb.org/downloads (67).*

The CPIC was set up in 2009 to provide specific guidance to clinicians and laboratories on pharmacogenetic testing in order to improve the difficulty of translating genetic results into clinical practice (73, 74). CPIC guidelines are written in collaboration with PharmGKB (68). Evidence from PharmGKB is chosen to be used in a guideline in a process based on clinical need, availability of strong supporting evidence, and the availability of genetic tests in a clinical setting (74, 75). All CPIC guidelines work from the underlying assumption that clinicians will one day have access to patients' genotypes before prescription as a matter of routine (74).

As of November 2020, CPIC has produced guidelines relating to 25 biomarker-drug pairs (76). Each guideline adheres to a standard format and synthesises information

from multiple sources on genes (including genetic test interpretation), drugs (link to genetic variability), and dosing. The quality of the evidence linking genotype to phenotype is also graded (77). Quality is assessed using a three-point scale (26):

- High: evidence includes consistent results from well-designed, well-conducted studies (although no standard for assessing this is provided)
- Moderate: evidence is sufficient to determine effects, but the strength of evidence is limited by the number, quality, or consistency of the individual studies, generalisability to routine practice, or indirect nature of the evidence
- Weak: Evidence is insufficient to assess the effects on health outcomes because of limited number or power of studies, important flaws in their design or conduct, gaps in the chain of evidence, or lack of information.

Despite the existence of this scale, specific details of how each category is assessed (e.g. what constitutes an 'important flaw' in a study?) could not be located.

Once these assessments are completed, the level of evidence is used to assign CPIC levels to specific gene-drug combinations. CPIC levels range from A to D (Table 2.2) (77), with higher grades for stronger recommendations based on higher quality evidence. Guidelines are normally only written for recommendations in levels A-B, although guidelines are sometimes published for those in level C (78).

As an example of this process, the association between *HLA-B*15:02* and carbamazepine is rated as level A, indicating that the evidence is so strong (and of good quality – although again, the methods for assessing this are not explicitly stated) that testing should be required before prescription (79). In contrast, level D includes an association between metformin and *C11orf65*, an association with little evidence for an association between variant and drug phenotype. This association is on Level 4 of the PharmGKB levels (80).

CPIC Level	Clinical context	Level of evidence	Strength of Recommendation	No. of recommendations*
Level A	Genetic information should be used to change prescribing of affected drug	Preponderance of evidence is high or moderate in favour of changing prescribing	At least one moderate or strong action recommended	72
Level B	Genetic information could be used to change prescribing of the affected drug because alternative therapies/dosing are extremely likely to be as effective and as safe as non-genetically based dosing	Preponderance of evidence is weak with little conflicting data	At least one optional action is recommended	58
Level C	There are published studies at varying levels of evidence, some with mechanistic rationale, but no prescribing actions are recommended because a) dosing based on genetics makes no convincing difference or b) alternatives are unclear, possibly less effective, more toxic, or otherwise impractical, or c) few published studies or mostly weak evidence and clinical actions are unclear	Evidence levels can vary	No prescribing actions recommended	95
Level D	There are few published studies, clinical actions are unclear, little mechanistic basis, mostly weak evidence, or substantial conflicting data. If the genes are not widely tested for clinically, evaluations are not needed	Evidence levels can vary	No prescribing actions recommended	101

*Table 2.2 - CPIC levels. *Number of recommendations in each category as of November 2020. The action recommended in this table is a change in prescribing. There are also levels between those shown. These are used at preliminary review stages. As of November 2020, there were 16 recommendations in A/B, 69 in B/C, and 5 in C/D. Data obtained from https://cpicpgx.org/genes-drugs/ (47).*

2.3 Regulatory bodies' views on biomarkers

The growth in the use of genetic biomarkers has strained regulatory processes that were designed for very different types of interventions (19). The complexity of regulatory processes has presented a significant barrier to industry in biomarker usage (60, 81).

A 2021 mini review outlined the views of the FDA and European Medicines Agency (EMA) on predictive biomarkers (82). The current recommendations given by the FDA focus on the amount of evidence for drug-biomarker pairings, sufficient to support medical decision making. This may involve decisions about safety, efficacy, and pharmacokinetics. Discussion of the EMA recommendations is briefer, as new regulations are due within 2022.

2.3.1 Medicines and Healthcare Products Regulatory Agency (MHRA), UK

The MHRA is responsible for the regulation of medicines, medical devices, and blood components for transfusion in the UK (83). The remit of the MHRA includes monitoring side-effects, reviewing evidence, and inspecting clinical trial sites (84). Presently, the MHRA works very closely with the EMA (85). The general regulatory framework currently used by the MHRA for biomarkers is based on the European *in vitro* Diagnostic Medical Devices Directive 98/79/EC (86).

The MHRA considers biomarkers and genetic tests as medical devices (3). A medical device is defined by the MHRA as “any instrument, apparatus, appliance, software, material or other article used alone or combined for humans to: diagnose, prevent, monitor, treat or alleviate disease; diagnose, monitor, treat, alleviate or compensate for an injury or handicap; investigate, replace or modify the anatomy or a physiological process; control conception” (87).

Three categories of medical device are defined: active implantable medical devices, general medical devices, and *in vitro* diagnostic medical devices (IVDs) (87). The first two of these are classified by the level of risk associated with their use (Figure 2.6B), assessed using 18 rules that assess invasiveness, local or systemic effects, and duration of use (14). These categories range from low risk, class I (everyday items such as plasters) to high risk, class III (implantable cardiac catheters). IVDs (under which most genetic tests and biomarkers fall) are grouped into four categories based on risk and specific usages (Figure 2.6B). Different evidentiary standards are required for the different categories.

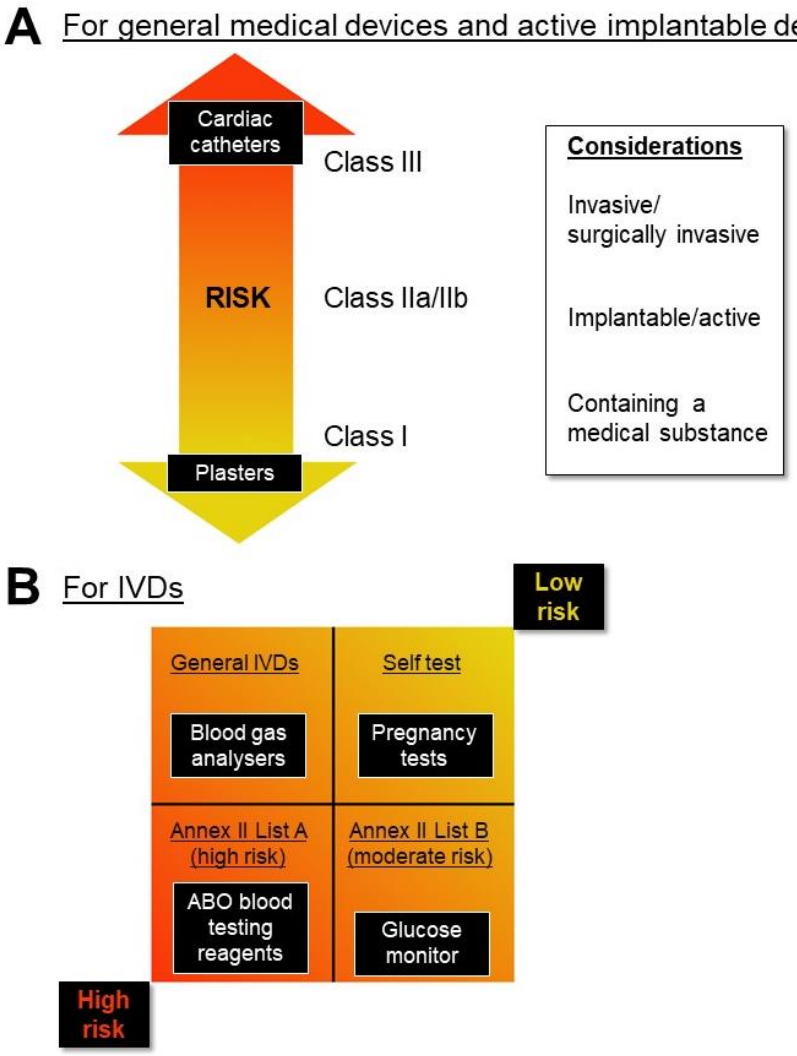


Figure 2.6 - MHRA risk categories for A) general medical devices and active implantable devices and B) IVDs. Black boxes contain examples, from the MHRA, of devices that fit within each category. Considerations box describes factors that go into the decision-making process (88). IVD = in vitro diagnostic.

The lowest risk category of the four is for 'self-test' items, such as home pregnancy tests. In contrast, the highest risk category contains devices for the assessment of blood groups and of blood borne diseases (89). These require additional evidence and scrutiny because many of the diseases are infectious and notifiable (need to be reported to the UK government since they present a significant risk to human health) (90). This is important for the assessment of biomarkers for monitoring HIV and hepatitis.

All IVDs must be registered with the MHRA (87). All categories other than 'general IVDs' must also have a conformity assessment carried out by a notified body. These

bodies are expected to assess whether manufacturers and medical devices meet the requirements set out in the legislation (91).

Despite being classed as IVDs, genetic tests are not specifically referred to in this list of devices. New guidance is clearly required, and these guidelines are currently being replaced by the new regulations, the Medical Device Regulation (MDR) and the In Vitro Diagnostic Medical Device Regulation (IVDR). These will be fully in place from 2022 (92). These regulations map existing categories into new classifications (Figure 2.7). Most biomarkers will now fall into the category of in vitro diagnostic medical devices. The available guidance suggests that most genetic tests will fall into Class C.

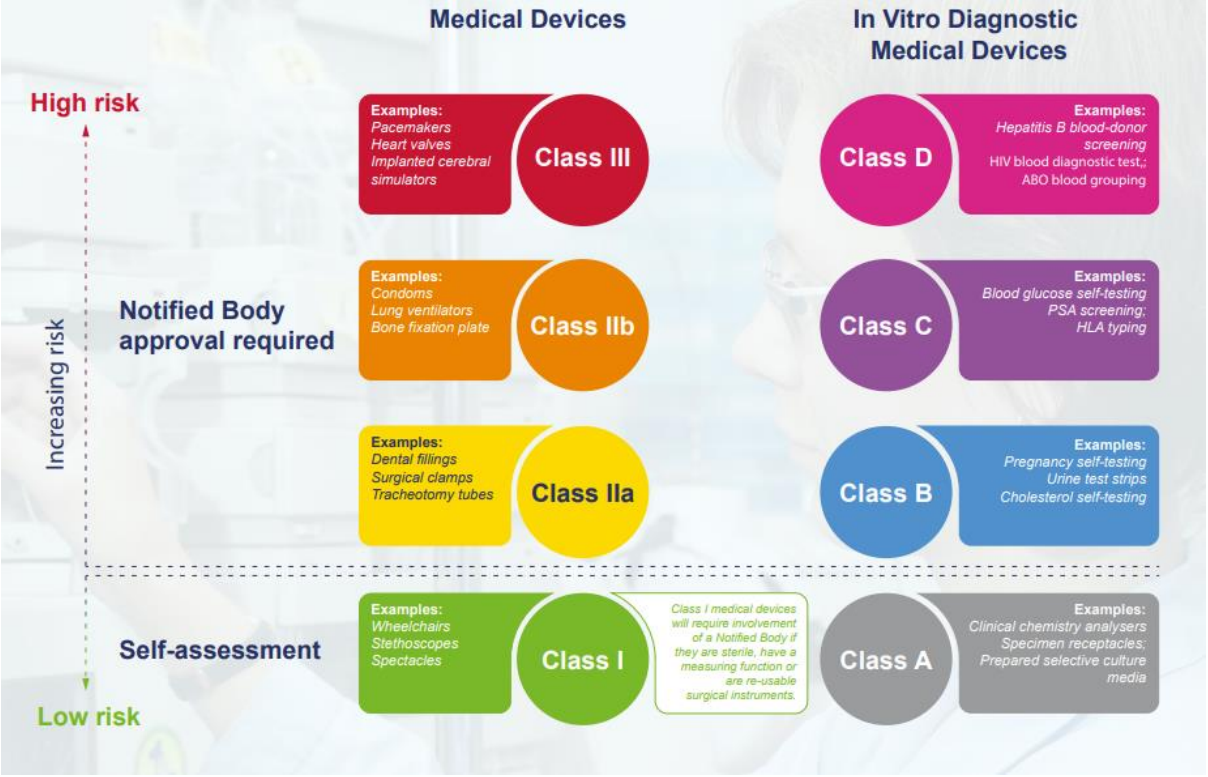


Figure 2.7 - new MHRA medical device and in vitro diagnostic medical device categories (92)

The approval of a notified body is still required for devices in classes B, C, and D. This approval requires a conformity assessment, which includes calibration, testing, certification, and inspection (92). Details of the technical requirements for receiving this approval can be found in Chapters I and III of Annex IX of the Regulation 2017/746 of the European Parliament (89). Class C and D devices are also required to submit periodic (at least annual) safety update reports throughout the lifetime of the device. This should include analyses of post-marketing surveillance data and details of the usage of the device (93).

The MHRA collaborates closely with the EMA (85). However, the uncertainty around the UK's exit from the EU ('Brexit') has made it difficult for regulators to plan for the future (94). The MHRA has published some advice on this, including general advice (covering marketing authorisations, import/export, and pharmacovigilance) (95) and advice specific to medical devices (96). Based on an amendment to the Medical Devices Regulations 2019 (EU Exit) (2019) the MHRA participation in the European regulatory network would end. However, some EU Directives on medical devices (directives 90/385/EEC, 93/42/EEC, and 98/79/EC) have already been transposed into UK law under the Medical Devices Regulations 2002 (97). So, although regulation post-Brexit is unclear, some EU directives are likely to still apply. The Medicines and Medical Devices Bill was introduced to parliament in 2020 (98). This bill will give the UK government powers to update existing regulatory frameworks.

The EMA published specific guidelines on pharmacogenetics and genetic biomarkers during the drug life cycle in 2013 (99, 100). The exact format for a submission package for biomarker qualification varies depending on the biomarker context, but in general includes several key attributes (Table 2.3) (100), based on requirements of the International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (101). These do not explicitly specify that participant ancestry should be included in the evaluation, and the importance of this is further discussed in Chapter 3.

Summary of the ICH guideline E16 on genomic biomarkers related to drug response: context, structure, and format of qualification submissions

Section 1: Regional Administrative Information	Documents specific to each region. This can be specified by relevant regulatory authorities	
Section 2: Summaries	**2.2.1. Biomarker qualification overview**	**2.2.2. Overall summaries**
	Introduction, Context of use, Data description, Critical appraisal of data and methods, Any additional data needed, Justification for the context of use	Analytical, non-clinical, clinical (as appropriate), Synopses of individual studies
Section 3: Quality reports	Product quality and manufacturing data	

Section 4: Nonclinical reports **Section 5: Clinical reports ***	Full study reports for biomarker qualification
	Information on compliance with GCP
	Number and classification of patients in study, Performance characteristics of the biomarker test used, Variables impacting on assay validity and interpretation, Methods used for analysing raw data, Criteria for determining sample quality, Methods used for determining gene expression, Criteria used for selection of candidate genes, Results of analyses of genomic biomarkers to international standards, Expert statements Evaluation reports issued by regulatory authorities Manufacturer technical descriptions Published articles in peer-reviewed journals (including meta-analyses)

*Table 2.3 - The information in this table comes from (100). *Nonclinical and clinical reports are listed together in the guidance.*
GCP = good clinical practice

Evidence for the performance characteristics of the biomarker test can be based on retrospective and/or prospective correlation with either nonclinical or clinical data (100). While submitters are required to provide a justification for the biomarker context of use, no examples are given of what information would be sufficient for this purpose.

However, a 2011 reflection paper published by the EMA acknowledged that evidentiary burdens are different depending on the circumstances of genomic biomarker usage (27). However, they state that “confirmation of findings obtained from early signal generating studies in a prospective pivotal clinical trial is expected”. The EMA also states that where such a trial is not possible, evidence from well-conducted case control studies, observational or epidemiological studies “might also serve the purpose” (27). In situations where the main evidence for a biomarker is retrospective, there are four requirements for the evidence to be persuasive:

1. The strength of the association should be high
2. The biological plausibility for the interaction should be strong
3. The marker status of the majority of the subjects in the dataset should be known, to avoid bias

4. The diagnostic performance of the marker for the measured outcome should be of acceptable level (27).

This paper also notes that bias is more likely in retrospective studies and needs to be reduced with proper study design and execution. Selection bias and measurement bias are singled out as particular risks. The authors also warn that a large sample size alone is not sufficient to remove bias (27).

These guidelines, only applicable where a prospective pivotal trial is not possible, provide a good basis for triallists planning trials for submission to European regulators. The evidentiary standards detailed above could be met using combinations of *in vitro* work and observational studies. However, this guidance is now almost a decade old, necessitating an update to reflect newer technologies. An update to these regulations is due in May 2022 (82), which will clarify the risk-based classification system, improve transparency, and enforce a unique device identifier system (102). Manufacturers have until May 2025 to fully comply. However, these guidelines also do not comment on any necessary quality of evidence, or assessment for risk of bias, something that is an essential additional check on a biomarker's readiness for trials or clinical use (103).

In the UK, many trials are run through the National Health Service (NHS). Policies from the NHS, Health Research Authority (HRA), and Department of Health are therefore influential in designing research. The UK Policy Framework for Health and Social Care Research, enacted in 2017, defines 19 principles to be followed by interventional health and social care research (104). Principle 16 states that any intervention must be "adequately supported by the available information (including evidence from previous research)" (104). This is the responsibility of the Chief Investigator, the research team, and the sponsor. While there is no specific guidance on what evidence is appropriate, the NHS Health Research Authority (HRA) recommends a systematic review is undertaken before setting the research question for a project. The information gathered should then be used in the design of the project (105).

2.3.2 Food and Drug Administration (FDA), USA

The FDA regulates medicines, medical devices, vaccines, food, cosmetics, tobacco, and other products in the United States (106). It is an agency within the US Department of Health and Human Services (107). As the regulator for the largest national drug market in the world by value (108), the FDA has a significant influence on drug regulation around the world. In 2004, the FDA issued a landmark report that

highlighted the potential of biomarkers to “drive rapid clinical development” and the importance of biomarkers in drug safety and effectiveness (109). This was followed up by the Biomarker Qualification program (see below) and guidance on biomarker qualification in drug development (31, 110, 111).

The Biomarkers, EndpointS, and other Tools (BEST) resource was then produced in 2016 by an FDA-NIH Working Group (25). The site includes FDA-approved definitions of biomarker-related terms, as well as information on the different types of biomarker and their application. The FDA uses the 2001 Biomarkers Definitions Working Group definition of a biomarker (112). The agency further defines a composite biomarker as “several individual biomarkers that are combined in a stated algorithm to reach a single interpretive readout” (113).

The FDA categorises biomarkers by their context-of-use (COU) (10, 25). The COU is the most important factor used to determine the level of evidence needed for biomarker qualification (114, 115). As previously discussed, the required level and quality of evidence also increases when there are greater risks associated with the biomarker’s use, or if it will be used to make critical decisions (10, 116). There are several ways to reach this required level of evidence.

Diverse types of data have been used for biomarker qualification. Data can be “retrospective or prospective, registry data, and/or randomised controlled trial data and should include an exploratory dataset and a confirmatory dataset” (117). A 2014 FDA paper stated that data for a premarket application may come from “clinical trials, appropriately curated databases, published literature, and/or other sources of valid scientific evidence”, and offered an example of a rare cystic fibrosis variant given clearance based on the use of a “well-curated third party database” (5, 118).

There are two ways a biomarker may then be accepted by the FDA for use in drug development (Figure 2.8).

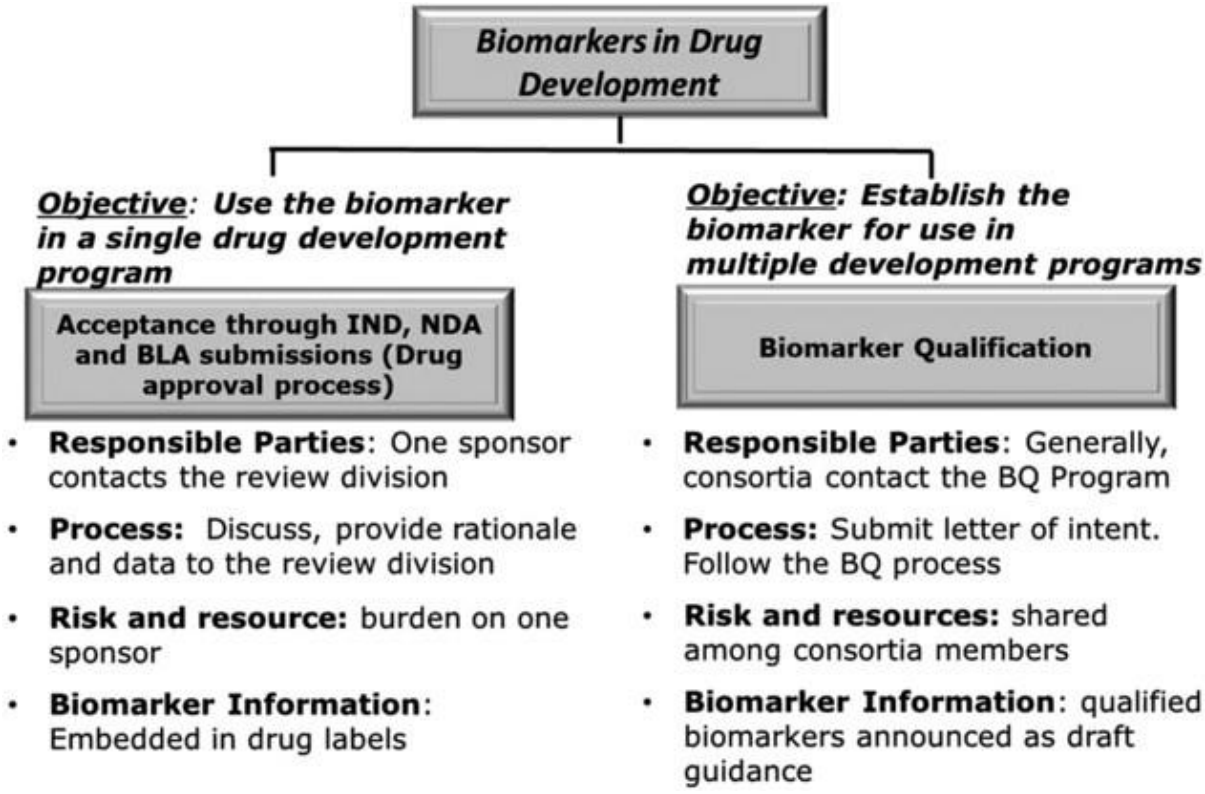


Figure 2.8 - FDA biomarker process, with two options for submitters. The first is the traditional option for using a biomarker in a single drug development program. The second is the dedicated Biomarker Qualification Program. BLA = Biologic Licence Application. BQ = biomarker qualification. IND = investigational new drug. NDA = new drug application. From Amur, et al. (2015) (10).

In the first pathway (left), a drug developer may reach an agreement with the FDA *during* the drug development process to allow the use of the biomarker clinically (10). In this pathway, developers present data through existing investigational new drug (IND), new drug application (NDA) or biologic licence application (BLA) processes. Full data submissions are required where biomarkers will be used for decision making in a clinical trial, for safety, efficacy, dosing, or pharmacology, or where the biomarker will be used on a drug label (116). This pathway also encompasses the Voluntary Genomic Data Submission (VGDS), a process whereby the FDA encourages developers to submit data voluntarily where IND approval is not required (such as for exploratory or research only biomarkers) (20). This first pathway is efficient for the drug developer but does not allow wider scientific scrutiny or peer review of the proposed biomarker.

The second pathway (right) is the Biomarker Qualification Programme (BQP) created by the FDA Center for Drug Evaluation and Research (CDER). This was created in 2009 to help biomarkers be developed for use in the drug development

process as tools by providing guidance and information to industry (114, 115, 119, 120). One of the aims of the programme is to enable transparency by making information on qualification publicly available (110). During this programme, a Biomarker Qualification Review Team (BQRT) will guide submitters through the initiation, consultation, and review stages of the BQP (114). In the initiation stage, the submitter sends a letter-of-intent including information about the biomarker and its potential use. The BQRT then makes recommendations for the submitter to address, before deciding whether to accept the biomarker into the BQP. If accepted, the BQP moves into the consultation stage. Here, the submitter must forward an initial briefing package that includes further information on the biomarker. The BQRT holds a formal meeting with the submitter to guide further biomarker development. The review stage takes place when the submitted data are “complete and adequate” (115). A qualification recommendation is made after reviewing the data, internal meetings, and possible additional information requests (114). The entire BQP process takes 2-3 years and has been criticised as being expensive and labour-intensive (110). However, once a biomarker is qualified, it can be used in drug development programs under the stated COU (114). Some examples of biomarkers approved through this process include plasmodium RNA/DNA measurement for the diagnosis of malaria, and total kidney volume as a prognostic biomarker for polycystic kidney disease (121). As of October 2020, no pharmacogenetic biomarkers have yet been approved through this programme.

More recent FDA approvals related to biomarkers include pembrolizumab and larotrectinib. These are notable as they were approved for any solid tumour harbouring specific biomarkers, regardless of tumour histology (34, 122-124). Pembrolizumab (Keytruda) was approved in 2016 for any solid tumour with high microsatellite instability or mismatch repair deficiency (123). This was the first drug approved by the FDA on the basis of mutations rather than a specific type of cancer (123) (although pembrolizumab was previously approved solely for metastatic non-small cell lung cancer (NSCLC) (125)). A 2016 statement by the FDA cited one trial that contributed to pembrolizumab’s approval, known as Keynote 001 (125). Keynote 001 was a complex Phase 1, open-label trial that included multiple amendments to add cohorts and subgroups (126-128). Keynote 001 was primarily a dose-escalation study, enrolling 550 NSCLC patients for treatment with pembrolizumab. The objective response rate (the proportion of patients with tumour size reduction of a predefined amount and for a minimum time period (129)) was 41% (95% CI 28.6-54.3%) (125). Pembrolizumab was later approved for treatment

of microsatellite-instability high (MSI-H) or mismatch repair deficient (dMMR) tumours, based on data pooled from five non-randomised enrichment basket trials (34, 130). MSI-H and dMMR tumours contain large numbers of mutations, including *PD-1* and *PD-L1*, the targets of pembrolizumab. The assays for detecting these tumour types were not standardised at the time of this 2017 approval, but was justified by the high unmet medical need for drugs targeting these pathways (34, 130).

Pembrolizumab was approved through the FDA accelerated approval program and granted priority review. A recent analysis of FDA approvals from 1983-2018 found that the mean annual number of new drug approvals has increased, while more and more drugs are approved through 'special' programs like the Orphan Drug Act (1983) and the Priority Review program (1992). Submitting drugs through these programs allows drugs to be approved with less supporting evidence (131). Drug efficacy claims made to the FDA are required to be supported by "adequate and well-controlled" trials. Darrow, *et al.* (2020) argued that this statement has become more flexible over the years (131). They provide the example of the 1962 drug approval statute requiring two adequate and well-controlled randomised trials, in contrast to the 1997 codification of the previously informal practice of accepting just one (131, 132).

A 2014 analysis of drug-biomarker pairs in the FDA Table of Pharmacogenomic Biomarkers in Drug Labelling found that only a minority provided convincing evidence for clinical utility and validity (9). The authors defined 'convincing' evidence as a systematic review/meta-analysis of RCTs showing consistency in their results, or at least one large RCT. Of 119 drug-biomarker combinations, only 43 (36.1%) provided 'convincing' evidence of clinical validity, and 18 (15.1%) did the same for clinical utility. These data pose the question of whether the FDA biomarker evidence evaluation process is fit-for-purpose. The authors proposed that a statement about the quality of evidence should be presented in drug labels, and that biomarkers should only be included in labels in the first place if 'compelling clinical utility information' has been generated.

This analysis was repeated in 2017 with similar results (133). These analyses suggest that the minimum standards for approval of biomarkers are unclear and so wide variation in levels of evidence is seen. There is a large body of literature on evidentiary standards yet no clear guidelines from regulatory agencies could be

located. A unified position on these issues is required to ensure patient safety and regulatory consistency.

Draft guidance published by the FDA in 2019 gives guidance on how to demonstrate evidence of effectiveness for industry (132). While not specific to biomarkers, the guidance offers details on the quality and quantity of evidence the FDA may accept for industry to prove the effectiveness of drugs and biological products (Figure 2.9). For example, in regards to trial designs, the use of a control group ‘generally provides strong evidence of effectiveness’, although bias could be introduced if blinding is not adequate. This reduces the quality of the evidence. The guidance includes an acknowledgement that just one well-controlled randomised trial may be used as evidence, along with appropriate confirmatory evidence. There is also provision for situations where RCTs may not be ethical or feasible, such as in rare diseases (< 200,000 cases in the USA per year) or conditions caused by toxic substances. In these cases, observational studies or work in animals may constitute acceptable evidence.

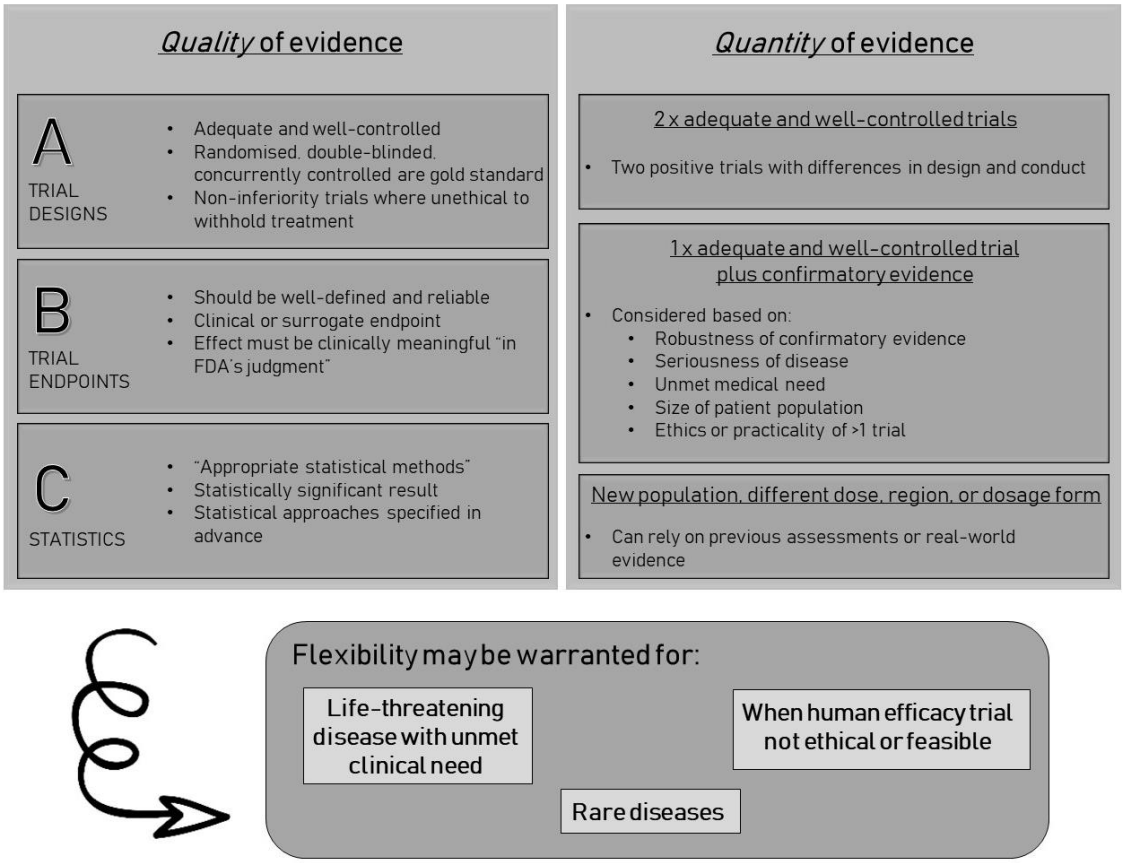


Figure 2.9 - FDA recommendations for industry on the quality and quantity of evidence required for demonstrating the effectiveness of human drug and biological products. Note that this paper is stated to contain 'nonbinding recommendations'. FDA, 2019 (132).

However, these recommendations are geared towards the regulation of drug products. A similar guidance for the use of biomarkers in clinical practice would be a valuable addition to the FDA regulatory arsenal.

An extreme example of regulatory challenge is the drug milasen, developed in 2019 for a single patient (134). An FDA editorial describes how the approval process was fast-tracked due to the patient's deteriorating condition (135). One month of animal studies and *in vitro* work were used as evidence for approval. This extreme end of personalised medicine also needs to be accounted for in regulatory development.

2.4 Challenges with genetic biomarker use in clinical trials contexts

As shown, regulators normally require clinical trial evidence prior to the approval of biomarkers for clinical use, and many innovative trial designs have been proposed for this purpose (15, 82, 136-139).

Freidlin, *et al.* (2012) proposed four potential designs for phase III trials after a biomarker has been shown to be useful in phase II trials (140). These designs (along with many others) are explored further in the online tool Biomarker-Guided Trial Designs (www.BiGTeD.org) (136). This tool, based on systematic reviews of the literature (141, 142), was developed to aid in the design and analysis of biomarker-guided trials by providing interactive overviews of trial characteristics, methodology, and evaluation of their advantages and disadvantages. The designs, based on systematic data extraction from 211 papers, are divided into adaptive and non-adaptive designs. Adaptive designs allow planned modifications during the trial, while non-adaptive are more traditional, fixed designs (Table 2.4) (136).

Adaptive design		Non-adaptive	
Pros	Cons	Pros	Cons
✓ Efficiency improvements (time, recruitment, costs) (143) ✓ Can increase trial attractiveness to participants (143) ✓ Optimise resource utilisation (144) ✓ Possibly more ethical (144)	× Greater design complexity (143) × Risk of type I/II error if trial is stopped early (145) × Some negative ethical implications (144) × Effective and specific trial infrastructure required (143) × Possibility of selection bias in investigators (144)	✓ Are often simpler to run (143) ✓ Better for long-term outcomes (143) ✓ Simpler to present to participants (144)	× Cannot easily adapt to new information (145) × Can be less efficient (time, recruitment, costs) (143) × More participants may receive ineffective interventions (144)

Table 2.4 - advantages and disadvantages of adaptive and non-adaptive trial designs

An example of adaptive design is the multi-arm, multi-stage (MAMS) design (Figure 2.10A). In a MAMS trial, there can be as many trial arms as needed, each evaluating a different intervention. In a biomarker context, patients within each arm, with a specific biomarker, are randomised to receive the experimental treatment or a control (which may be common to all arms). Each arm is effectively a ‘mini-trial’. Results are analysed at pre-defined interim analysis points, and any arm where the experimental treatment is inferior can be dropped, while the larger trial continues. A successful example is the ongoing STAMPEDE trial in prostate cancer, which has enrolled over 11,000 participants over 11 trial arms (146, 147).

An example of a non-adaptive design is an enrichment trial (Figure 2.10B). In an enrichment trial, a population is screened and the study population is selected from those who possess the required biomarker (137). In this way, the study population is enriched, particularly useful when the biomarker of interest is a rare variant. This design is also used when there is evidence that a treatment is beneficial only in a biomarker subgroup. Strong evidence for the biomarker's utility is required before choosing this design (34). The SHIVA trial is an example of an enrichment design. In SHIVA, patients with any cancer type were screened and entered into the trial only if their tumours possessed a mutation in one of three available pathways. Patients were then randomised 1:1 to receive biomarker-guided or physician-guided treatment (17).

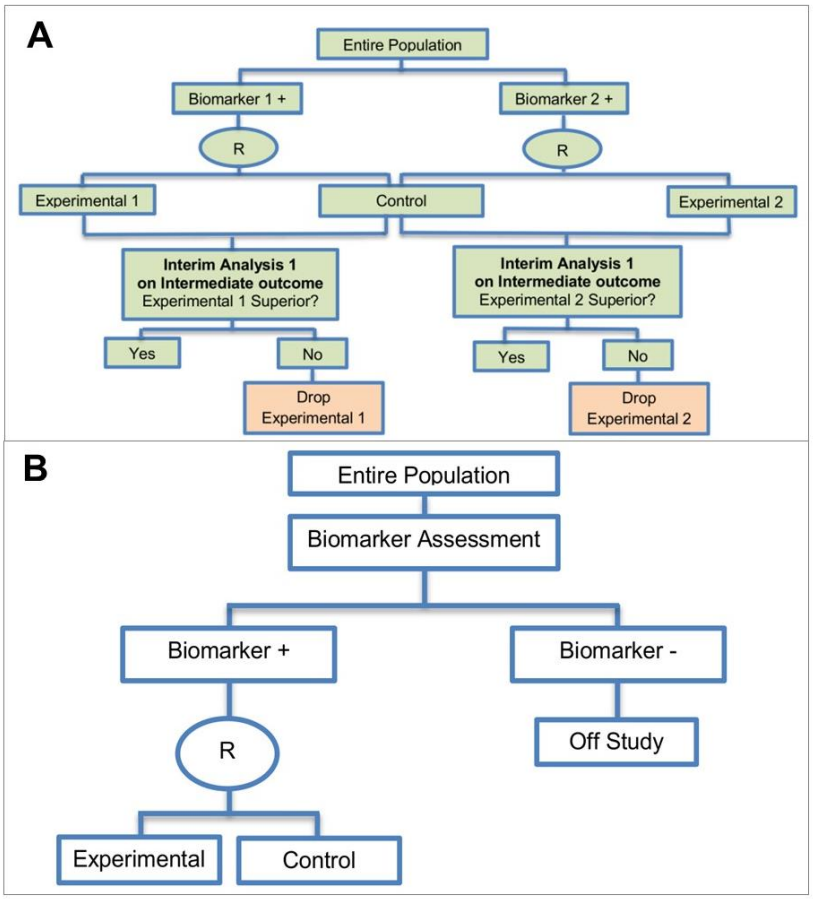


Figure 2.10 – Example of two BiGTeD designs. A) Multi-arm, multi-stage (MAMS) trial design, used for evaluating multiple interventions at once. B) Enrichment design, where only patients with the biomarker are randomised to receive the intervention or control, and those without the biomarker do not progress with the study. R = randomisation. From www.bigted.org (136).

Despite the many different options for designing a biomarker-guided trial, conducting clinical trials in pharmacogenetics can often prove difficult. One difficulty

is that genetic biomarkers are often very rare, meaning that demonstrating clinical validity and utility can quickly become impossible as a variant's rarity increases (5, 19, 148-152). Early examples of personalised medicine (e.g. trastuzumab in HER2+ breast cancer) had the advantage of having large populations for testing (8). The rarer the mutation, the more difficult it is to design and recruit for a study with sufficient power and for extremely rare variants, even a trial cohort of 1 million people would be too small to establish validity (19).

Further issues with trials using biomarkers include: increased costs, ethical issues (including disclosure to participants, incidental findings), recruitment problems, and the need for specialist staff (for a full review see Antoniou, *et al.* (2019) (153)). While RCTs are the 'gold standard' of evidence (154-156), they are not always suitable for pharmacogenetics. Questions have also been asked about the ethics of randomising patients who carry known actionable mutations to a drug or dose that could cause an ADR (151). For this and other reasons (mainly in scenarios where a variant or phenotype is very rare), an RCT is not always an appropriate tool (150, 151). There are other methods for assessing the utility of a genetic biomarker, and this issue is discussed further in Chapter 4.

The BiGTeD project is one example of trial methodologies adapting to the new challenges of pharmacogenetics. These sorts of biomarker-guided trial designs often fall into the definition of 'complex clinical trials' used by the Clinical Trials Facilitation and Coordination Group (CTFG), a group within the European Heads of Medicines Agencies (HMA), an independent network that advises the EMA (157, 158). A complex clinical trial is defined as a design consisting of separate parts, that would themselves constitute individual clinical trials, and/or one that has extensive prospective adaptations planned (such as the addition of trial arms) (157). These are often defined by the inclusion of biomarkers. A 2019 recommendation by a working group of the HMA stated eight key recommendations for the design and conduct of complex clinical trials (157) (Table 2.5). These recommendations are likely to become widely used in the European Union, and stakeholders conducting biomarker-guided trials should ensure compliance.

Additionally, a 2020 consensus statement by a group of UK researchers provided 10 recommendations for the conduct of complex innovative design (CID) cancer trials (159). These are also applicable to other disease areas. Both of these sets of guidelines make similar recommendations, valuing patient and public involvement, *a priori* trial planning, and encouraging dissemination of results (Table 2.5).

HMA CTFG (157)	UK CID trials working group (159)
Clearly describe and justify design	Engage with regulators
Maintain scientific integrity	Standards for protocol development
Ensure quality of trial conduct and optimise clinical feasibility	Patient and public involvement
Ensure safety of trial subjects	Standards for patient-facing documentation
Maintain data integrity	Statistical considerations, heavier workload
Reassess benefit-risk balance at critical steps throughout clinical trial	Define leadership and oversight
Validate companion diagnostics	Disseminate results
Consider data transparency	Consider higher levels of staff training
	Regulators to consider post-marketing commitments from manufacturers
	Public health impact analyses

Table 2.5 - Comparison of HMA CTFG and UK CID trials working group recommendations for complex trial designs. CID = complex innovative design. CTFG = Clinical Trials Facilitation and Coordination Group. HMA = Heads of Medicines Agencies.

These new frameworks will form the foundation of regulatory processes in the future, particularly as trials continue to become larger and more complex (122, 160).

2.5 Discussion

This review of the methods available for assessing genetic biomarkers and the views of regulatory authorities reveals shortcomings in the available guidance. While many working groups and committees have formulated guidance, the resulting guidelines form a patchwork of conflicting guidance that is a significant barrier to the wider implementation of biomarkers and pharmacogenetics (12).

The ACCE guidelines provided an excellent basis for many national and international assessment efforts (45). This was built on and considered by many other groups, including the PhRMA committee (13). These efforts paved the way for the creation of the PharmGKB and CPIC resources, used for detailing and summarising the totality of the evidence behind genetic biomarkers (47, 67).

A lack of clear regulatory standards is a significant barrier to the wider implementation of pharmacogenetics (21, 161-164). Regulatory opinions on biomarker evidence standards were more difficult to locate and examine. From analyses of the levels of evidence behind FDA drug labels, it is clear that there is significant heterogeneity in the evidence that has allowed biomarkers past the approval process (9, 133). These studies raise further questions about the suitability of current regulatory processes for modern biomarker research. For example, the rapid changes in these fields require faster approvals and subsequent drug label updates (133).

There are many challenges involved in regulating genetic biomarkers. Firstly, although evidence standards are unclear, trials of some kind are undoubtedly required. Designing trials that are able to incorporate biomarkers and adapt to rapid changes as the research evolves is one of the central challenges of evidence gathering in this area (153, 165, 166). Precision medicine necessitates the recruitment of smaller populations with specific characteristics. This is a challenge for designing trials that will be sufficiently powered to detect meaningful effect sizes. Innovative trial designs are required (136, 153), as is regulatory approval of these sorts of trials. There needs to be sufficient flexibility in the evidence requirements in cases where full randomised trials are not possible (150, 151). An example of this is when an outcome or variant is very rare. In these cases, it is not always possible to collect the RCT data normally required for regulatory approval. Combining data from well-conducted observational studies is a potential method of solving this problem (167, 168). I have explored this issue further in Chapter 4.

Secondly, there is a need for guidance that is applicable internationally, to reflect the multinational drug market and trial recruitment areas (169, 170). At present, different regulatory agencies and frameworks use different terminologies and criteria to describe biomarkers, evidence, and trial standards. This may even require agencies to produce two sets of guidelines – one for drug development, and another for clinical implementation.

Third, there is very little discussion of the quality of evidence or risk of bias in any of the listed frameworks. The framework produced by the PhRMA committee includes assessment of quality in their evidence map, but there is no discussion of the risk of bias of that evidence (13). PharmGKB provides details of how quality is assessed, but not risk of bias, and CPIC guidelines provide neither. The EMA guidelines for evidence where RCTs are not possible does account for the quality of evidence and

takes steps to address the risk of bias. Since a regulator does assess these issues, and because of the relevance of such an assessment to the reliability of any estimates of effect, it would clearly be useful for PharmGKB and CPIC, widely used frameworks, to incorporate these into their work.

Finally, the cost and cost-effectiveness of many of these interventions needs to be considered, particularly in a wider societal context (13, 24). Calculating the monetary values of these costs and benefits is difficult in practice (13, 171). This necessitates the development of a framework that, while acknowledging cost-effectiveness, is not defined by it. The ACCE framework did include an economic evaluation domain, which assesses the costs associated with testing and the economic benefits associated with testing (28). However, this is a small part of the overall evaluation of the biomarker, in a societal sense. Industry and drug developers clearly have a strong interest in cost. Frameworks proposed by these groups focus heavily on the cost associated with the usage (or non-usage) of biomarkers in trials or clinical practice (13, 24).

This conflict echoes more widely throughout the field of evidence gathering. There is a clear incentive for industry to suggest genes as ‘actionable’ with weaker evidence than other stakeholders would require. Actionable tumour genes are found in industry trials more often than in government-funded trials (8, 172). This underlines the need for regulatory intervention in industry-focussed evaluations of biomarkers.

One possible solution to these issues would be the setting of minimum standards. A regulator could provide a unified evidence framework similar to that of the 2007 PhRMA Committee, and state that a biomarker under prognostic, predictive, diagnostic, or response categories should meet at least, e.g. Grade B. Flexibility could be granted in certain settings such as in diseases with unmet clinical need. Ideas from rare disease regulation could be applied to biomarker-guided trials, where the biomarker positive population is small. For example, the FDA allows one RCT to be used as evidence (as opposed to the usual two) in some cases. They also allow observational studies or *in vitro* work to be used as sole evidence where an RCT is not ethical or feasible (132).

A 2014 FDA paper discussed the clearance of a variant for clinical testing based on a “well-curated third party database” (5, 118). The term “well-curated” is key here – evidence needs to be of sufficient quality to be applicable. However, the standard of this quality is not formally defined by the FDA, leaving applicants unclear about requirements.

A unified framework, based on a full systematic review of existing frameworks, would make the process of evidence gathering clear and transparent to triallists, clinicians and industry. The well-designed and validated work of PharmGKB and CPIC should be used to frame recommendations. The framework would also need to include provision for frequent updates as the field progresses. Innovative and well-designed trials will form the cornerstone of this framework. International considerations, including testing in multiple ancestry groups, would also form part of this framework, as would cost-effectiveness calculations. The creation of this framework by an influential regulatory agency such as the FDA or EMA would be the impetus for its use in other parts of the world. Importantly, the assessment of study quality and risk of bias should be incorporated into this unified framework.

It is also ethically and clinically important to account for the preferences of patients and the general public when making regulatory decisions. The PhRMA committee included patients in its discussions (13), and the FDA have stated their belief that patient experiences can help evaluate the benefit-risk profile of new interventions (173). Policy making around patient preferences benefits from quantitative methods, such as the discrete choice experiment (DCE) (173-175). I have explored this further in Chapters 6 and 7.

There is evidence that the regulatory environment is changing. The approval of pembrolizumab by the FDA for use in any tumour showing a particular mutation is an important development (34). While the minimal evidence used in this approval is notable, it is also important to realise the shift that has taken place in the approval of a histology-agnostic drug. The indication for pembrolizumab being defined by biomarker is a new direction for regulatory authorities that will undoubtedly become more common in the future (34).

2.5.1 Strengths and Limitations

This review focuses on methods available for the assessment of biomarkers and published regulatory views of biomarkers. I include discussion of frameworks developed by both academics and industry, and provide an overview of the widely used PharmGKB and CPIC resources. This combination of resources provides several powerful options for assessing biomarker evidence.

Another strength of the review is the focus on clinical trials and evidence, a discussion missing from many reviews of evidence gathering. While the need for

trials as evidence is often discussed, the unique challenges of conducting trials with genetic biomarkers are often not addressed.

One of the limitations of this chapter is the focus on the UK and US regulatory systems. Other systems may have very different methods for assessing biomarkers and/or provide more detailed guidance about their use. However, I focussed on the UK as this is my home system. I additionally included the US FDA as this has been called the “most powerful regulatory agency in the world” (176). Further, the focus on EU regulations applies to the systems of many European countries.

This project was not a full systematic review of all available regulatory guidance. A full systematic review would be a challenging but worthwhile project in the future. This would allow an unbiased view of current guidance and show any gaps in regulatory systems. A piece of work similar to that of Pitini, *et al. (45)* but also incorporating the perspectives of multiple stakeholders (including patients, clinicians, regulators, and industry) could form the basis of recommendations.

2.6 Conclusion

Ultimately, the lack of a standardised evaluation framework or pathway for biomarker qualification leaves individual clinicians and institutions as the decision-makers (8). This introduces bias into decisions that should be objective, and leaves inherent inconsistencies in the process. There is a balance to be struck between the need for patient safety and the need for innovative new biomarkers.

In the next chapter, I will investigate how existing trials have justified inclusion of their genetic biomarkers, and provide recommendations on how future trials can better provide evidence for including the biomarker in their trial design. I will later propose one solution to the problem of how evidence can be collected for an extremely rare biomarker, without conducting a full RCT. Final chapters discuss the views of the general public on the levels of evidence required for biomarker implementation into clinical practice.

2.7 References

1. Hayes DF. Precision Medicine and Testing for Tumor Biomarkers-Are All Tests Born Equal? JAMA Oncol. 2018;4(6):773-4.
2. Ratanawijitrasin S, Wondemagegnehu E. Effective drug regulation: A multicountry study. World Health Organization; 2002.

3. Zimmern RL. Testing challenges: evaluation of novel diagnostics and molecular biomarkers. Clinical medicine (London, England). 2009;9(1):68-73.
4. Food and Drug Administration (US). Context of Use 2018 [Available from: https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/BiomarkerQualificationProgram/ucm535395.htm.
5. Food and Drug Administration (US). Optimizing FDA's Regulatory Oversight of Next Generation Sequencing Diagnostic Tests—Preliminary Discussion Paper 2014 [Available from: https://www.fda.gov/media/90403/download.
6. Janssens AC, Deverka PA. Useless until proven effective: the clinical utility of preemptive pharmacogenetic testing. Clinical pharmacology and therapeutics. 2014;96(6):652-4.
7. University of Maryland's Center of Excellence in Regulatory Science and Innovation, Food and Drug Administration (US), Critical Path Institute (C-Path). Evidentiary Considerations for Integration of Biomarkers in Drug Development 2015 [Available from: https://c-path.org/wp-content/uploads/2015/09/Proceedings-EvidentiaryConsiderationsSymposium-FINAL-20160104.pdf.
8. Moscow JA, Fojo T, Schilsky RL. The evidence framework for precision cancer medicine. Nat Rev Clin Oncol. 2018;15(3):183-92.
9. Wang B, Canestaro WJ, Choudhry NK. Clinical evidence supporting pharmacogenomic biomarker testing provided in US Food and Drug Administration drug labels. JAMA internal medicine. 2014;174(12):1938-44.
10. Amur S, LaVange L, Zineh I, Buckman-Garner S, Woodcock J. Biomarker Qualification: Toward a Multiple Stakeholder Framework for Biomarker Development, Regulatory Acceptance, and Utilization. Clinical pharmacology and therapeutics. 2015;98(1):34-46.
11. Yee LM, Lively TG, McShane LM. Biomarkers in early-phase trials: fundamental issues. Bioanalysis. 2018;10(12):933-44.
12. Lee JW, Devanarayan V, Barrett YC, Weiner R, Allinson J, Fountain S, et al. Fit-for-purpose method development and validation for successful biomarker measurement. Pharmaceutical research. 2006;23(2):312-28.
13. Altar CA, Amakye D, Bounos D, Bloom J, Clack G, Dean R, et al. A prototypical process for creating evidentiary standards for biomarkers and diagnostics. Clinical pharmacology and therapeutics. 2008;83(2):368-71.
14. European Commission. MEDICAL DEVICES: Guidance Document - Classification of Medical Devices 2010 [Available from: https://ec.europa.eu/docsroom/documents/10337/attachments/1/translations.
15. Pennello GA. Analytical and clinical evaluation of biomarkers assays: when are biomarkers ready for prime time? Clin Trials. 2013;10(5):666-76.
16. Dancey JE, Dobbin KK, Groshen S, Jessup JM, Hruszkewycz AH, Koehler M, et al. Guidelines for the development and incorporation of biomarker studies in early clinical trials of novel agents. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010;16(6):1745-55.
17. Le Tourneau C, Delord JP, Goncalves A, Gavoille C, Dubot C, Isambert N, et al. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. Lancet Oncol. 2015;16(13):1324-34.
18. Mosley SA, Hicks JK, Portman DG, Donovan KA, Gopalan P, Schmit J, et al. Design and rational for the precision medicine guided treatment for cancer pain pragmatic clinical trial. Contemp Clin Trials. 2018;68:7-13.
19. Evans BJ, Burke W, Jarvik GP. The FDA and genomic tests--getting regulation right. The New England journal of medicine. 2015;372(23):2258-64.
20. Food and Drug Administration (US). Guidance for Industry Pharmacogenomic Data Submissions 2005 [Available from: https://www.fda.gov/media/122944/download.

21. Klein ME, Parvez MM, Shin JG. Clinical Implementation of Pharmacogenomics for Personalized Precision Medicine: Barriers and Solutions. J Pharm Sci. 2017;106(9):2368-79.
22. Hippman C, Nislow C. Pharmacogenomic Testing: Clinical Evidence and Implementation Challenges. J Pers Med. 2019;9(3).
23. Walley T, Haycox A, Boland A. Pharmacoeconomics: Elsevier Health Sciences; 2004.
24. Williams SA, Slavin DE, Wagner JA, Webster CJ. A cost-effectiveness approach to the qualification and acceptance of biomarkers. Nature reviews Drug discovery. 2006;5(11):897-902.
25. Food and Drug Administration (US). FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource 2016 [updated 2017 Nov 14. Available from: https://www.ncbi.nlm.nih.gov/books/NBK338448/
26. Valdes R, Payne DA, Linder MW. Laboratory analysis and application of pharmacogenetics to clinical practice: American Association for Clinical Chemistry; 2010.
27. European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP). Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection (Draft) 2011 [Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-methodological-issues-associated-pharmacogenomic-biomarkers-relation-clinical_en.pdf.
28. Haddow JE, Palomaki GE. ACCE: A model process for evaluating data on emerging genetic tests. In: Khoury MJ, Little J, Burke W, editors. Human Genome Epidemiology. New York: Oxford University Press; 2004. p. 217-33.
29. Yu C, Bashaw ED. Regulatory perspective of biomarker bioanalysis during drug development. Bioanalysis. 2019;11(7):607-10.
30. Sanderson S, Zimmern R, Kroese M, Higgins J, Patch C, Emery J. How can the evaluation of genetic tests be enhanced? Lessons learned from the ACCE framework and evaluating genetic tests in the United Kingdom. Genet Med. 2005;7(7):495-500.
31. Chau CH, Rixe O, McLeod H, Figg WD. Validation of analytic methods for biomarkers used in drug development. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008;14(19):5967-76.
32. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med. 2009;11(1):3-14.
33. Gillis NK, Innocenti F. Evidence required to demonstrate clinical utility of pharmacogenetic testing: the debate continues. Clinical pharmacology and therapeutics. 2014;96(6):655-7.
34. Jorgensen JT. Site-agnostic biomarker-guided oncology drug development. Expert Rev Mol Diagn. 2020;20(6):583-92.
35. Ioannidis JP. Biomarker failures. Clin Chem. 2013;59(1):202-4.
36. Ransohoff DF. Lessons from controversy: ovarian cancer screening and serum proteomics. Journal of the National Cancer Institute. 2005;97(4):315-9.
37. Diamandis EP. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. Journal of the National Cancer Institute. 2004;96(5):353-6.
38. Evaluation of Genomic Applications in P, Prevention Working G. The EGAPP initiative: lessons learned. Genet Med. 2014;16(3):217-24.
39. Baggerly KA, Coombes KR. What information should be required to support clinical "omics" publications? Clin Chem. 2011;57(5):688-90.

40. Baggerly KA, Coombes KR. Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. The Annals of Applied Statistics. 2009:1309-34.
41. PharmGKB. Citalopram 2020 [Available from: https://www.pharmgkb.org/chemical/PA449015/labelAnnotation.
42. Mrazek DA, Biernacka JM, O'Kane DJ, Black JL, Cunningham JM, Drews MS, et al. CYP2C19 variation and citalopram response. Pharmacogenetics and genomics. 2011;21(1):1-9.
43. Herrlin K, Yasui-Furukori N, Tybring G, Widen J, Gustafsson LL, Bertilsson L. Metabolism of citalopram enantiomers in CYP2C19/CYP2D6 phenotyped panels of healthy Swedes. British journal of clinical pharmacology. 2003;56(4):415-21.
44. Kavanagh J, Oliver S, Lorenc T. Reflections on developing and using PROGRESS-Plus. Equity Update. 2008;2:1-3.
45. Pitini E, De Vito C, Marzuillo C, D'Andrea E, Rosso A, Federici A, et al. How is genetic testing evaluated? A systematic review of the literature. Eur J Hum Genet. 2018;26(5):605-15.
46. Barbarino JM, Whirl-Carrillo M, Altman RB, Klein TE. PharmGKB: A worldwide resource for pharmacogenomic information. Wiley Interdiscip Rev Syst Biol Med. 2018;10(4):e1417.
47. Clinical Pharmacogenetics Implementation Consortium. What is CPIC? 2020 [Available from: https://cpicpgx.org/.
48. Kroese M, Zimmern RL, Sanderson S. Genetic tests and their evaluation: can we answer the key questions? Genet Med. 2004;6(6):475-80.
49. Center for Surveillance EaLSC, Genomics PH. ACCE Model Process for Evaluating Genetic Tests 2010 [Available from: https://www.cdc.gov/genomics/gtesting/ACCE/index.htm.
50. Centers for Disease Control and Prevention (US). ACCE Model List of 44 Targeted Questions 2010 [Available from: https://www.cdc.gov/genomics/gtesting/acce/acce_proj.htm.
51. National Academies of Sciences E, Medicine. An evidence framework for genetic testing: National Academies Press; 2017.
52. Wald N, Cuckle H. Reporting the assessment of screening and diagnostic tests. Br J Obstet Gynaecol. 1989;96(4):389-96.
53. Secretary's Advisory Committee on Genetic Testing. Enhancing the Oversight of Genetic Tests: Recommendations of the SACGT 2000 [Available from: https://osp.od.nih.gov/wp-content/uploads/2013/11/oversight_report.pdf.
54. Kroese M, Zimmern RL, Farndon P, Stewart F, Whittaker J. How can genetic tests be evaluated for clinical use? Experience of the UK Genetic Testing Network. European Journal of Human Genetics. 2007;15(9):917-21.
55. Centers for Disease Control and Prevention (US). Evaluation of Genomic Applications in Practice and Prevention (EGAPP™) 2016 [Available from: https://www.cdc.gov/genomics/gtesting/index.htm#EGAPP.
56. Evaluation of Genomic Applications in P, Prevention Working G. Recommendations from the EGAPP Working Group: does the use of Oncotype DX tumor gene expression profiling to guide treatment decisions improve outcomes in patients with breast cancer? Genet Med. 2016;18(8):770-9.
57. Leptak C, Menetski JP, Wagner JA, Aubrecht J, Brady L, Brumfield M, et al. What evidence do we need for biomarker qualification? Sci Transl Med. 2017;9(417):eaal4599.
58. Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nature reviews Drug discovery. 2015;14(7):475-86.
59. Faulkner E, Holtorf AP, Walton S, Liu CY, Lin H, Biltaj E, et al. Being Precise About Precision Medicine: What Should Value Frameworks Incorporate to Address Precision Medicine? A Report of the Personalized Precision Medicine Special

Interest Group. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2020;23(5):529-39.
60. Lavezzari G, Womack AW. Industry perspectives on biomarker qualification. Clinical pharmacology and therapeutics. 2016;99(2):208-13.
61. Daud ANA, Bergsma EL, Bergman JEH, De Walle HEK, Kerstjens-Frederikse WS, Bijker BJ, et al. Knowledge and attitude regarding pharmacogenetics among formerly pregnant women in the Netherlands and their interest in pharmacogenetic research. BMC Pregnancy Childbirth. 2017;17(1):120.
62. Baruch S, Hudson K. Civilian and military genetics: nondiscrimination policy in a post-GINA world. American journal of human genetics. 2008;83(4):435-44.
63. Harrison M, Milbers K, Hudson M, Bansback N. Do patients and health care providers have discordant preferences about which aspects of treatments matter most? Evidence from a systematic review of discrete choice experiments. BMJ open. 2017;7(5):e014719.
64. Hudson KL, Holohan MK, Collins FS. Keeping pace with the times--the Genetic Information Nondiscrimination Act of 2008. The New England journal of medicine. 2008;358(25):2661-3.
65. Moldrup C. Ethical, social and legal implications of pharmacogenomics: a critical review. Community Genet. 2001;4(4):204-14.
66. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clinical pharmacology and therapeutics. 2012;92(4):414-7.
67. PharmGKB. About Us 2020 [Available from: https://www.pharmgkb.org/about.
68. PharmGKB. What is PharmGKB? 2020 [Available from: https://www.pharmgkb.org/whatIsPharmgkb.
69. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clinical pharmacology and therapeutics. 2012;92(4):414-7.
70. Tilleman L, Weymaere J, Heindryckx B, Deforce D, Nieuwerburgh FV. Contemporary pharmacogenetic assays in view of the PharmGKB database. Pharmacogenomics. 2019;20(4):261-72.
71. Lever J, Barbarino JM, Gong L, Huddart R, Sangkuhl K, Whaley R, et al., editors. PGxMine: Text mining for curation of PharmGKB. Pacific Symposium on Biocomputing; 2020.
72. PharmGKB. Scoring of PharmGKB Variant Annotations 2021 [Available from: https://www.pharmgkb.org/page/varAnnScoring.
73. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. Clinical pharmacology and therapeutics. 2011;89(3):464-7.
74. Caudle KE, Klein TE, Hoffman JM, Muller DJ, Whirl-Carrillo M, Gong L, et al. Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process. Current drug metabolism. 2014;15(2):209-17.
75. Clinical Pharmacogenetics Implementation Consortium. Prioritization of CPIC Guidelines 2021 [Available from: https://cpicpgx.org/prioritization-of-cpic-guidelines/.
76. Clinical Pharmacogenetics Implementation Consortium. Guidelines 2019 [Available from: https://cpicpgx.org/guidelines/.
77. Clinical Pharmacogenetics Implementation Consortium. Levels of Evidence 2016 [Available from: https://cpicpgx.org/levels-of-evidence/.
78. Crews KR, Monte AA, Huddart R, Caudle KE, Kharasch ED, Gaedigk A, et al. Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2D6, OPRM1, and COMT Genotypes and Select Opioid Therapy. Clinical pharmacology and therapeutics. 2021;110(4):888-96.

79. Phillips EJ, Sukasem C, Whirl-Carrillo M, Muller DJ, Dunnenberger HM, Chantratita W, et al. Clinical Pharmacogenetics Implementation Consortium Guideline for HLA Genotype and Use of Carbamazepine and Oxcarbazepine: 2017 Update. Clinical pharmacology and therapeutics. 2018;103(4):574-81.
80. PharmGKB. C11orf65 2021 [Available from: https://www.pharmgkb.org/gene/PA144596484/clinicalAnnotation/981237900.
81. (US) FaDA. Personalized medicine: A biological approach to patient treatment, A conversation with Issam Zineh, Director of CDER's Office of Clinical Pharmacology 2016 [Available from: https://www.fda.gov/Drugs/NewsEvents/ucm325475.htm.
82. Jorgensen JT. Predictive biomarkers and clinical evidence. Basic Clin Pharmacol Toxicol. 2021;128(5):642-8.
83. Medicines and Healthcare products Regulatory Agency (UK). About Us 2019 [Available from: https://www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency/about.
84. Medicines and Healthcare products Regulatory Agency (UK). Medicines & Medical Devices Regulation: What you need to know 2012 [Available from: http://www.mhra.gov.uk/home/groups/comms-ic/documents/websiteresources/con2031677.pdf.
85. Richards N, Hudson I. UK medicines regulation: responding to current challenges. British journal of clinical pharmacology. 2016;82(6):1471-6.
86. Dati F. The New European Directive on in vitro Diagnostics. Clinical Chemistry and Laboratory Medicine2003. p. 1289.
87. Medicines and Healthcare products Regulatory Agency (UK). Medical Devices: how to comply with the legal requirements 2019 [Available from: https://www.gov.uk/guidance/medical-devices-how-to-comply-with-the-legal-requirements.
88. Medicines and Healthcare products Regulatory Agency (UK). Guidance on the In Vitro Diagnostic Medical Devices Directive 98/79/EC 2013 [Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/404335/In_vitro_diagnostic_medical_devices_-_guidance_on_legislation.pdf.
89. European Parliament CotEU. Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. In: DG03, editor. 1998. p. P. 0001 - 37.
90. Public Health England. Notifiable diseases and causative organisms: how to report 2020 [Available from: https://www.gov.uk/guidance/notifiable-diseases-and-causative-organisms-how-to-report.
91. Medicines and Healthcare products Regulatory Agency (UK). Notified bodies for medical devices 2017 [Available from: https://www.gov.uk/government/publications/notified-bodies-for-medical-devices/notified-bodies-for-medical-devices.
92. Medicines and Healthcare products Regulatory Agency. An introductory guide to the medical device regulation (MDR) and the in vitro diagnostic medical device regulation (IVDR) 2020 [Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/640404/MDR_IVDR_guidance_Print_13.pdf.
93. European Union. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. Official Journal of the European Union. 2017.
94. House of Commons Library. Brexit and medicines regulation 2019 [Available from: https://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-8148.
95. Medicines and Healthcare products Regulatory Agency (UK). MHRA guidance and publications on a possible no deal scenario 2019 [Available from:

https://www.gov.uk/government/collections/mhra-guidance-and-publications-on-a-possible-no-deal-scenario.
96. Medicines and Healthcare products Regulatory Agency (UK). Regulating medical devices in the event of a no deal scenario 2019 [Available from: https://www.gov.uk/guidance/regulating-medical-devices-in-the-event-of-a-no-deal-scenario.
97. The Medical Devices Regulations 2002, Stat. SI 2002/618 (2002).
98. Parliament UK. Medicines and Medical Devices Bill 2019-20 2020 [Available from: https://services.parliament.uk/bills/2019-21/medicinesandmedicaldevices.html.
99. European Medicines Agency. ICH E15 Definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories 2007 [Available from: https://www.ema.europa.eu/en/ich-e15-definitions-genomic-biomarkers-pharmacogenomics-pharmacogenetics-genomic-data-sample-coding.
100. European Medicines Agency. ICH E16 Genomic biomarkers related to drug response: context, structure and format of qualification submissions 2013 [Available from: https://www.ema.europa.eu/en/ich-e16-genomic-biomarkers-related-drug-response-context-structure-format-qualification-submissions.
101. European Medicines Agency. ICH guidelines 2020 [Available from: https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/ich-guidelines.
102. European Commission. Questions and Answers on the progressive roll-out of the new In Vitro Diagnostic Medical Devices Regulation 2021 [Available from: https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_5210.
103. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ, et al. What is "quality of evidence" and why is it important to clinicians? BMJ. 2008;336(7651):995-8.
104. NHS Health Research Authority. UK policy framework for health and social care research 2017 [Available from: https://www.hra.nhs.uk/media/documents/Final_Accessibility_uk-policy-framework-health-social-care-research_.pdf.
105. NHS Health Research Authority. Systematic review 2018 [Available from: https://www.hra.nhs.uk/planning-and-improving-research/research-planning/systematic-review/.
106. Food and Drug Administration (US). What does FDA regulate? 2018 [Available from: https://www.fda.gov/AboutFDA/Transparency/Basics/ucm194879.htm.
107. Food and Drug Administration (US). FDA Organization 2019 [Available from: https://www.fda.gov/aboutfda/centersoffices/default.htm
108. The Association of the British Pharmaceutical Industry. Top 10 pharmaceutical markets by value (USD) 2016 [Available from: https://www.abpi.org.uk/facts-and-figures/global-pharmaceutical-market/top-10-pharmaceutical-markets-by-value-usd/.
109. US Department of Health and Human Services. Innovation or stagnation? Challenge and opportunity on the critical path to new medical products 2004 [Available from: http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm077262.htm

http://wayback.archive-it.org/7993/20180125035500/https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/UCM113411.pdf
110. Harpur E, Walker EG. Qualification of safety biomarkers for use in drug development: What has been achieved and what is the path forward? Current Opinion in Toxicology. 2017;4:66-73.

111. Parekh A, Buckman-Garner S, McCune S, R ON, Geanacopoulos M, Amur S, et al. Catalyzing the Critical Path Initiative: FDA's progress in drug development activities. Clinical pharmacology and therapeutics. 2015;97(3):221-33.
112. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clinical pharmacology and therapeutics. 2001;69(3):89-95.
113. Brady L. Biomarker Development and Validation: Establishing Standards of Evidence for their Context of Use in Clinical Trials 2014 [Available from: https://www.nimh.nih.gov/funding/grant-writing-and-application-process/concept-clearances/2014/biomarker-development-and-validation-establishing-standards-of-evidence-for-their-context-of-use-in-clinical-trials.shtml.
114. Amur SG, Sanyal S, Chakravarty AG, Noone MH, Kaiser J, McCune S, et al. Building a roadmap to biomarker qualification: challenges and opportunities. Biomark Med. 2015;9(11):1095-105.
115. Woodcock J, Buckman S, Goodsaid F, Walton MK, Zineh I. Qualifying biomarkers for use in drug development: a US Food and Drug Administration overview. Expert Opin Med Diagn. 2011;5(5):369-74.
116. Salerno RA, Lesko LJ. Pharmacogenomic data: FDA voluntary and required submission guidance. Pharmacogenomics. 2004;5(5):503-5.
117. Food and Drug Administration (US). Biomarker Qualification—collaborative effort to Qualify a Drug Development tool 2017 [Available from: https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/BiomarkerQualificationProgram/UCM552881.pdf.
118. Sosnay PR, Siklosi KR, Van Goor F, Kaniecki K, Yu H, Sharma N, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. Nat Genet. 2013;45(10):1160-7.
119. Food and Drug Administration (US). CDER Biomarker Qualification Program 2018 [Available from: https://www.fda.gov/drugs/drug-development-tool-qualification-programs/cder-biomarker-qualification-program.
120. Beaver JA, Tzou A, Blumenthal GM, McKee AE, Kim G, Pazdur R, et al. An FDA Perspective on the Regulatory Implications of Complex Signatures to Predict Response to Targeted Therapies. Clinical cancer research : an official journal of the American Association for Cancer Research. 2017;23(6):1368-72.
121. Food and Drug Administration (US). List of Qualified Biomarkers 2018 [Available from: https://www.fda.gov/drugs/cder-biomarker-qualification-program/list-qualified-biomarkers.
122. Mittra A, Moscow JA. Future Approaches to Precision Oncology-Based Clinical Trials. Cancer J. 2019;25(4):300-4.
123. Food and Drug Administration (US). Pembrolizumab (KEYTRUDA) 2016 [Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/pembrolizumab-keytruda.
124. Food and Drug Administration (US). FDA approves larotrectinib for solid tumors with NTRK gene fusions 2018 [Available from: https://www.fda.gov/drugs/fda-approves-larotrectinib-solid-tumors-ntrk-gene-fusions-0.
125. Sul J, Blumenthal GM, Jiang X, He K, Keegan P, Pazdur R. FDA Approval Summary: Pembrolizumab for the Treatment of Patients With Metastatic Non-Small Cell Lung Cancer Whose Tumors Express Programmed Death-Ligand 1. The oncologist. 2016;21(5):643-50.
126. National Institutes of Health, US National Library of Medicine. Study of Pembrolizumab (MK-3475) in Participants With Progressive Locally Advanced or Metastatic Carcinoma, Melanoma, or Non-small Cell Lung Carcinoma (P07990/MK-3475-001/KEYNOTE-001) (KEYNOTE-001). NCT01295827 2019 [Available from: https://clinicaltrials.gov/ct2/show/NCT01295827.

127. Leighl NB, Hellmann MD, Hui R, Carcereny E, Felip E, Ahn MJ, et al. Pembrolizumab in patients with advanced non-small-cell lung cancer (KEYNOTE-001): 3-year results from an open-label, phase 1 study. The Lancet Respiratory medicine. 2019;7(4):347-57.
128. Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. The New England journal of medicine. 2015;372(21):2018-28.
129. Food and Drug Administration (US). Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics: Guidance for Industry 2018 [Available from: https://www.fda.gov/media/71195/download.
130. Lemery S, Keegan P, Pazdur R. First FDA Approval Agnostic of Cancer Site - When a Biomarker Defines the Indication. The New England journal of medicine. 2017;377(15):1409-12.
131. Darrow JJ, Avorn J, Kesselheim AS. FDA Approval and Regulation of Pharmaceuticals, 1983-2018. JAMA. 2020;323(2):164-76.
132. Food and Drug Administration (US). Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products. Guidance for Industry 2019 [Available from: https://www.fda.gov/media/133660/download.
133. Chin L, Devine B, Baradaran S, Keyloun K, Canestaro W, Pham J. Characterizing the Strength of Evidence in FDA Labels for Pharmacogenomic Biomarker-Guided Medication Use. AMIA Summits on Translational Science Proceedings. 2017;2017:30-9.
134. Kim J, Hu C, Moufawad El Achkar C, Black LE, Douville J, Larson A, et al. Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. The New England journal of medicine. 2019;381(17):1644-52.
135. Woodcock J, Marks P. Drug Regulation in the Era of Individualized Therapies. The New England journal of medicine. 2019;381(17):1678-80.
136. Antoniou M, Jorgensen AL, Kolamunnage-Dona R. Biomarker-guided trial designs (BiGTeD): An online tool to help develop personalised medicine [Available from: http://www.bigted.org/.
137. Freidlin B, Korn EL. Biomarker enrichment strategies: matching trial design to biomarker credentials. Nat Rev Clin Oncol. 2014;11(2):81-90.
138. Mandrekar SJ, An MW, Sargent DJ. A review of phase II trial designs for initial marker validation. Contemp Clin Trials. 2013;36(2):597-604.
139. Mandrekar SJ, Sargent DJ. Design of clinical trials for biomarker research in oncology. Clin Investig (Lond). 2011;1(12):1629-36.
140. Freidlin B, McShane LM, Polley MY, Korn EL. Randomized phase II trial designs with biomarkers. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012;30(26):3304-9.
141. Antoniou M, Jorgensen AL, Kolamunnage-Dona R. Biomarker-Guided Adaptive Trial Designs in Phase II and Phase III: A Methodological Review. PLoS One. 2016;11(2):e0149803.
142. Antoniou M, Kolamunnage-Dona R, Jorgensen AL. Biomarker-Guided Non-Adaptive Trial Designs in Phase II and Phase III: A Methodological Review. J Pers Med. 2017;7(1):1.
143. Wason JMS, Brocklehurst P, Yap C. When to keep it simple – adaptive designs are not always useful. BMC Medicine. 2019;17(1):152.
144. Legocki LJ, Meurer WJ, Frederiksen S, Lewis RJ, Durkalski VL, Berry DA, et al. Clinical trialist perspectives on the ethics of adaptive clinical trials: a mixed-methods analysis. BMC Medical Ethics. 2015;16(1):27.
145. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Medicine. 2018;16(1):29.
146. Carthon BC, Antonarakis ES. The STAMPEDE trial: paradigm-changing data through innovative trial design. Transl Cancer Res. 2016;5(3 Suppl):S485-S90.

147. Gilson C, Chowdhury S, Parmar MKB, Sydes MR, Investigators S. Incorporating Biomarker Stratification into STAMPEDE: an Adaptive Multi-arm, Multi-stage Trial Platform. Clin Oncol (R Coll Radiol). 2017;29(12):778-86.
148. Vineis P, Schulte P, McMichael AJ. Misconceptions about the use of genetic tests in populations. Lancet. 2001;357(9257):709-12.
149. Shirts BH, Jacobson A, Jarvik GP, Browning BL. Large numbers of individuals are required to classify and define risk for rare variants in known cancer risk genes. Genet Med. 2014;16(7):529-34.
150. Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. Statistics in medicine. 2014;33(24):4186-201.
151. Huddart R, Sangkuhl K, Whirl-Carrillo M, Klein TE. Are Randomized Controlled Trials Necessary to Establish the Value of Implementing Pharmacogenomics in the Clinic? Clinical pharmacology and therapeutics. 2019;106(2):284-6.
152. Eng KH. Randomized reverse marker strategy design for prospective biomarker validation. Statistics in medicine. 2014;33(18):3089-99.
153. Antoniou M, Kolamunnage-Dona R, Wason J, Bathia R, Billingham C, Bliss JM, et al. Biomarker-guided trials: Challenges in practice. Contemp Clin Trials Commun. 2019;16:100493.
154. Barton S. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. BMJ. 2000;321(7256):255-6.
155. Hyde P. Fool's Gold: Examining the Use of Gold Standards in the Production of Research Evidence. British Journal of Occupational Therapy. 2016;67(2):89-94.
156. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the Gold Standard--Lessons from the History of RCTs. The New England journal of medicine. 2016;374(22):2175-81.
157. Heads of Medicines Agencies, Clinical Trials Facilitation and Coordination Group. Recommendation Paper on the Initiation and Conduct of Complex Clinical Trials 2019 [Available from: https://www.hma.eu/fileadmin/dateien/Human_Medicines/01-About_HMA/Working_Groups/CTFG/2019_02_CTFG_Recommendation_paper_on_Complex_Clinical_Trials.pdf.
158. Heads of Medicines Agencies. About HMA 2020 [Available from: https://www.hma.eu/abouthma.html.
159. Blagden SP, Billingham L, Brown LC, Buckland SW, Cooper AM, Ellis S, et al. Effective delivery of Complex Innovative Design (CID) cancer trials-A consensus statement. British journal of cancer. 2020;122(4):473-82.
160. Malik L, Lu D. Increasing complexity in oncology phase I clinical trials. Invest New Drugs. 2019;37(3):519-23.
161. Lam YW. Scientific challenges and implementation barriers to translation of pharmacogenomics in clinical practice. ISRN Pharmacol. 2013;2013:641089.
162. Arwood MJ, Chumnumwat S, Cavallari LH, Nutescu EA, Duarte JD. Implementing Pharmacogenomics at Your Institution: Establishment and Overcoming Implementation Challenges. Clin Transl Sci. 2016;9(5):233-45.
163. Limaye N. Pharmacogenomics, Theranostics and Personalized Medicine - the complexities of clinical trials: challenges in the developing world. Appl Transl Genom. 2013;2:17-21.
164. Turner RM, Newman WG, Bramon E, McNamee CJ, Wong WL, Misbah S, et al. Pharmacogenomics in the UK National Health Service: opportunities and challenges. Pharmacogenomics. 2020;21(17):1237-46.
165. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(24):4027-34.

166. Jorgensen AL, Williamson PR. Methodological quality of pharmacogenetic studies: issues of concern. Statistics in medicine. 2008;27(30):6547-69.
167. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. The New England journal of medicine. 2000;342(25):1887-92.
168. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. PLoS Med. 2011;8(5):e1001026.
169. Gardner HR, Treweek S, Gillies K. Using evidence when planning for trial recruitment: An international perspective from time-poor trialists. PLoS One. 2019;14(12):e0226081.
170. Luo J, Wu M, Chen W. Geographical Distribution and Trends of Clinical Trial Recruitment Sites in Developing and Developed Countries. Journal of Health Informatics in Developing Countries. 2017;11(1).
171. Hatz MH, Schremser K, Rogowski WH. Is individualized medicine more cost-effective? A systematic review. Pharmacoeconomics. 2014;32(5):443-55.
172. Khan SS, Chen AP, Takebe N. Impact of NCI-MATCH: a Nationwide Oncology Precision Medicine Trial. Expert Review of Precision Medicine and Drug Development. 2019;4(4):251-8.
173. Food and Drug Administration (US). Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling: Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders 2016 [Available from: https://www.fda.gov/media/92593/download.
174. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2011;14(4):403-13.
175. Bridges JFP, Kinter ET, Kidane L, Heinzen RR, McCormick C. Things are Looking up Since We Started Listening to Patients. The Patient: Patient-Centered Outcomes Research. 2008;1(4):273-82.
176. Carpenter D. Reputation and power: Organizational image and pharmaceutical regulation at the FDA: Princeton University Press; 2014.

Chapter 3: Evidence to support the inclusion of pharmacogenetic biomarkers in randomised controlled trials

Parts of this chapter have been published in:

Johnson D, Hughes D, Pirmohamed M, Jorgensen A. Evidence to Support Inclusion of Pharmacogenetic Biomarkers in Randomised Controlled Trials. J Pers Med. 2019;9(3):42. Published 2019 Sep 1. doi:10.3390/jpm9030042

3.1 Evidence base in pharmacogenetics

As discussed, pharmacogenetics has the potential to impact healthcare in improving drug efficacy, reducing drug side-effects, and improving drug cost-effectiveness. However, the full potential of pharmacogenetics is not currently being exploited (1). One study of prescriptions in a US health system over 3 years found that of 8718 medication orders with recommended or required testing, only 129 pharmacogenetic tests were performed (1.5%) (2).

As far back as 2003, a White Paper by the UK Department of Health laid out the potential for pharmacogenetics to improve healthcare in the National Health Service (NHS) (3): preventing disease, preventing adverse drug reactions; and as predictors of drug response. This paper predicted that new pharmacogenetic products would be common in the NHS within 5 years. Almost 20 years later, the implementation of pharmacogenetics has not lived up to expectations.

There are many reasons for this, including the regulatory hurdles discussed in Chapter 2 (4-11). A review of 229 published papers also found that issues with information technology and scientific barriers were the most common obstacles hindering the wider implementation of pharmacogenetics (5). These technology obstacles included: alert fatigue by clinicians receiving frequent pharmacogenetic information on prescribing, lack of infrastructure for decision support, and incompatibility with existing electronic health records. Scientific barriers to pharmacogenetics included: long turnaround time of tests leading to treatment delays, cost, and a lack of randomised controlled trials (RCTs) demonstrating the efficacy and utility of pharmacogenetic testing (Figure 3.1) (5). Other barriers identified were education (of both clinicians and the general public),

ethical/legal/social/regulation issues, and reimbursement (in terms of national health system payers or health insurance) (5).

There are also significant geographical disparities in implementation. A 2017 review of pharmacogenetics implementation found that pharmacogenetics projects were mostly based in North America and Europe, but highlighted significant ongoing work in Australia, Japan, and South Korea (5). Implementation is also complex in low- and middle-income countries. Challenges such as under-resourced health systems, a relative lack of research infrastructure, and socio-cultural barriers contribute to these disparities (12, 13).

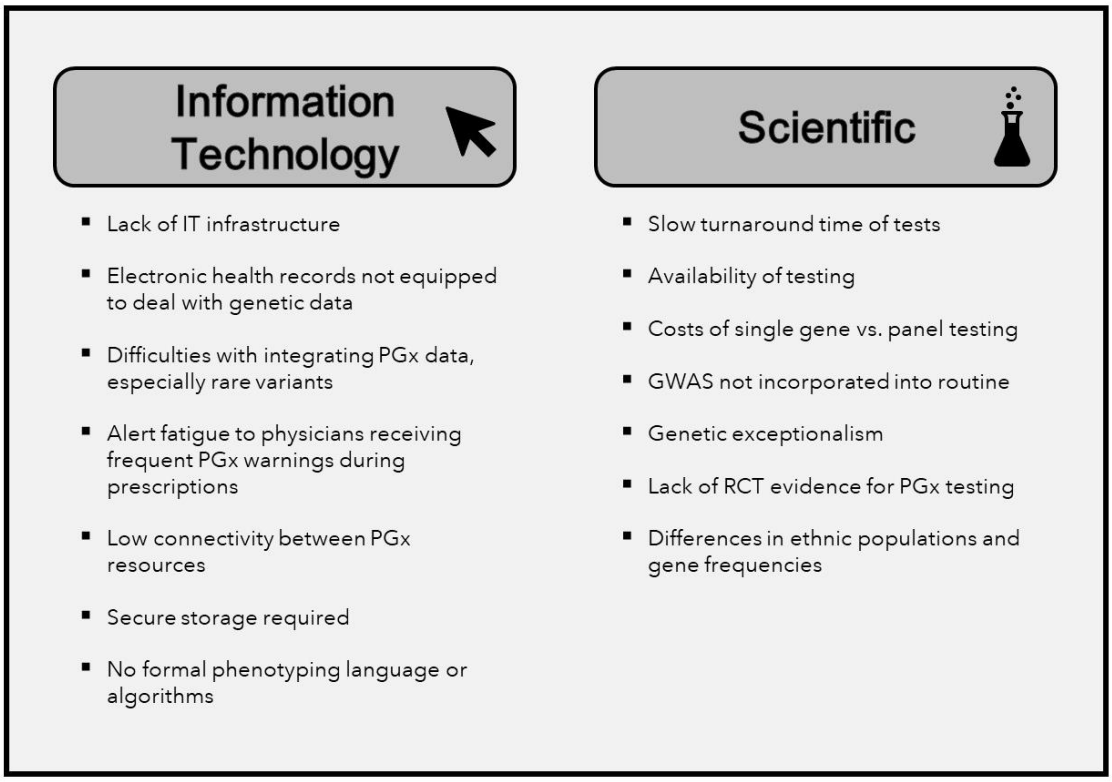


Figure 3.1 -Information Technology and Scientific barriers to the wider implementation of pharmacogenetics (5)

Over a decade ago, McKinnon, *et al.* (2007) wrote about the barriers to pharmacogenetics implementation (14). One of these was cost - a barrier that has been significantly lowered by the falling cost of sequencing since then (1, 15). However, a more recent publication by Hippman & Nislow (2019) divides barriers into two categories: whether testing should be performed at all, and challenges to

pharmacogenetic integration into clinical systems (1). Lack of evidence for pharmacogenetic testing forms the majority of the first category. If there is insufficient good quality evidence for a pharmacogenetic test, there may be negative utility associated with its use – e.g. wasting time and resources. A lack of robust evidence of clinical utility is the largest impediment to the wider implementation of pharmacogenetics (5, 8, 16-18), since a lack of well-designed trials limits the ability of regulators and payers to evaluate the evidence for a biomarker's efficacy (9, 17-20).

McKinnon, *et al.* (2007) also called for large, randomised controlled trials of pharmacogenetic approaches, in multiple ethnic groups (14) which is essential for the clinical implementation of pharmacogenetics. However, before any trial can take place, there needs to be sufficient evidence that there is an association between the pharmacogenetic biomarker and clinical outcome. This is to ensure time and money invested into a trial are not wasted, but more importantly, participants are not put at undue risk.

The evidence that trials do cite can take many forms. For biomarker-guided trials, specific evidence of the biomarker's clinical utility and analytic validity is required (9, 21-24). Clinical utility (the risks and benefits associated with a biomarker's introduction to practice) (25), and analytic validity (the ability of the biomarker to accurately assess genotypes)(25, 26) are distinct concepts that require different evidence. While analytic validity can be shown in *in vitro* or observational studies, clinical utility often requires an RCT (27).

Despite the importance of this evidence gathering prior to commencing a biomarker-guided trial, the nature and extent of evidence required, and how it should be compiled, is unclear. Although a biomarker assay is an integral component of many trials, there exists more guidance on the evidence required for the inclusion of the actual intervention than the biomarkers themselves (24, 28). Trialists have therefore been left to justify biomarker inclusion in their own ways.

Since there appears to be no specific guidance on how evidence of a biomarker's validity should be compiled before proceeding to a clinical trial, I conducted a review with the aim of identifying how existing RCTs have justified inclusion of biomarkers within their trials. I chose 5 different trials that each represented a different area of biomarker use – prevention of adverse drug reactions (ADRs), improving drug efficacy, choosing targeted therapies, improving medication adherence, and improving patient's health-related quality of life. I explored the extent and nature of

the evidence used to justify biomarker inclusion, and reflected on how evidence could be compiled by those planning biomarker-guided trials in the future. Parts of this work were published in the *Journal of Personalised Medicine* in 2019 (10).

3.2 Methods

For this review, I identified five trials that explored different genetic biomarker applications – prevention of ADRs (29), improving efficacy (30), choosing targeted therapies (31), improving medication adherence (32, 33), and improving health-related quality of life (34).

The first trial (TPMT: AZA Response to Genotyping and Enzyme Testing, TARGET, 2011) explored whether *TPMT* genotyping helped prevent ADRs associated with azathioprine (29, 35). A second trial (European Pharmacogenetics of Anticoagulant Therapy, EU-PACT, 2013) tested whether a genotype-guided approach to calculating therapeutic dose of the anticoagulant, warfarin, led to improved efficacy and reduced the incidence of ADRs (30). The third trial (SHIVA, 2015) explored the utility of an approach that used genotyping to match patients to molecularly targeted therapies (31). A fourth trial (Genotype-guided statin therapy, GIST statin trial, 2018) explored whether using genotype testing improved medication adherence and subsequently statin efficacy (32, 33, 36). The final trial (NCT02664350) investigated the use of genotyping to reduce pain associated with cancer (34) (Appendix Table 2). Since I wanted to evaluate what evidence for each biomarker's validity was available when the trial was planned or commenced, I used published protocols or design papers where available. Where trials did not have this information, I contacted trial authors or personnel to obtain protocols.

This review focussed on RCTs as they are the 'gold-standard' of evidence (37) and are likely to be considered highly by regulators compiling evidence for a biomarker's approval (38, 39). The review does not consider trial results, and whilst they are reported here for completeness, trials were not specifically chosen to favour a pharmacogenetics approach to treatment.

For each trial, each piece of evidence referenced in the introduction section of the published protocol or design paper that justified including the biomarker in the trial was identified. Where meta-analyses were cited, I evaluated whether these meta-analyses assessed studies for quality before inclusion. This was used as a proxy of quality of the meta-analyses themselves.

For each trial, details of the publication year, study design, drug of interest, biomarker used, sample size, country of origin, and the age, gender, and ethnicity of participants were also extracted. For TARGET protocols were located by contacting the authors. For SHIVA, the protocol was contained in the supplementary information of the results paper. Figures were made using RStudio (version 1.1.453, RStudio Team, Boston MA) (40), particularly the 'formattable' package (41).

3.3 Results

The timings of the evidence used by each trial were evaluated. Trials published evidence from varying time periods, but all cited evidence from within at least 3 years of their publication or protocol date (Figure 3.2). The full lists of evidence for each trial are included in Appendix Tables 3-7.

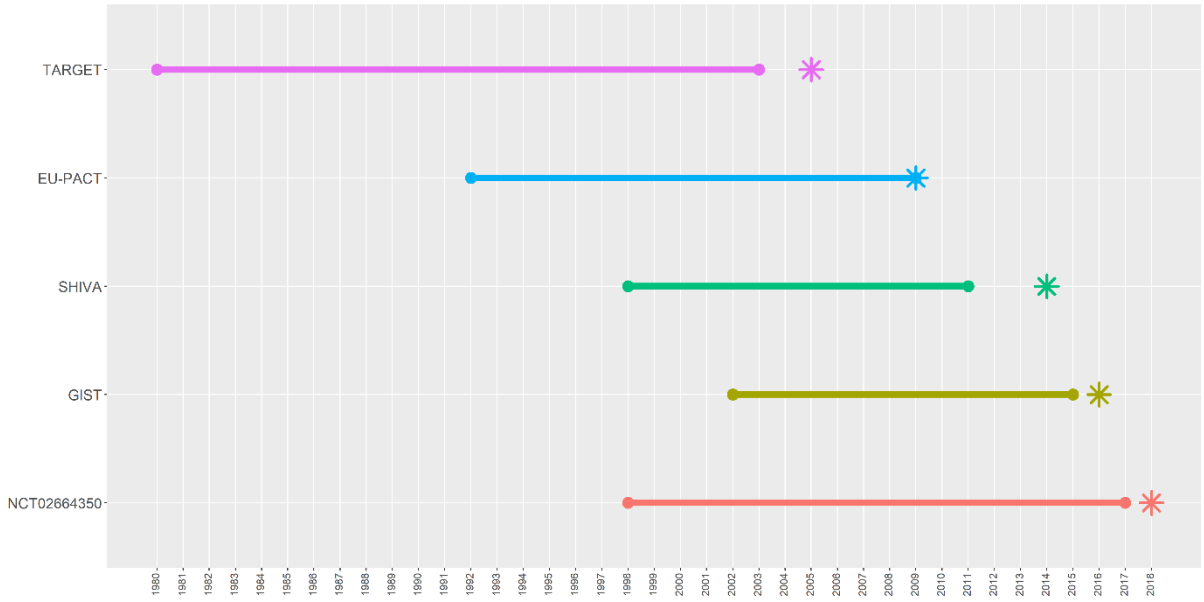


Figure 3.2 - Timings of publications cited by each trial in this review. Stars indicate the year of publication of the paper or protocol references were extracted from. Note that results from NCT02664350 are not yet published.

3.3.1 TARGET

3.3.1.1 Background

The TARGET trial (ISRCTN30748308) used *TPMT* genotyping to guide azathioprine treatment of patients with inflammatory disease (29, 42). Azathioprine is a thiopurine immunosuppressant medication that can cause profound neutropenia as a side-

effect of treatment (43). Patients that have *TPMT*2, TPMT*3A,* and *TPMT*3C* alleles are more likely to suffer neutropenia due to a deficiency of the TPMT enzyme (44). The *TPMT* gene is now recognised as a 'very important pharmacogene' by the Pharmacogenomics Knowledgebase (PharmGKB) (45).

While measurement of enzyme activity is a regular part of clinical practice, it has several limitations. Measurement can be affected by recent blood transfusions, and several common drugs can affect the result (such as aspirin and sulphasalazine) (46). Genotyping the *TPMT* gene instead removes these limitations.

Recruitment to TARGET began in 2005, and a protocol was published contemporaneously (35). This protocol was used to evaluate the evidence available at the time of the trial.

3.3.1.2 Methods

TARGET used a biomarker strategy design without biomarker assessment in the control arm (47). A total of 333 participants recruited from rheumatology and gastroenterology were randomised 1:1 to genotyping or non-genotyping arms (35). In the genotyping arm, the patient's *TPMT* status was revealed to their clinicians. This was accompanied by information on how this status would affect azathioprine dosing. In the non-genotyping arm, participants' clinicians did not receive this information and dosing was calculated according to standard procedures.

3.3.1.3 Results

Patients in the genotyping arm received significantly lower starting azathioprine doses than those in the non-genotyping arm. However, there was no significant difference between arms in rates of stopping treatment due to ADRs. A later cost-effectiveness analysis found that genotyping was associated with a cost saving, but also a slight negative effect on health status (42).

3.3.1.4 Evidence used to justify biomarker

The evidence used to justify the use of the *TPMT* biomarker spanned from 1980 to 2003 (Figure 3.3). The oldest evidence cited was a 1980 observational cohort study that first proposed a monogenic inheritance pattern for the TPMT enzyme (48). A 1989 case-control study comparing TPMT enzyme activity in patients with ADRs from thiopurines to a control group was also cited (49).

Eleven observational studies were cited, consisting of 9 cohort studies (44, 48, 50-56), 1 case control study (49), and 1 study of enzymatic assay use in the UK (57).

The cohort studies show the progression from enzymatic testing to genetic testing. A paper from 1994 compared TPMT enzyme activity across black and white populations, finding lower median activity in black subjects (51). The same group performed a similar study in 1999, using genetic testing to determine *TPMT* allele frequencies in children with leukaemia (53). From this point, most of the observational studies focussed on genotype over enzymatic studies.

One systematic review was cited, but this was a general pharmacogenetics review, not specific to azathioprine or *TPMT* (58). This was cited to underline the utility of a pharmacogenetics approach.

Figure 3.3 - Evidence cited by the TARGET trial to justify inclusion of the TPMT biomarker. The numbers at the top represent years relative to the publication of the protocol in 2005 (35).

The most recent citation was a 2003 expert opinion from a paediatric gastroenterologist on the use of *TPMT* to monitor azathiopurine levels in patients with inflammatory bowel disease (59). Other evidence included: a 2002 cost-effectiveness analysis of *TPMT* genotyping (60), a 1997 questionnaire of UK clinicians on azathioprine usage (61), and a case study of a patient with azathioprine ADR and *TPMT* mutant alleles, from 2000 (62). A guideline by the British Society of Rheumatology from 2000 was also cited, but could not be located online for evaluation.

3.3.1.5 Discussion

Overall, the TARGET trial cited a variety of evidence types to justify inclusion of the biomarker within its trial, from a wide time range. The citations spanned the longest time frame of all the trials included here. For this trial, I investigated the evidence cited in a trial protocol provided by the authors. It is important to note that whilst the trial protocol was investigated for evidence cited, due to character and reference limits from the funders or sponsors, the authors may have been unable to include all relevant references used to justify biomarker inclusion.

3.3.2 EU-PACT

3.3.2.1 Background

EU-PACT (NCT01119300) was a single-blind, randomised European trial of genotype-guided warfarin dosing (30, 63-66). Warfarin is an antiplatelet drug used for the treatment of: rheumatic heart disease, atrial fibrillation, patients with prosthetic heart valves, venous thrombosis, pulmonary embolism, and transient ischaemic attacks (67). It is widely used in the UK, particularly in older populations (68). Warfarin works by inhibiting the synthesis of clotting factors, reducing the risk of blood clots (69). However, warfarin patients require regular monitoring due to the drug's narrow therapeutic window (69, 70). Dosing of warfarin to maintain this therapeutic range is complex and depends on clinical algorithms that incorporate interacting medications, food, and alcohol, among others (71-74). Too high a dose increases the risk of bleeding, while too low a dose increases the risk of thromboembolic events (68, 70) (Figure 3.4). Warfarin is a common causative agent in hospital admissions due to ADRs (75).

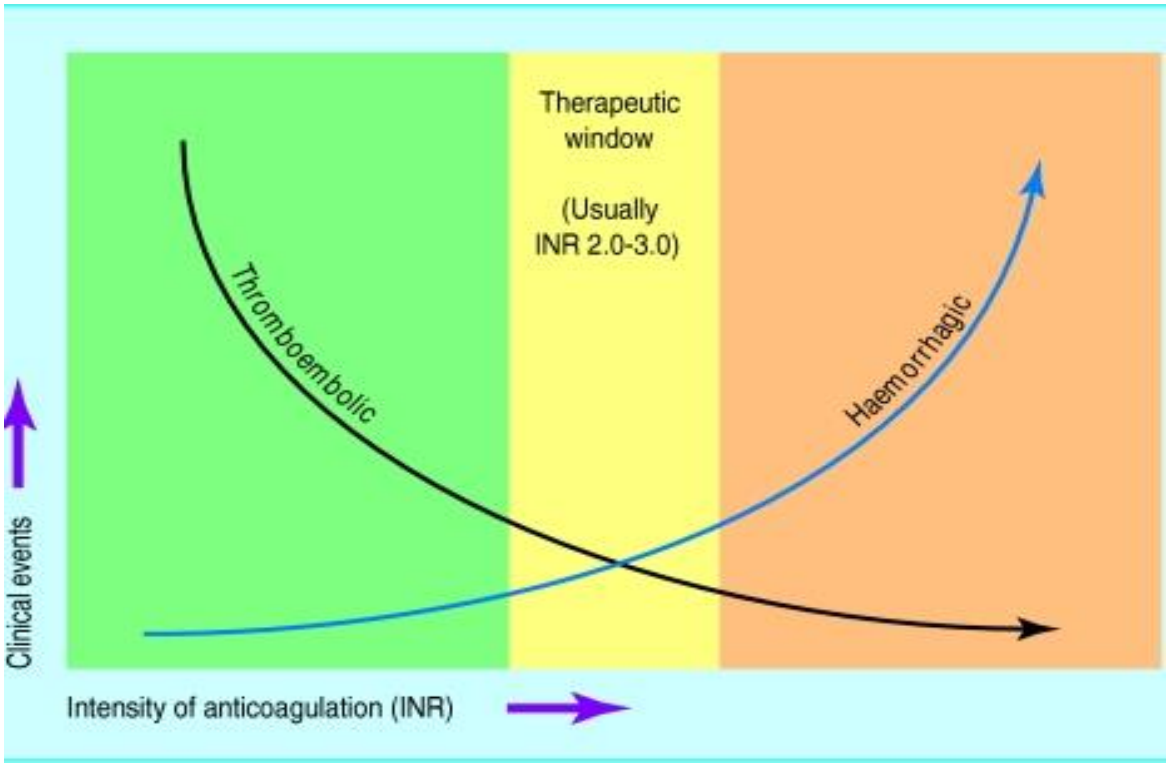


Figure 3.4 - warfarin dosing and the importance of the therapeutic window (67, 68, 71). Adapted from Blann, et al. *(2003) (70). INR = International Normalised Ratio, a measure of intensity of anticoagulation.*

More recently, there has been research into genetic factors that can affect warfarin dosing (76). Over 30 genes have been identified as being linked to the warfarin

mode of action, but *CYP2C9* and *VKORC1* have been acknowledged as the most important (74, 76-78). Both *CYP2C9* and *VKORC1* are now included in PharmGKB's list of very important pharmacogenes (79, 80).

Patients with *CYP2C9*2* and *CYP2C9*3* alleles have reduced metabolism of warfarin, requiring a lower daily dose (76, 81, 82). Those with variants in *VKORC1* also have different warfarin requirements (76, 82). These genes, combined with age and height, account for 55% of the variance in warfarin dosage requirements (68, 83).

The EU-PACT trial was a 2013 RCT that compared genotype-guided dosing to conventional dosing. A protocol was published in 2009 and recruitment began in 2011 (30, 66).

3.3.2.2 Methods

Participants in the UK and Sweden were randomised 1:1 to genotype-guided or conventional dosing (control) groups, stratified by centre and treatment indication. Those in the genotype-guided group were genotyped for *CYP2C9* and *VKORC1* and dosed according to an algorithm including both genetic and clinical factors. The control group received a standard dosing regimen. All participants had not received previous warfarin treatment and suffered from either atrial fibrillation or venous thromboembolism (30). EU-PACT used a biomarker strategy design without biomarker assessment in the control arm (47).

3.3.2.3 Results

In the 427 participants, genotype-guided dosing was associated with an increased percentage of time in therapeutic range, a key measure of anticoagulation success (7.0% increase, 95% CI 3.3-10.6, $p<0.001$). Participants in the genotype-guided group also reached this therapeutic range faster than those in the control group (1.43 days faster, 95% CI 1.17-1.76, $p<0.001$). There were fewer dose adjustments required in the genotype guided group (4.9 compared to 5.4 in the control group, $p=0.02$) (30).

3.3.2.4 Evidence used to justify biomarkers

The published 2009 protocol cited mostly observational studies as evidence (Figure 3.5). These consisted of 19 cohort studies (77, 83-100) and 4 case-control studies (101-104). The cohort studies included various anticoagulants with similar mechanisms of action to warfarin. They investigated the association between warfarin dosing and *CYP2C9*2, CYP2C9*3,* and *VKORC1.* One paper also included

GGCX (77). Many studies were testing algorithms that included both clinical (age, body surface area, smoking status) and genetic factors (83, 96-98, 100). One examined the incidence of over-anticoagulation in one anticoagulation clinic (85). Another compared the efficacy of algorithms across black and white populations (96).

The authors also cited a 2009 genome-wide association study (GWAS) that showed the implications of specific *CYP2C9*, *VKORC1*, and *CYP4F2* genes on warfarin dosing (94).

Also cited were editorials (68, 105), cost-effectiveness analyses of genotype-guided vs standard dosing of warfarin (87, 106), and a literature review of economic evaluations of warfarin dosing (107). No previous RCTs were cited.

Figure 3.5 - Evidence cited by the EU-PACT trial to justify inclusion of the biomarkers. The numbers at the top represent years relative to the protocol publication date of 2009 (66).

3.3.2.5 Discussion

EU-PACT was a large, international undertaking and one of the first RCTs to investigate genotype-guided dosing of warfarin. Other RCTs had been published by 2009 (108, 109) but had shown little or no benefit to genotype-guided warfarin dosing (110). Daly (2013) proposed that this was due to these studies being underpowered to detect all genetic effects (110). No pharmacogenetic studies of a similar size to EU-PACT had been published by 2009. The amount of observational evidence cited as justification for including biomarkers within this trial provides strong rationale for the RCT.

3.3.3 SHIVA

3.3.3.1 Background

SHIVA (NCT01771458) was a French phase II trial of targeted agents in oncology, published in 2015 (31, 111, 112). The trial was histology-agnostic, meaning it recruited patients with any tumour type, but with the molecular mechanism being the

inclusion criteria (113). For example, patients were recruited to SHIVA if they had tumours with mutations in the hormone receptors pathway, regardless of the tissue (31).

3.3.3.2 Methods

SHIVA included drugs previously approved in France for targeted use: erlotinib, sorafenib, imatinib, dasatinib, vemurafenib, everolimus, abiraterone, letrozole, tamoxifen, trastuzumab, and lapatinib. The study protocol was included in the supplementary data of the 2015 results paper (31). After analysis of their tumour, patients with mutations that matched one of these drugs' targets were randomised 1:1 to receive the targeted treatment, or to receive their physician's choice of treatment. Randomisation was stratified by signalling pathways and prognoses (31). Drugs were assigned based on an algorithm, taking into account the possibility of multiple mutations in each patient. The primary outcome of SHIVA was progression-free survival (PFS).

SHIVA used an enrichment trial design (47), where only patients with actionable biomarkers were randomised into the trial.

3.3.3.3 Results

While 716 patients underwent tumour sampling, only 293 had tumours with actionable mutations and were enrolled in the trial. Breast adenocarcinoma was the most common tumour type. There was no significant different in median PFS in the experimental group compared to the control group (hazard ratio 0.88, 95% CI 0.65-1.19, p=0.41) (31). Median PFS was not significant in subgroup analyses by molecular pathway. There was also no statistically significant difference in rates of adverse events between groups.

3.3.3.4 Evidence used to justify biomarkers

The evidence cited in the protocol ranged from 1998 to 2011 (Figure 3.6). Four RCTs were cited (114-117). Two of these were trials of gefinitib in lung cancer (114, 115). Another RCT cited was an investigation of trastuzumab in HER2+ breast cancer patients, a combination that was investigated in SHIVA (116). The final RCT was the BATTLE trial, which was an adaptive biomarker-based study in lung cancer patients (118). This trial was similar to SHIVA in that patients received treatments based on their biomarkers, but was specific to lung cancer.

Two observational studies were cited. The first examined patients with metastatic colorectal cancer for *KRAS* mutations (119). The second was a pilot study with

similar traits to the SHIVA protocol – 66 patients with metastatic cancer were matched to targeted agents based on their tumour molecular profile. This small study reported an improvement in PFS, using patients' previous regimens as controls (120). SHIVA authors also cited a contemporaneous editorial commenting on this study (121).

The paper reporting on the results of this trial included an 'Evidence before this study' box (31). This detailed a literature search performed prior to the start of the trial, which identified several additional observational cohort studies (120, 122-125) and RCTs (126-128) . These were not the same papers cited in the protocol as evidence for the inclusion of the biomarkers.

Figure 3.6 - SHIVA trial evidence cited for biomarker justification. The numbers at the top represent years relative to the publication of the 2014 protocol (included in Supplementary of a 2015 paper (31)). RCT = randomised controlled trial

3.3.3.5 Discussion

Due to the number of targeted agents evaluated, it would have been difficult for the SHIVA trial to provide extensive evidence for each genetic marker-drug combination. However, the amount of evidence cited in the trial protocol is sparse and only concerns a small number of drugs. Since at least some evidence for each drug is likely available, the authors could have presented this in a table or supplementary figure. The limited scope of a protocol clearly constrained the full citation of justifications, as further justifications are included in the results paper.

3.3.4 GIST

3.3.4.1 Background

The US *SLCO1B1* genotype informed statin therapy (GIST) trial (NCT01894230) investigated the utility of using genotyping to increase adherence to statins and promote lower cholesterol in patients with cardiovascular disease and a history of statin-induced side effects (32, 33, 36).

Non-adherence to statins in patients with cardiovascular disease is a known problem (129, 130). Non-adherent patients face higher risks of hospitalisation and mortality (adjusted hazard ratio 1.36, 95% CI 1.34-1.38 in the least adherent patients compared to the most adherent) (131). While some have reported 70-80% adherence in clinical trial settings (132, 133), real-world usage results in much lower rates of adherence (134, 135).

The aim of the *SLCO1B1* GIST trial was to improve adherence by showing patients that treatment includes an assessment of the risks (real and perceived) of statin-induced side-effects (32). A trial protocol with rationale and design details was published in 2016 (32), and the results published in 2018 (33).

3.3.4.2 Methods

Recruitment to the trial focussed on patients that had previously discontinued statin therapy due to suspected side-effects. Patients were genotyped for *SLCO1B1* and then randomised 1:1 to receive genotype information as part of their care, or to receive usual care alone. The trial was unblinded and the primary outcome was patient-reported adherence to therapy. Secondary outcomes included low-density lipoprotein cholesterol (LDLc), number of new statin prescriptions, and patient-reported quality of life.

The trial used a biomarker strategy with biomarker assessment in the control arm design (47).

3.3.4.3 Results

While the trial initially recruited 159 participants, only 62 were available for analysis of adherence at 3 months (higher numbers were available for analysis of other outcomes). There was no statistically significant difference in patient-reported adherence between the genotype and control arms (p=0.96). This remained true at the 8 month time point. However, LDLc levels were significantly lower in the genotyped group compared to the control group at 3 months (131.9 vs 144.4, p=0.05). Analysis of the total study population revealed an interaction between *SLCO1B1* status, randomisation, and LDLc. This may indicate a psychological effect of having an ‘actionable’ test (positive result, adjustment of statin dosage according to genotype) compared to a reassuring but negative test result.

3.3.4.4 Evidence used to justify biomarker

The trial cited a large number of references, dated from 2002 to 2015 (Figure 3.7).

Five sets of guidelines from four separate bodies were cited (136-140). These included two Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines on statin therapy in patients with the *SLCO1B1* polymorphism rs4149056 (from 2012 and 2014) (136, 137), broad guidelines on lowering cardiovascular risk (138), on statin treatment (139), and European guidelines on statin-induced muscular symptoms (140).

Also included was a report on heart disease and stroke epidemiology from the American Heart Association (141). Seven literature reviews were cited (129, 142-147), alongside two editorials regarding lowering cholesterol (148) and statin effects on muscular ADRs (149) . This trial also cited the largest number of observational studies, a total of 11 (consisting of 1 case control study (150), 9 cohort studies (151-159), and 1 cohort/meta-analysis study (160)). These included a large study in the US, defining how many people would be eligible for statins under current guidelines and the impact of this (151). There were 4 cohort studies that examined the relationship between statin adherence and cardiovascular outcomes (152-155).

Also included was a small cohort study that served as a pilot for GIST (157). Two cohort studies were in healthy participants, examining the genetics of statin-induced myopathy (156), and how the pharmacokinetics of two common statins differs between people of different ancestries (159).

In contrast to the large amount of observational study evidence, the trial cited only one RCT (161). However, two further references were genetic sub-studies of larger RCTs (162, 163). A 2013 Cochrane review was also cited (164).

Figure 3.7 – Evidence cited by the GIST statin trial to justify inclusion of the SLCO1B1 *biomarker. The numbers at the top represent years relative to the publication of the trial protocol in 2016 (32).*

The authors cited one systematic review (165) and three meta-analyses (166-168).

The systematic review of adherence in statin patients (165) assessed the quality of included studies using guidelines from the International Society for Pharmacogenomics and Outcomes Research (169).

Two meta-analyses were published by the Cholesterol Treatment Trialists' Collaborators (CTTC) group (166, 167), a group established in 1994 to perform meta-analyses of long-term and large-scale trials of lipid intervention therapies (170). These analysed the risk of major vascular events in statin users (166), and compared the risk of vascular events between male and female statin users (167).

The final meta-analysis was about the risk of statin-related myopathy (168) and evaluated quality using the Newcastle-Ottawa scale (171).

The meta-analyses by the CTTC group were both done on the same large data set of n=174,149 participants from 27 RCTs investigating the impact of statins on cardiovascular risk (166, 167). Each RCT had to have a recruitment target of >1000 participants, and have a minimum 2-year treatment duration. Whilst these meta-analyses did not assess the quality of the included studies, they both collated individual participant data (IPD).

3.3.4.5 Discussion

Even though GIST did not find that pharmacogenetics improved statin adherence, a modest effect on LDLc was observed. The group observed for adherence was small, especially compared to the original sample size. The reason so few patients completed adherence data was not addressed in the paper. The authors considered that there may be other, unobserved barriers to adherence. There are also known issues with using self-reported measures of adherence (172).

GIST cited the highest quantity of evidence out of the included trials. This also included high-quality evidence in the form of meta-analyses and a Cochrane review. This is a clear case where high-quality supporting evidence did not guarantee a significant result when testing the biomarker-guided treatment approach within a clinical trial.

SLCO1B1 is designated as a very important pharmacogene by PharmGKB (173). The PharmGKB summary paper of this addition was published in 2010, which would have provided additional high-quality evidence for this trial.

3.3.5 Precision Medicine Guided Treatment for Cancer Pain

3.3.5.1 Background

This trial (NCT02664350) investigated the effect of pharmacogenetic information on the treatment of pain with opioids in cancer patients (34). The trial used the *CYP2D6* gene as a biomarker. This gene encodes the CYP2D6 enzyme, one of the most widely investigated liver enzymes (174). It has been estimated that this enzyme affects metabolism of approximately 25% of marketed drugs (174, 175). As of 2020, there are over 300 known variations in *CYP2D6* linked to specific phenotypes (176). In relation to opioids, patients can be categorised by the number and functionality of their *CYP2D6* alleles (Table 3.1) (34).

Phenotype	Genetics	Enzyme activity	Clinical outcome
Ultra-rapid metabolisers	Multiple gene copies	Many times more active	Increased production of metabolites associated with toxicity (177, 178)
Normal metabolisers	At least 1 fully functional allele or 2 partially functioning alleles	Normal	Normal dosing (179)
Intermediate metabolisers	1 loss-of-function allele and 1 reduced function allele	Significantly impaired	Risk of decreased analgesia and ADRs (180, 181)
Poor metabolisers	No functional alleles	Little to no active enzyme	Greater risk of decreased analgesia and ADRs (179, 180)

Table 3.1 - Different CYP2D6 *phenotypes and their underlying genetics and enzyme activity (34, 179).*

This trial used *CYP2D6* genotyping to guide the dosing of opioids for treating cancer pain in patients with metastatic solid tumours. The trial began in 2016, and was completed in 2019, but results have not yet been published (182).

3.3.5.2 Methods

The trial aimed to recruit 200 participants with metastatic solid tumours and pain scores >=4 on a scale of 1-10 (Brief Pain Inventory scale) (183). Participants were randomised 1:1 to receive either *CYP2D6*-guided or conventional selection of pain medication. All participants were genotyped, but only those in the first group had their results entered into their electronic health record, accessed by their physician. This was accompanied by an interpretation of how their genotype would affect their

opioid metabolism and a recommendation for changes to usual prescribing. Those in the control group received standard of care cancer pain treatment. Pain questionnaires were completed at baseline, 2, 4, 6, and 8 weeks, with change in pain severity the primary outcome.

The trial used a biomarker strategy without biomarker assessment in the control arm design (47).

3.3.5.3 Results

The results of the trial have not yet been published (182). In response to an email enquiry, the authors stated that the results are currently being written up and prepared for publication (September 2020) (184). The results were not yet published as of June 2021.

3.3.5.4 Evidence used to justify biomarker

The authors cited evidence ranging from 1998 to 2017 (Figure 3.8). The oldest evidence was a 1998 RCT (185), that randomised healthy volunteers to high and low doses of opioids, stratified by their *CYP2D6* genotypes. This was cited alongside four other RCTs that investigated the effects of randomised doses of opioids in healthy participants with varying *CYP2D6* genotypes (186-189), and an RCT of palliative care in patients with metastatic lung cancer (190).

The most recent evidence was 2017 guidelines on adult cancer pain from the National Comprehensive Cancer Network (191). Interestingly, the trial cited three case studies; one in a patient with the poor metabolizer phenotype (192), and two with patients with the ultra-rapid metabolizer phenotype (193, 194). No evidence for normal or intermediate metabolism was presented.

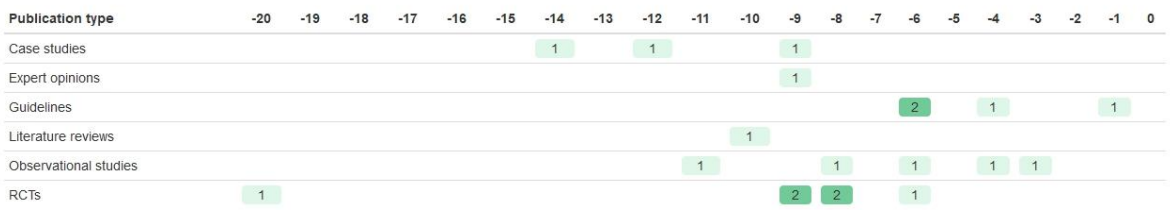


Figure 3.8 – Evidence cited by the Precision Medicine Guided Treatment for Cancer Pain trial to justify inclusion of the CYP2D6 *biomarker. The numbers at the top represent years relative to the publication of the protocol in 2008 (34).*

3.3.5.5 Discussion

This trial is an interesting example of using pharmacogenetics to improve patient quality of life. Although the results of this trial are not currently available, a similar trial published results in 2019 (180). This trial, also using *CYP2D6*-guided dosing but in patients with any chronic pain, found a significant reduction in pain scores in genotype-guided vs. conventional dosing groups (p=0.016). This was more pronounced in intermediate/poor metabolisers than in normal metabolisers. The previously performed RCTs cited by this trial are all from more than 5 years prior to its start. There is therefore a rationale here for an updated trial, although a meta-analysis of the previous trials should have been conducted where appropriate.

CYP2D6 was designated a 'very important pharmacogene' by PharmGKB in 2015 (195). This would have provided additional high-quality evidence for its inclusion in this trial.

3.4 Discussion and Recommendations

There does not appear to be a standard approach for gathering evidence for justifying biomarker inclusion within a biomarker-guided trial, and the trials in this review all used different approaches to do so. Of the trials examined, all cited evidence from within at least 3 years of their publication, but evidence was also cited from much earlier than that (Figure 3.2). The oldest evidence compared to trial start date was cited by the TARGET trial, which cited work from 25 years prior to its 2005 protocol date (48).

The evidence types used included systematic reviews/meta-analyses, RCTs, qualitative research, guidelines, recommendations, editorials, and case studies. According to national and international guidelines (37, 196-198), the randomised controlled trial (RCT) is the 'gold standard' of evidence. Regulatory bodies still usually require high quality RCT evidence as a minimum to accept a new technology. For example, of 795 European Medicines Agency (EMA) approvals from 1999-2004, only 44 (5.5%) were products for which there were no RCT results (39).

A hierarchy of evidence was used by Concato, *et al.* (2000) (199), referencing an older US guideline (200). In this guideline, the top grade of evidence is that obtained from "at least one properly randomised, controlled trial". Lower tiers include trials without randomisation, cohort and case-control studies, and time series, with clinical

opinions and case reports at the bottom. However, Concato, *et al.* demonstrated that the results of non-randomised trials often correlate closely with those from RCTs (199). This is explored further in Chapter 4.

The traditional 'evidence pyramid' is often used to rank evidence types, with meta-analyses and systematic reviews at the top, and case studies and *in vitro* evidence near the base (Figure 3.9A) (201). However, this has seen some modification in recent years, notably the viewing of evidence through the 'lens' of systematic reviews and meta-analyses, ensuring that the quality of included studies is evaluated (202). The rest of the evidence should then be viewed through the lens of these studies – using them as "tools to consume and apply the evidence" (202) (Figure 3.9B). In this iteration, a meta-analysis based on weak evidence, suffering from bias, is not automatically seen as superior evidence to a well-conducted observational study.

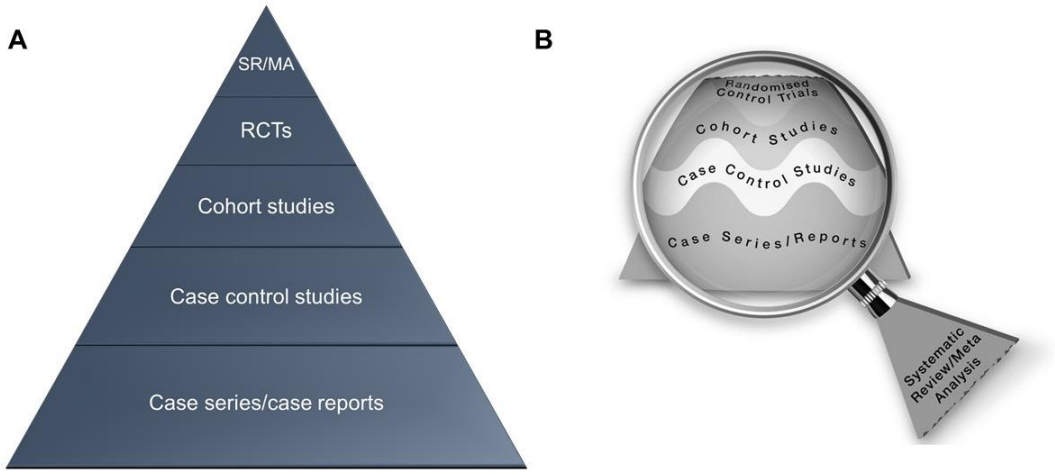


Figure 3.9 – A) Traditional pyramid of evidence and B) An updated version put forward by Murad et al (2016). RCTs = randomised controlled trials. SR/MA = Systematic review/meta-analysis

Looking at these pyramids, one might expect that most or all of the trials cited here would include a systematic review or meta-analysis in their justification for biomarker inclusion. In practice, only two of the trials did (TARGET and GIST). Further, only two of the trials cited existing RCTs as evidence (SHIVA and GIST). Observational studies were the most common evidence type cited. This category encompasses any non-randomised trial, including cohort, retrospective, and case-control studies.

To explore the type and extent of evidence compiled to justify including biomarkers in previous biomarker-guided trials, I have examined the previous literature on biomarker associations referenced within the trial design paper or protocol. This represents a relatively straightforward method of assessing the evidence used to justify a biomarker's inclusion in a trial, however, it has some inherent limitations. First, this method will not necessarily capture the entire evidence base upon which inclusion of the biomarker was justified, since the authors may not have provided a complete and accurate snapshot of the evidence they explored and used. Second, journal rules on the number of references in a paper and word count restrictions could mean that the references included do not represent the totality of evidence used. Similar restrictions on references and word counts may limit the representation of the literature in protocols.

Publicly available published protocols were used as a source for evaluating biomarker evidence for three trials in this review (EU-PACT, GIST, and NCT02664350). For the TARGET and SHIVA trials, formal protocols, obtained from the authors, and from supplementary sections of papers were used. Protocols like these are primarily used for reference by triallists and while they should be publicly available, they are not always made so (203). While there are guidelines for writing protocols, they do not usually have a set structure. The Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) checklist, published in 2013, specifically recommends that protocols include a section describing the 'justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention' (204). All trials examined fulfilled this brief to some extent. The question remains, what forms should this justification take in order to provide sufficient evidence for beginning a clinical trial? The SPIRIT checklist does not provide specific guidance in this regard.

3.4.1 Limitations

In this review, I aimed to look at the evidence available at the time the trial began. Only one of the included trials included their own systematic review of the evidence. To get a true picture of the evidence available before a trial started, I may have considered conducting systematic reviews of the evidence available for each, setting the trial start date as a cut-off. This was considered unfeasible in light of the wider PhD aims. Furthermore, I believe the method used is one that could easily be reproduced by other researchers reporting on trials.

At the outset of this review, I initially planned to conduct a thorough systematic review of all previously conducted genetic biomarker-guided trials. Preliminary scoping searches yielded over 20,000 results and after narrowing the selection by title and abstract screening, when around 500 papers still remained for full text evaluation, the approach was abandoned as not feasible. The decision was made that a better approach would be to use the collected list of trials to choose five RCTs representing different biomarker applications, and perform an in-depth analysis of these trials' evidence for biomarker inclusion.

3.4.2 Recommendations

While the ideal level of evidence prior to proceeding to a biomarker-guided clinical trial is a well-conducted meta-analysis/systematic review of good quality RCTs, including a rigorous assessment of their quality, this is not always available or feasible. In particular, where a biomarker is very new, there may be limited previous evidence to underpin its use. This evidence may take the form of case series or previous case studies. If this is the only evidence available, then this may be the 'best' evidence to justify including the biomarker in a trial. It would be important to consider, in such circumstances, whether the proposed RCT would be premature and that the science should first of all be allowed to mature.

It may be that different standards of evidence are necessary for different biomarker types (196, 205). For example, evidence standards could be based on risk, with biomarkers for lower risk applications requiring less evidence and regulatory oversight than those for high risk applications (196). Recommendations could also be based on the disease being treated, similar to how orphan drugs for rare diseases are given accelerated approvals (206, 207). Biomarkers used for more serious indications could be allowed to proceed to trial with less or lower quality evidence than biomarkers for less serious conditions. Novelty of the biomarker will also influence the extent of evidence available – a biomarker first utilised in 1980 is likely to have accumulated much more evidence than one first described in 2015.

In the UK, a 2018 workshop attended by academics, representatives of CPIC and PharmGKB, and clinicians evaluated the barriers to implementation of pharmacogenetics in the NHS (8). Variability in the evidence for the effectiveness of testing for different gene-drug combinations was identified as a key barrier. The group proposed that initially only associations with the best evidence are phased into clinical practice, with others being released if the evidence for their use improves. They acknowledged the difficulty that can be inherent in conducting RCTs

in pharmacogenetics (18, 208, 209), and recommended the use of novel study designs (including the use of real-world electronic health data).

It is also important to ensure that genetic biomarkers are not subject to higher evidentiary requirements than other types of biomarkers. This 'genetic exceptionalism' and the higher burden of evidence for genetic tests has been shown to be a significant barrier to clinical implementation (9, 205, 210, 211). The level of evidence required for a genetic biomarker before its use in a trial should be the same as an equivalent non-genetic biomarker. Finally, biomarkers that are integral to a trial's conduct require more evidence than biomarkers used on an exploratory basis (see Chapter 2) (24).

With these factors in mind, my recommendations related to compiling evidence to justify proceeding to a genetic biomarker-guided trials consist of three steps: a systematic review before embarking on a trial: more guidance from regulatory authorities; and the need for all stakeholders to consider diversity in recruiting to trials.

1. *Systematic review before embarking on a trial*

I would recommend an initial systematic review is undertaken prior to the start of any trial. While other authors have also recommended this (204, 212, 213), few RCTs include systematic reviews of the evidence for their choice of intervention (212). The Lancet journal now requires all research papers to include a 'Research in Context' panel that shows the evidence available prior to the study, and how the authors searched for this information (214, 215). Many top journals require authors to follow and submit a Consolidated Standards of Reporting Trials (CONSORT) checklist (see below) with the trial publication (Table 3.2).

Journal	Policy last updated	Summary of policy
PLOS One	Not found (216)	Trials must adhere to the CONSORT statement (217). Does not explicitly require a systematic review.
The New England Journal of Medicine	Not found (218)	Allows up to 40 references. Asks authors to include a trial's protocol with submission, but no specific reference to systematic reviews.
Annals of Internal Medicine	Not found (219)	Allows up to 75 references. Endorses the CONSORT statement for reporting RCTs(217). Does not explicitly require a systematic review.

The Lancet	November 2015 (215)	'Research in context' panel required, that should include 'a description of all the evidence that the authors considered before undertaking this study'. This also states that authors should include the search terms used, inclusion/exclusion criteria, and an evaluation of the quality of that evidence. A meta-analysis of the evidence should also be included if appropriate. Reports should also conform to the CONSORT guidelines (217).
The BMJ	May 2018 (220)	All papers should include a summary box, that incorporates 2-3 single sentence bullet points on the state of scientific knowledge on the topic, before the study started. Clinical trial submissions should use the CONSORT statement (217). Does not explicitly require a systematic review. The TIDieR checklist is also recommended (221).
JAMA: The Journal of the American Medical Association	October 2020 (222)	Authors must include a copy of the trial protocol and should use the CONSORT statement (217). Does not explicitly require a systematic review.
Nature	Not found (223)	30-50 references, as a guide. A completed CONSORT checklist must be included with all clinical data submissions (217).
Neurology	Not found (224)	Limit of 50 references. Authors must submit a CONSORT checklist (217).
Proceedings of the National Academy of Sciences of the United States of America	August 2020 (225)	Limit of 50 references. No other details found.
Science	Not found (226)	Expected to contain about 40 references. Reports should include a completed CONSORT checklist. RCTs not conforming to CONSORT guidelines may be returned to authors for revision.

Table 3.2 - Policies of journals on evidence required for randomised controlled trials. Top ten journals as cited by Jemielniak et al 2019 (227). Cochrane Database of Systematic Reviews excluded as it does not publish RCTs, replaced with the next highest ranking, across three time periods, not already included (Science).
CONSORT guidelines include provision for 'explanation of rationale' (217). However, this does not explicitly require a systematic review.

The CONSORT statement includes a requirement that papers should 'explain the scientific background and rationale for their trial' (217). While the CONSORT checklist is essential for ensuring reportability and reproducibility in trials, journals should consider explicitly stating the need for a quality systematic review of the evidence available prior to a trial's start date. Funding sources also have a role to play in enforcing this in applications and grants.

Further, if the systematic review reveals a sufficient number of previous RCTs or observational studies, authors should consider conducting a meta-analysis to assess the current evidence quantitatively. This would help ascertain whether there was sufficient uncertainty surrounding the current evidence to justify the planned RCT. An example of where this could have been implemented is in the fifth trial examined (34). Authors can also utilise funnel plots to examine any potential bias in the publication of included studies (228), and explore any heterogeneity between studies. They can also incorporate their new data into the analysis, putting their work into context amongst the literature. This was completed successfully in a prospective study and meta-analysis by Genin, *et al.* (2013) (229).

Regardless of the type of evidence identified in the systematic review, it is recommended that the quality of that evidence is also assessed when justifying including the biomarker, and I suggest that design-specific tools are used for this purpose (e.g. the Cochrane Collaboration's Risk of Bias tool for RCTs) (230). Several study type-specific methods for doing this are available (171, 197, 230-233) and have been reviewed by Zeng, et al. (2015) (234). Additionally, the quality of pharmacogenetic studies should be assessed using the guidelines proposed by Jorgensen and Williamson (2008) (18). Wang, *et al.* (2014) accepted a systematic review/meta-analysis as 'convincing' evidence only when there was consistency in the results (235). Reviews showing heterogeneity were downgraded to 'adequate' evidence. This is a useful heuristic for the evaluation of data, although as discussed, it is important to also consider the quality of included papers.

Authors should also consider including an analysis where their trial data is incorporated into the existing literature using a meta-analysis. This has been used successfully in several trial reports (160, 236-238), but remains rare. A 2010 review of 29 RCTs (not all in pharmacogenetics) found that only 1 contained an updated systematic review integrating the RCT results (212). This approach should be more widely used.

When synthesising evidence already existing from previous studies, it is also important to consider the age and ethnicities of the populations of the previous studies compared to the proposed trial's population to ensure that the evidence is relevant. This is further explored below.

2. Guidelines are required

Given the lack of standardisation across biomarker trials in terms of how inclusion of biomarkers is justified, I recommend that guidelines are developed to aid researchers in compiling and presenting evidence to justify their inclusions. This will not only ensure that sufficient evidence exists prior to embarking on a biomarker trial, thus avoiding waste of resources, but will also serve as a useful guide to those planning a biomarker trial and provide transparency in the trial report.

As previously discussed in Chapter 2, there is little guidance from regulatory bodies on the evidence required before proceeding to a pharmacogenetic trial. Biomarkers should undergo validation before their use in a clinical trial. This process is not well-defined by regulators, however public and private consortia have developed various guidelines that might help in this regard.

For example the CPIC provides guidelines for the implementation of pharmacogenetics (11). The guidelines provide a grading of the level of evidence given in support of the biomarker's implementation ('high', 'moderate' or 'weak') (239). The CPIC levels are based on PharmGKB criteria, where the evidence for a gene-drug association is rated on a six-point scale between 1A (guidelines endorsed by a medical society or major health system) to 4 (in vitro, case study, or nonsignificant study evidence) (21). This scale is based on clinical annotations obtained from PubMed, produced by manually combining and summarising associations from several publications (240). These clinical annotations are then given a 'level of evidence' score based on replication of the association, P-value, and odds ratio. The score is determined by PharmGKB curators (21). This process has produced excellent results, but is labour-intensive and time-consuming. A new alternative is the automated text-mining process outlined by Lever, *et al.* (2020) (241). In time, this will further improve PharmGKB as a source of pharmacogenetic information. The guidelines produced by PharmGKB are influential. The database is widely used by clinicians (240), and this will include principal investigators for potential pharmacogenetic trials.

Whilst these guidelines are for implementation of biomarkers into clinical practice, a similar approach could be developed for justification of inclusion in an RCT. One

paper was located that discussed the incorporation of biomarkers into early phase clinical trials (24), but I recommend that this needs to contribute to the formation of formal guidelines for biomarker trials similar to CPIC guidelines for biomarker implementation.

3. Diversity in clinical trials

If most pharmacogenetic studies are conducted in just one ancestry group, any resulting clinical data will have limited usage in other groups. This must be considered when compiling the evidence for a biomarker's use in a trial. While equity of access is important in all research, this is an especially important consideration in pharmacogenetics, since groups of different ancestries will have different frequencies of actionable alleles (12, 242-244). The genomes of African populations are the most underrepresented in genetic research (12, 242, 245, 246). Bentley, *et al.* (2017) provided an overview of why genomic research in diverse populations is important, including: the potential to gain novel insights into health disparities and human biology, improving clinical care, informing genetic diagnoses, and promoting a better understanding of human history (247).

A 2010 study underlined the importance of including diverse cohorts in pharmacogenetic studies (248). The authors compared the performance of published pharmacogenetic warfarin algorithms by using them to calculate dosage in a database of warfarin-treated patients. Algorithms derived from studies in mixed-ethnicity populations outperformed those from non-mixed populations. None of the selected algorithms were derived from African American populations. Accordingly, the algorithms all had their highest mean absolute error when applied to African American patients.

A cohort study of 274 warfarin treated African Americans found similar dosing errors when using standard algorithms (249), and found that genetic markers associated with warfarin dose requirements in African Americans were not captured in pharmacogenetic dosing algorithms developed in cohorts including patients from other ethnic groups.

Clearly, stronger evidence for some 'established' biomarkers (such as *CYP2C9* and *VKORC1* in warfarin pharmacogenetics (248-250)) is needed in certain ethnic groups. There are particular challenges with implementing pharmacogenetics in the developing world (5, 7). While under-resourced health systems are a leading factor

in this, the surplus of pharmacogenetic data in European and Asian populations makes further study in other populations more difficult (12, 248). Novel variants are harder to predict and are subsequently under-researched (12).

A study of pharmacogenetics in 141 Ghanaian warfarin patients did not detect any non-wild type *CYP2C9*2* or **3* alleles (251), two alleles commonly used in other dosing algorithms (252). A similar study in Caribbean Hispanic Puerto Ricans found that the rs2860905 variant in *CYP2C9* was a stronger predictor of warfarin dosing than the other, more well-known *CYP2C9* alleles (253).

Trials that include patients from traditionally underserved areas are therefore required to provide high-quality evidence for pharmacogenetics use in these populations. The GUARDD-US trial (NCT04191824) is recruiting participants in the USA with African ancestry to determine the effect of providing pharmacogenetic knowledge to participants and clinicians on blood pressure management (254). Other upcoming trials and initiatives (such as H3Africa, a biobanking initiative (255) and the War-PATH study in warfarin pharmacogenetics (256)) should improve access to pharmacogenetics. This will allow the construction of a strong evidence base that includes multiple ethnic groups.

3.5 Conclusion

This work has shown that there is currently no standard approach for collecting evidence to justify the inclusion of a biomarker in a biomarker-guided trial. Each of the trials here took a different approach. GIST was able to rely on large and robust meta-analyses for evidence, while there was less evidence cited by SHIVA for all their drug-gene combinations.

This variability underlines the need for guidelines for trialists on how to compile evidence to justify the inclusion of a biomarker within a trial. Best practice should include a systematic review of the evidence before a trial commences, and a meta-analysis where appropriate. Trials also need to be conscious of differences between ethnic groups, and to ensure that the evidence is based on the same ethnic groups as those being studied within the trial.

It is clear that the full potential of pharmacogenetics cannot be unlocked without significant work to remove barriers to implementation. Many commentators have identified a lack of well conducted studies demonstrating clinical utility of personalised approaches to treatment as one of the largest roadblocks to

implementation (4-6, 12, 257-259), and the gold-standard for demonstrating such clinical utility is the randomised controlled trial. With these recommendations, it is hoped that such trials will be based on a strong and robust evidence base, thus ensuring that biomarker-guided trials are only conducted when there is sufficient preliminary evidence that the biomarker may be useful in personalising treatment. This will increase the chance of success whilst minimising waste in resources.

Finally, the conclusions and recommendations above assume that a trial is indeed required. It is possible that when compiling the evidence to justify inclusion of a biomarker in a trial that it is so overwhelmingly in favour of the biomarker's clinical utility that it may be unethical to restrict its use to a randomised trial. This loss of clinical equipoise is something important to consider and indeed clinical implementation may be recommended and accepted without the need for a biomarker trial in such cases. This is a consideration that has been explored by several authors (260-263) and is considered further in Chapter 4.

3.6 References

1. Hippman C, Nislow C. Pharmacogenomic Testing: Clinical Evidence and Implementation Challenges. J Pers Med. 2019;9(3).
2. Mathias PC, Hendrix N, Wang WJ, Keyloun K, Khelifi M, Tarczy-Hornoch P, et al. Characterizing Pharmacogenomic-Guided Medication Use With a Clinical Data Repository. Clinical pharmacology and therapeutics. 2017;102(2):340-8.
3. Department of Health. Our Inheritance, Our Future Realising the Potential of Genetics in the NHS: HM Stationery Office; 2003.
4. Lam YW. Scientific challenges and implementation barriers to translation of pharmacogenomics in clinical practice. ISRN Pharmacol. 2013;2013:641089.
5. Klein ME, Parvez MM, Shin JG. Clinical Implementation of Pharmacogenomics for Personalized Precision Medicine: Barriers and Solutions. J Pharm Sci. 2017;106(9):2368-79.
6. Arwood MJ, Chumnumwat S, Cavallari LH, Nutescu EA, Duarte JD. Implementing Pharmacogenomics at Your Institution: Establishment and Overcoming Implementation Challenges. Clin Transl Sci. 2016;9(5):233-45.
7. Limaye N. Pharmacogenomics, Theranostics and Personalized Medicine - the complexities of clinical trials: challenges in the developing world. Appl Transl Genom. 2013;2:17-21.
8. Turner RM, Newman WG, Bramon E, McNamee CJ, Wong WL, Misbah S, et al. Pharmacogenomics in the UK National Health Service: opportunities and challenges. Pharmacogenomics. 2020;21(17):1237-46.
9. van der Wouden CH, Cambon-Thomsen A, Cecchin E, Cheung KC, Davila-Fajardo CL, Deneer VH, et al. Implementing Pharmacogenomics in Europe: Design and Implementation Strategy of the Ubiquitous Pharmacogenomics Consortium. Clinical pharmacology and therapeutics. 2017;101(3):341-58.
10. Johnson D, Hughes D, Pirmohamed M, Jorgensen A. Evidence to Support Inclusion of Pharmacogenetic Biomarkers in Randomised Controlled Trials. J Pers Med. 2019;9(3):42.

11. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. Clinical pharmacology and therapeutics. 2011;89(3):464-7.
12. E BT, M AA, M SP. Barriers to Implementing Clinical Pharmacogenetics Testing in Sub-Saharan Africa. A Critical Review. Pharmaceutics. 2020;12(9):809.
13. Zhong A, Darren B, Loiseau B, He LQB, Chang T, Hill J, et al. Ethical, social, and cultural issues related to clinical genetic testing and counseling in low- and middle-income countries: a systematic review. Genet Med. 2021;23(12):2270-80.
14. McKinnon RA, Ward MB, Sorich MJ. A critical analysis of barriers to the clinical implementation of pharmacogenomics. Ther Clin Risk Manag. 2007;3(5):751-9.
15. Manolio TA. Incorporating Whole-Genome Sequencing Into Primary Care: Falling Barriers and Next Steps. Annals of internal medicine. 2017;167(3):204-5.
16. Chin L, Devine B, Baradaran S, Keyloun K, Canestaro W, Pham J. Characterizing the Strength of Evidence in FDA Labels for Pharmacogenomic Biomarker-Guided Medication Use. AMIA Summits on Translational Science Proceedings. 2017;2017:30-9.
17. Vivot A, Boutron I, Ravaud P, Porcher R. Guidance for pharmacogenomic biomarker testing in labels of FDA-approved drugs. Genet Med. 2015;17(9):733-8.
18. Jorgensen AL, Williamson PR. Methodological quality of pharmacogenetic studies: issues of concern. Statistics in medicine. 2008;27(30):6547-69.
19. Vivot A, Boutron I, Beraud-Chaulet G, Zeitoun JD, Ravaud P, Porcher R. Evidence for Treatment-by-Biomarker interaction for FDA-approved Oncology Drugs with Required Pharmacogenomic Biomarker Testing. Sci Rep. 2017;7(1):6882.
20. Poste G. Bring on the biomarkers. Nature. 2011;469(7329):156-7.
21. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clinical pharmacology and therapeutics. 2012;92(4):414-7.
22. Lesko LJ, Atkinson AJ, Jr. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. Annu Rev Pharmacol Toxicol. 2001;41(1):347-66.
23. Rolan P. The contribution of clinical pharmacology surrogates and models to drug development--a critical appraisal. British journal of clinical pharmacology. 1997;44(3):219-25.
24. Yee LM, Lively TG, McShane LM. Biomarkers in early-phase trials: fundamental issues. Bioanalysis. 2018;10(12):933-44.
25. Haddow JE, Palomaki GE. ACCE: A model process for evaluating data on emerging genetic tests. In: Khoury MJ, Little J, Burke W, editors. Human Genome Epidemiology. New York: Oxford University Press; 2004. p. 217-33.
26. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med. 2009;11(1):3-14.
27. Gillis NK, Innocenti F. Evidence required to demonstrate clinical utility of pharmacogenetic testing: the debate continues. Clinical pharmacology and therapeutics. 2014;96(6):655-7.
28. Hayes DF, Allen J, Compton C, Gustavsen G, Leonard DG, McCormack R, et al. Breaking a vicious cycle. Sci Transl Med. 2013;5(196):196cm6.
29. Newman WG, Payne K, Tricker K, Roberts SA, Fargher E, Pushpakom S, et al. A pragmatic randomized controlled trial of thiopurine methyltransferase genotyping prior to azathioprine treatment: the TARGET study. Pharmacogenomics. 2011;12(6):815-26.
30. Pirmohamed M, Burnside G, Eriksson N, Jorgensen AL, Toh CH, Nicholson T, et al. A randomized trial of genotype-guided dosing of warfarin. The New England journal of medicine. 2013;369(24):2294-303.

31. Le Tourneau C, Delord JP, Goncalves A, Gavoille C, Dubot C, Isambert N, et al. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. Lancet Oncol. 2015;16(13):1324-34.
32. Singh K, Peyser B, Trujillo G, Milazzo N, Savard D, Haga SB, et al. Rationale and design of the SLCO1B1 genotype guided statin therapy trial. Pharmacogenomics. 2016;17(17):1873-80.
33. Peyser B, Perry EP, Singh K, Gill RD, Mehan MR, Haga SB, et al. Effects of Delivering SLCO1B1 Pharmacogenetic Information in Randomized Trial and Observational Settings. Circulation Genomic and precision medicine. 2018;11(9):e002228.
34. Mosley SA, Hicks JK, Portman DG, Donovan KA, Gopalan P, Schmit J, et al. Design and rational for the precision medicine guided treatment for cancer pain pragmatic clinical trial. Contemp Clin Trials. 2018;68:7-13.
35. Ollier B, Newman B, Payne K, Poulton K, Andrews J, Elliott R, et al. The TARGET Study TPMT: AZA Response to Genotyping and Enzyme Testing Protocol: A prospective randomised controlled trial of thiopurine methyltransferase (TPMT) genotyping in the management of patients, prior to commencement of AZA treatment. In: University of Manchester, editor. 2005.
36. Deepak Voora EP, Kavisha Singh, Gloria Trujillo, Nicholas Milazzo, Dillon Savard, Susanne Haga, Michael D Musty, Yi-Ju Li, Bruce Peyser. SLCO1B1 Genotype-Guided Statin Therapy Lowers LDL Cholesterol in Patients With Statin-Intolerance-A Randomized Controlled Trial. Circulation. 2016;134.
37. Barton S. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. BMJ. 2000;321(7256):255-6.
38. Amur SG, Sanyal S, Chakravarty AG, Noone MH, Kaiser J, McCune S, et al. Building a roadmap to biomarker qualification: challenges and opportunities. Biomark Med. 2015;9(11):1095-105.
39. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999-2014. BMJ open. 2016;6(6):e011666.
40. R Studio Team. RStudio: Integrated Development for R. RStudio, Inc. Boston, MA2016.
41. Ren K, Russell K. Package 'formattable' 2016 [Available from: https://cran.r-project.org/web/packages/formattable/formattable.pdf.
42. Thompson AJ, Newman WG, Elliott RA, Roberts SA, Tricker K, Payne K. The cost-effectiveness of a pharmacogenetic test: a trial-based evaluation of TPMT genotyping for azathioprine. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2014;17(1):22-33.
43. Fraser AG, Orchard TR, Jewell DP. The efficacy of azathioprine for the treatment of inflammatory bowel disease: a 30 year review. Gut. 2002;50(4):485-9.
44. Yates CR, Krynetski EY, Loennechen T, Fessing MY, Tai HL, Pui CH, et al. Molecular diagnosis of thiopurine S-methyltransferase deficiency: genetic basis for azathioprine and mercaptopurine intolerance. Annals of internal medicine. 1997;126(8):608-14.
45. Wang L, Pelleymounter L, Weinshilboum R, Johnson JA, Hebert JM, Altman RB, et al. Very important pharmacogene summary: thiopurine S-methyltransferase. Pharmacogenetics and genomics. 2010;20(6):401-5.
46. Ford LT, Berg JD. Thiopurine S-methyltransferase (TPMT) assessment prior to starting thiopurine drug treatment; a pharmacogenomic test whose time has come. J Clin Pathol. 2010;63(4):288-95.
47. Antoniou M, Jorgensen AL, Kolamunnage-Dona R. Biomarker-guided trial designs (BiGTeD): An online tool to help develop personalised medicine [Available from: http://www.bigted.org/.

48. Weinshilboum RM, Sladek SL. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. American journal of human genetics. 1980;32(5):651-62.
49. Lennard L, Van Loon JA, Weinshilboum RM. Pharmacogenetics of acute azathioprine toxicity: relationship to thiopurine methyltransferase genetic polymorphism. Clinical pharmacology and therapeutics. 1989;46(2):149-54.
50. Dubinsky MC, Lamothe S, Yang HY, Targan SR, Sinnett D, Theoret Y, et al. Pharmacogenomics and metabolite measurement for 6-mercaptopurine therapy in inflammatory bowel disease. Gastroenterology. 2000;118(4):705-13.
51. McLeod HL, Lin JS, Scott EP, Pui CH, Evans WE. Thiopurine methyltransferase activity in American white subjects and black subjects. Clinical pharmacology and therapeutics. 1994;55(1):15-20.
52. Bloomfeld RS, Onken JE. Mercaptopurine metabolite results in clinical gastroenterology practice. Alimentary pharmacology & therapeutics. 2003;17(1):69-73.
53. McLeod HL, Coulthard S, Thomas AE, Pritchard SC, King DJ, Richards SM, et al. Analysis of thiopurine methyltransferase variant alleles in childhood acute lymphoblastic leukaemia. British journal of haematology. 1999;105(3):696-700.
54. Black AJ, McLeod HL, Capell HA, Powrie RH, Matowe LK, Pritchard SC, et al. Thiopurine methyltransferase genotype predicts therapy-limiting severe toxicity from azathioprine. Annals of internal medicine. 1998;129(9):716-8.
55. Pandya B, Thomson W, Poulton K, Bruce I, Payne D, Qasim F. Azathioprine toxicity and thiopurine methyltransferase genotype in renal transplant patients. Transplant Proc. 2002;34(5):1642-5.
56. Murphy LA, Atherton D. A retrospective evaluation of azathioprine in severe childhood atopic eczema, using thiopurine methyltransferase levels to exclude patients at high risk of myelosuppression. The British journal of dermatology. 2002;147(2):308-15.
57. Holme SA, Duley JA, Sanderson J, Routledge PA, Anstey AV. Erythrocyte thiopurine methyl transferase assessment prior to azathioprine use in the UK. QJM : monthly journal of the Association of Physicians. 2002;95(7):439-44.
58. Phillips KA, Veenstra DL, Oren E, Lee JK, Sadee W. Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. JAMA. 2001;286(18):2270-9.
59. Seidman EG. Clinical use and practical application of TPMT enzyme and 6-mercaptopurine metabolite monitoring in IBD. Reviews in gastroenterological disorders. 2003;3 Suppl 1:S30-8.
60. Marra CA, Esdaile JM, Anis AH. Practical pharmacogenetics: the cost effectiveness of screening for thiopurine s-methyltransferase polymorphisms in patients with rheumatological conditions treated with azathioprine. The Journal of rheumatology. 2002;29(12):2507-12.
61. Tan BB, Lear JT, Gawkrodger DJ, English JS. Azathioprine in dermatology: a survey of current practice in the U.K. The British journal of dermatology. 1997;136(3):351-5.
62. Tavadia SM, Mydlarski PR, Reis MD, Mittmann N, Pinkerton PH, Shear N, et al. Screening for azathioprine toxicity: a pharmacoeconomic analysis based on a target case. Journal of the American Academy of Dermatology. 2000;42(4):628-32.
63. Pirmohamed M BG, Stoddern J, Prince C, Toh CH, Nicholson T, Kesteven P, Jorgensen A, Daly A, Maitland-van der Zee AH, Williamson P, Eriksson N, Avery P, Kamali F, Wadelius M. A Randomized Trial Comparing Genotype-Guided Dosing of Warfarin to Standard Dosing: The EU Pharmacogenetics of Anticoagulant Therapy (EU-PACT) Warfarin Study. 2013.
64. Verhoef TI, Ragia G, de Boer A, Barallon R, Kolovou G, Kolovou V, et al. A randomized trial of genotype-guided dosing of acenocoumarol and phenprocoumon. The New England journal of medicine. 2013;369(24):2304-12.

65. Manolopoulos VG, Ragia G, Tavridou A, Kolovou V, Kolovou G, Maltezos E, et al. Effectiveness of Acenocoumarol Genetic and Clinical Dosing Algorithms in Predicting Stable Dose in the Greek Cohort of the Eu-Pact Trial. Clinical Therapeutics. 2015;37(8):E6-E.
66. van Schie RM, Wadelius MI, Kamali F, Daly AK, Manolopoulos VG, de Boer A, et al. Genotype-guided dosing of coumarin derivatives: the European pharmacogenetics of anticoagulant therapy (EU-PACT) trial design. Pharmacogenomics. 2009;10(10):1687-95.
67. British National Formulary. WARFARIN SODIUM 2016 [Available from: https://bnf.nice.org.uk/drug/warfarin-sodium.html.
68. Pirmohamed M. Warfarin: almost 60 years old and still causing problems. British journal of clinical pharmacology. 2006;62(5):509-11.
69. Burn J, Pirmohamed M. Direct oral anticoagulants versus warfarin: is new always better than the old? Open Heart. 2018;5(1):e000712.
70. Blann AD, Fitzmaurice DA, Lip GY. Anticoagulation in hospitals and general practice. Bmj. 2003;326(7381):153-6.
71. National Institute for Health and Care Excellence (NICE) U. Scenario: Warfarin 2020 [Available from: https://cks.nice.org.uk/topics/anticoagulation-oral/management/warfarin/.
72. Holbrook A, Schulman S, Witt DM, Vandvik PO, Fish J, Kovacs MJ, et al. Evidence-based management of anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141(2 Suppl):e152S-e84S.
73. Holbrook AM, Pereira JA, Labiris R, McDonald H, Douketis JD, Crowther M, et al. Systematic overview of warfarin and its drug and food interactions. Archives of internal medicine. 2005;165(10):1095-106.
74. Wadelius M, Pirmohamed M. Pharmacogenetics of warfarin: current status and future challenges. Pharmacogenomics J. 2007;7(2):99-111.
75. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ. 2004;329(7456):15-9.
76. Jorgensen AL, FitzGerald RJ, Oyee J, Pirmohamed M, Williamson PR. Influence of CYP2C9 and VKORC1 on patient response to warfarin: a systematic review and meta-analysis. PLoS One. 2012;7(8):e44064.
77. Wadelius M, Chen LY, Downes K, Ghori J, Hunt S, Eriksson N, et al. Common VKORC1 and GGCX polymorphisms associated with warfarin dose. Pharmacogenomics J. 2005;5(4):262-70.
78. McClain MR, Palomaki GE, Piper M, Haddow JE. A rapid-ACCE review of CYP2C9 and VKORC1 alleles testing to inform warfarin dosing in adults at elevated risk for thrombotic events to avoid serious bleeding. Genet Med. 2008;10(2):89-98.
79. Van Booven D, Marsh S, McLeod H, Carrillo MW, Sangkuhl K, Klein TE, et al. Cytochrome P450 2C9-CYP2C9. Pharmacogenetics and genomics. 2010;20(4):277-81.
80. Owen RP, Gong L, Sagreiya H, Klein TE, Altman RB. VKORC1 pharmacogenomics summary. Pharmacogenetics and genomics. 2010;20(10):642-4.
81. Sanderson S, Emery J, Higgins J. CYP2C9 gene variants, drug dose, and bleeding risk in warfarin-treated patients: a HuGEnet systematic review and meta-analysis. Genet Med. 2005;7(2):97-104.
82. Zhang J, Chen Z, Chen C. Impact of CYP2C9, VKORC1 and CYP4F2 genetic polymorphisms on maintenance warfarin dosage in Han-Chinese patients: A systematic review and meta-analysis. Meta Gene. 2016;9:197-209.
83. Sconce EA, Khan TI, Wynne HA, Avery P, Monkhouse L, King BP, et al. The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics

upon warfarin dose requirements: proposal for a new dosing regimen. Blood. 2005;106(7):2329-33.
84. James AH, Britt RP, Raskino CL, Thompson SG. Factors affecting the maintenance dose of warfarin. J Clin Pathol. 1992;45(8):704-6.
85. Penning-van Beest FJ, van Meegen E, Rosendaal FR, Stricker BH. Characteristics of anticoagulant therapy and comorbidity related to overanticoagulation. Thromb Haemost. 2001;86(2):569-74.
86. Carlquist JF, Horne BD, Muhlestein JB, Lappe DL, Whiting BM, Kolek MJ, et al. Genotypes of the cytochrome p450 isoform, CYP2C9, and the vitamin K epoxide reductase complex subunit 1 conjointly determine stable warfarin dose: a prospective study. Journal of thrombosis and thrombolysis. 2006;22(3):191-7.
87. Schalekamp T, Boink GJ, Visser LE, Stricker BH, de Boer A, Klungel OH. CYP2C9 genotyping in acenocoumarol treatment: is it a cost-effective addition to international normalized ratio monitoring? Clinical pharmacology and therapeutics. 2006;79(6):511-20.
88. Schalekamp T, Brasse BP, Roijers JF, Chahid Y, van Geest-Daalderop JH, de Vries-Goldschmeding H, et al. VKORC1 and CYP2C9 genotypes and acenocoumarol anticoagulation status: interaction between both genotypes affects overanticoagulation. Clinical pharmacology and therapeutics. 2006;80(1):13-22.
89. Schalekamp T, van Geest-Daalderop JH, Kramer MH, van Holten-Verzantvoort AT, de Boer A. Coumarin anticoagulants and co-trimoxazole: avoid the combination rather than manage the interaction. Eur J Clin Pharmacol. 2007;63(4):335-43.
90. Gage BF, Eby C, Milligan PE, Banet GA, Duncan JR, McLeod HL. Use of pharmacogenetics and clinical factors to predict the maintenance dose of warfarin. Thromb Haemost. 2004;91(1):87-94.
91. Bodin L, Verstuyft C, Tregouet DA, Robert A, Dubert L, Funck-Brentano C, et al. Cytochrome P450 2C9 (CYP2C9) and vitamin K epoxide reductase (VKORC1) genotypes as determinants of acenocoumarol sensitivity. Blood. 2005;106(1):135-40.
92. Wadelius M, Chen LY, Lindh JD, Eriksson N, Ghori MJ, Bumpstead S, et al. The largest prospective warfarin-treated cohort supports genetic forecasting. Blood. 2009;113(4):784-92.
93. D'Andrea G, D'Ambrosio RL, Di Perna P, Chetta M, Santacroce R, Brancaccio V, et al. A polymorphism in the VKORC1 gene is associated with an interindividual variability in the dose-anticoagulant effect of warfarin. Blood. 2005;105(2):645-9.
94. Takeuchi F, McGinnis R, Bourgeois S, Barnes C, Eriksson N, Soranzo N, et al. A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. PLoS genetics. 2009;5(3):e1000433.
95. Caldwell MD, Awad T, Johnson JA, Gage BF, Falkowski M, Gardina P, et al. CYP4F2 genetic variant alters required warfarin dose. Blood. 2008;111(8):4106-12.
96. Schelleman H, Chen J, Chen Z, Christie J, Newcomb CW, Brensinger CM, et al. Dosing algorithms to predict warfarin maintenance dose in Caucasians and African Americans. Clinical pharmacology and therapeutics. 2008;84(3):332-9.
97. International Warfarin Pharmacogenetics C, Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. The New England journal of medicine. 2009;360(8):753-64.
98. Perini JA, Struchiner CJ, Silva-Assuncao E, Santana IS, Rangel F, Ojopi EB, et al. Pharmacogenetics of warfarin: development of a dosing algorithm for brazilian patients. Clinical pharmacology and therapeutics. 2008;84(6):722-8.
99. Tham LS, Goh BC, Nafziger A, Guo JY, Wang LZ, Soong R, et al. A warfarin-dosing model in Asians that uses single-nucleotide polymorphisms in

vitamin K epoxide reductase complex and cytochrome P450 2C9. Clinical pharmacology and therapeutics. 2006;80(4):346-55.
100. Gage BF, Eby C, Johnson JA, Deych E, Rieder MJ, Ridker PM, et al. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. Clinical pharmacology and therapeutics. 2008;84(3):326-31.
101. Hylek EM, Skates SJ, Sheehan MA, Singer DE. An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with nonrheumatic atrial fibrillation. The New England journal of medicine. 1996;335(8):540-6.
102. Penning-van Beest FJ, Geleijnse JM, van Meegen E, Vermeer C, Rosendaal FR, Stricker BH. Lifestyle and diet as risk factors for overanticoagulation. Journal of clinical epidemiology. 2002;55(4):411-7.
103. Schalekamp T, Klungel OH, Souverein PC, de Boer A. Increased bleeding risk with concurrent use of selective serotonin reuptake inhibitors and coumarins. Archives of internal medicine. 2008;168(2):180-5.
104. Reitsma PH, van der Heijden JF, Groot AP, Rosendaal FR, Buller HR. A C1173T dimorphism in the VKORC1 gene determines coumarin sensitivity and bleeding risk. PLoS Med. 2005;2(10):e312.
105. Rosendaal FR. The Scylla and Charybdis of oral anticoagulant treatment. The New England journal of medicine. 1996;335(8):587-9.
106. Eckman MH, Rosand J, Greenberg SM, Gage BF. Cost-effectiveness of using pharmacogenetic information in warfarin dosing for patients with nonvalvular atrial fibrillation. Annals of internal medicine. 2009;150(2):73-83.
107. Hughes DA, Pirmohamed M. Warfarin pharmacogenetics: economic considerations. Pharmacoeconomics. 2007;25(11):899-902.
108. Anderson JL, Horne BD, Stevens SM, Grove AS, Barton S, Nicholas ZP, et al. Randomized trial of genotype-guided versus standard warfarin dosing in patients initiating oral anticoagulation. Circulation. 2007;116(22):2563-70.
109. Caraco Y, Blotnick S, Muszkat M. CYP2C9 genotype-guided warfarin prescribing enhances the efficacy and safety of anticoagulation: a prospective randomized controlled study. Clinical pharmacology and therapeutics. 2008;83(3):460-70.
110. Daly AK. Optimal dosing of warfarin and other coumarin anticoagulants: the role of genetic polymorphisms. Archives of toxicology. 2013;87(3):407-20.
111. Tarcic G, Kamal M, Edelheit O, Barbash Z, Vidne M, Miron B, et al. Functional mutational analysis to assess the oncogenic activity of variant of uncertain significance (VUS) detected in patients included in the SHIVA trial. European Journal of Cancer. 2016;69:S6-S7.
112. Kamal M, Servant N, Pierron G, Callens C, Gentien D, Lermine A, et al. Abstract 1524: Mutations and gene copy number variations landscape of metastases of various cancer types from patients enrolled in the SHIVA trial. Cancer Research. 2016;76(14 Supplement):1524-.
113. Lacombe D, Burock S, Bogaerts J, Schoeffski P, Golfinopoulos V, Stupp R. The dream and reality of histology agnostic cancer clinical trials. Molecular oncology. 2014;8(6):1057-63.
114. Thatcher N, Chang A, Parikh P, Rodrigues Pereira J, Ciuleanu T, von Pawel J, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). Lancet. 2005;366(9496):1527-37.
115. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. The New England journal of medicine. 2009;361(10):947-57.
116. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic

breast cancer that overexpresses HER2. The New England journal of medicine. 2001;344(11):783-92.
117. Kim ES, Herbst RS, Wistuba, II, Lee JJ, Blumenschein GR, Jr., Tsao A, et al. The BATTLE trial: personalizing therapy for lung cancer. Cancer discovery. 2011;1(1):44-53.
118. Herbst RS GD, Hirsch FR, Redman MW, LeBlanc M, Mack PC, Schwartz LH, Vokes E, Ramalingam S, Bradley JD, Sparks D. Lung Master Protocol (Lung-MAP)-A Biomarker-Driven Protocol for Accelerating Development of Therapies for Squamous Cell Lung Cancer: SWOG S1400. Clinical cancer research : an official journal of the American Association for Cancer Research. 2015.
119. Lievre A, Bachet JB, Boige V, Cayre A, Le Corre D, Buc E, et al. KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(3):374-9.
120. Von Hoff DD, Stephenson JJ, Jr., Rosen P, Loesch DM, Borad MJ, Anthony S, et al. Pilot study using molecular profiling of patients' tumors to find potential targets and select treatments for their refractory cancers. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(33):4877-83.
121. Doroshow JH. Selecting systemic cancer therapy one patient at a time: is there a role for molecular profiling of individual patients with advanced solid tumors? Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(33):4869-71.
122. Sosman JA, Kim KB, Schuchter L, Gonzalez R, Pavlick AC, Weber JS, et al. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. The New England journal of medicine. 2012;366(8):707-14.
123. Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. The New England journal of medicine. 2001;344(14):1031-7.
124. Sekulic A, Migden MR, Oro AE, Dirix L, Lewis KD, Hainsworth JD, et al. Efficacy and safety of vismodegib in advanced basal-cell carcinoma. The New England journal of medicine. 2012;366(23):2171-9.
125. Tsimberidou AM, Iskander NG, Hong DS, Wheler JJ, Falchook GS, Fu S, et al. Personalized medicine in a phase I clinical trials program: the MD Anderson Cancer Center initiative. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012;18(22):6373-83.
126. Shaw AT, Kim DW, Nakagawa K, Seto T, Crino L, Ahn MJ, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. The New England journal of medicine. 2013;368(25):2385-94.
127. Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M, et al. Adjuvant trastuzumab in HER2-positive breast cancer. The New England journal of medicine. 2011;365(14):1273-83.
128. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. The New England journal of medicine. 2010;362(25):2380-8.
129. Hirsh BJ, Smilowitz NR, Rosenson RS, Fuster V, Sperling LS. Utilization of and Adherence to Guideline-Recommended Lipid-Lowering Therapy After Acute Coronary Syndrome: Opportunities for Improvement. J Am Coll Cardiol. 2015;66(2):184-92.
130. Gomez Sandoval YH, Braganza MV, Daskalopoulou SS. Statin discontinuation in high-risk patients: a systematic review of the evidence. Curr Pharm Des. 2011;17(33):3669-89.
131. Rodriguez F, Maron DJ, Knowles JW, Virani SS, Lin S, Heidenreich PA. Association of Statin Adherence With Mortality in Patients With Atherosclerotic Cardiovascular Disease. JAMA Cardiol. 2019;4(3):206-13.

132. Vonbank A, Drexel H, Agewall S, Lewis BS, Dopheide JF, Kjeldsen K, et al. Reasons for disparity in statin adherence rates between clinical trials and real-world observations: a review. Eur Heart J Cardiovasc Pharmacother. 2018;4(4):230-6.
133. Serruys PW, de Feyter P, Macaya C, Kokott N, Puel J, Vrolix M, et al. Fluvastatin for prevention of cardiac events following successful first percutaneous coronary intervention: a randomized controlled trial. JAMA. 2002;287(24):3215-22.
134. Slejko JF, Ho M, Anderson HD, Nair KV, Sullivan PW, Campbell JD. Adherence to statins in primary prevention: yearly adherence changes and outcomes. J Manag Care Pharm. 2014;20(1):51-7.
135. Kern DM, Balu S, Tunceli O, Anzalone D. Statin treatment patterns and clinical profile of patients with risk factors for coronary heart disease defined by National Cholesterol Education Program Adult Treatment Panel III. Curr Med Res Opin. 2014;30(12):2443-51.
136. Ramsey LB, Johnson SG, Caudle KE, Haidar CE, Voora D, Wilke RA, et al. The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update. Clinical pharmacology and therapeutics. 2014;96(4):423-8.
137. Wilke RA, Ramsey LB, Johnson SG, Maxwell WD, McLeod HL, Voora D, et al. The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. Clinical pharmacology and therapeutics. 2012;92(1):112-7.
138. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2014;63(25 Pt B):2889-934.
139. Pasternak RC, Smith SC, Bairey-Merz CN, Grundy SM, Cleeman JI, Lenfant C. ACC/AHA/NHLBI clinical advisory on the use and safety of statins. Journal of the American College of Cardiology. 2002;40(3):567-72.
140. Stroes ES, Thompson PD, Corsini A, Vladutiu GD, Raal FJ, Ray KK, et al. Statin-associated muscle symptoms: impact on statin therapy-European Atherosclerosis Society Consensus Panel Statement on Assessment, Aetiology and Management. Eur Heart J. 2015;36(17):1012-22.
141. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Executive Summary: Heart Disease and Stroke Statistics—2015 Update. Circulation. 2015;131(4):434-41.
142. Alfirevic A, Neely D, Armitage J, Chinoy H, Cooper RG, Laaksonen R, et al. Phenotype standardization for statin-induced myotoxicity. Clinical pharmacology and therapeutics. 2014;96(4):470-6.
143. Ong FS, Deignan JL, Kuo JZ, Bernstein KE, Rotter JI, Grody WW, et al. Clinical utility of pharmacogenetic biomarkers in cardiovascular therapeutics: a challenge for clinical implementation. Pharmacogenomics. 2012;13(4):465-75.
144. Voora D, Ginsburg GS. Clinical application of cardiovascular pharmacogenetics. J Am Coll Cardiol. 2012;60(1):9-20.
145. Patel J, Superko HR, Martin SS, Blumenthal RS, Christopher-Stine L. Genetic and immunologic susceptibility to statin-related myopathy. Atherosclerosis. 2015;240(1):260-71.
146. Osterberg L, Blaschke T. Adherence to medication. The New England journal of medicine. 2005;353(5):487-97.
147. Niemi M, Pasanen MK, Neuvonen PJ. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. Pharmacol Rev. 2011;63(1):157-81.
148. Greenland P, Lauer MS. Cholesterol Lowering in 2015: Still Answering Questions About How and in Whom. JAMA. 2015;314(2):127-8.

149. Thompson PD, Clarkson PM, Rosenson RS, National Lipid Association Statin Safety Task Force Muscle Safety Expert P. An assessment of statin safety by muscle experts. Am J Cardiol. 2006;97(8A):69C-76C.
150. Group SC, Link E, Parish S, Armitage J, Bowman L, Heath S, et al. SLCO1B1 variants and statin-induced myopathy--a genomewide study. The New England journal of medicine. 2008;359(8):789-99.
151. Pencina MJ, Navar-Boggan AM, D'Agostino RB, Sr., Williams K, Neely B, Sniderman AD, et al. Application of new cholesterol guidelines to a population-based sample. The New England journal of medicine. 2014;370(15):1422-31.
152. Bermingham M, Hayden J, Dawkins I, Miwa S, Gibson D, McDonald K, et al. Prospective analysis of LDL-C goal achievement and self-reported medication adherence among statin users in primary care. Clin Ther. 2011;33(9):1180-9.
153. Ho PM, Magid DJ, Shetterly SM, Olson KL, Maddox TM, Peterson PN, et al. Medication nonadherence is associated with a broad range of adverse outcomes in patients with coronary artery disease. Am Heart J. 2008;155(4):772-9.
154. Vodonos A, Ostapenko I, Toledano R, Henkin Y, Zahger D, Wolak T, et al. Statin adherence and LDL cholesterol levels. Should we assess adherence prior to statin upgrade? European journal of internal medicine. 2015;26(4):268-72.
155. Franklin JM, Krumme AA, Tong AY, Shrank WH, Matlin OS, Brennan TA, et al. Association between trajectories of statin adherence and subsequent cardiovascular events. Pharmacoepidemiol Drug Saf. 2015;24(10):1105-13.
156. Pasanen MK, Backman JT, Neuvonen PJ, Niemi M. Frequencies of single nucleotide polymorphisms and haplotypes of organic anion transporting polypeptide 1B1 SLCO1B1 gene in a Finnish population. Eur J Clin Pharmacol. 2006;62(6):409-15.
157. Li JH, Joy SV, Haga SB, Orlando LA, Kraus WE, Ginsburg GS, et al. Genetically guided statin therapy on statin perceptions, adherence, and cholesterol lowering: a pilot implementation study in primary care patients. J Pers Med. 2014;4(2):147-62.
158. Donnelly LA, Doney AS, Tavendale R, Lang CC, Pearson ER, Colhoun HM, et al. Common nonsynonymous substitutions in SLCO1B1 predispose to statin intolerance in routinely treated individuals with type 2 diabetes: a go-DARTS study. Clinical pharmacology and therapeutics. 2011;89(2):210-6.
159. Birmingham BK, Bujac SR, Elsby R, Azumaya CT, Wei C, Chen Y, et al. Impact of ABCG2 and SLCO1B1 polymorphisms on pharmacokinetics of rosuvastatin, atorvastatin and simvastatin acid in Caucasian and Asian subjects: a class effect? Eur J Clin Pharmacol. 2015;71(3):341-55.
160. de Keyser CE, Peters BJ, Becker ML, Visser LE, Uitterlinden AG, Klungel OH, et al. The SLCO1B1 c. 521T> C polymorphism is associated with dose decrease or switching during statin therapy in the Rotterdam Study. Pharmacogenetics and genomics. 2014;24(1):43-51.
161. Voora D, Shah SH, Spasojevic I, Ali S, Reed CR, Salisbury BA, et al. The SLCO1B1*5 genetic variant is associated with statin-induced side effects. J Am Coll Cardiol. 2009;54(17):1609-16.
162. Danik JS, Chasman DI, MacFadyen JG, Nyberg F, Barratt BJ, Ridker PM. Lack of association between SLCO1B1 polymorphisms and clinical myalgia following rosuvastatin therapy. Am Heart J. 2013;165(6):1008-14.
163. Martin NG, Li KW, Murray H, Putt W, Packard CJ, Humphries SE. The effects of a single nucleotide polymorphism in SLCO1B1 on the pharmacodynamics of pravastatin. British journal of clinical pharmacology. 2012;73(2):303-6.
164. Taylor F, Huffman MD, Macedo AF, Moore TH, Burke M, Davey Smith G, et al. Statins for the primary prevention of cardiovascular disease. The Cochrane database of systematic reviews. 2013(1):CD004816.

165. De Vera MA, Bhole V, Burns LC, Lacaille D. Impact of statin adherence on cardiovascular disease and mortality outcomes: a systematic review. British journal of clinical pharmacology. 2014;78(4):684-98.
166. Cholesterol Treatment Trialists Collaborators. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. The Lancet. 2012;380(9841):581-90.
167. Cholesterol Treatment Trialists Collaborators. Efficacy and safety of LDL-lowering therapy among men and women: meta-analysis of individual data from 174 000 participants in 27 randomised trials. The Lancet. 2015;385(9976):1397-405.
168. Hou Q, Li S, Li L, Li Y, Sun X, Tian H. Association Between SLCO1B1 Gene T521C Polymorphism and Statin-Related Myopathy Risk: A Meta-Analysis of Case-Control Studies. Medicine. 2015;94(37):e1268.
169. Peterson AM, Nau DP, Cramer JA, Benner J, Gwadry-Sridhar F, Nichol M. A checklist for medication compliance and persistence studies using retrospective databases. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2007;10(1):3-12.
170. Cholesterol Treatment Trialists Collaborators. CTT Collaboration 2019 [Available from: https://www.cttcollaboration.org/.
171. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol. 2010;25(9):603-5.
172. Voils CI, Hoyle RH, Thorpe CT, Maciejewski ML, Yancy WS, Jr. Improving the measurement of self-reported medication nonadherence. Journal of clinical epidemiology. 2011;64(3):250-4.
173. Oshiro C, Mangravite L, Klein T, Altman R. PharmGKB very important pharmacogene: SLCO1B1. Pharmacogenetics and genomics. 2010;20(3):211-6.
174. Teh LK, Bertilsson L. Pharmacogenomics of CYP2D6: molecular genetics, interethnic differences and clinical importance. Drug Metab Pharmacokinet. 2012;27(1):55-67.
175. Kroemer HK, Eichelbaum M. "It's the genes, stupid". Molecular bases and clinical consequences of genetic cytochrome P450 2D6 polymorphism. Life Sci. 1995;56(26):2285-98.
176. National Center for Biotechnology Information, ClinVar. Search results for variants in the CYP2D6 gene 2020 [Available from: https://www.ncbi.nlm.nih.gov/clinvar/?term=CYP2D6[gene].
177. Kirchheiner J, Keulen JT, Bauer S, Roots I, Brockmoller J. Effects of the CYP2D6 gene duplication on the pharmacokinetics and pharmacodynamics of tramadol. J Clin Psychopharmacol. 2008;28(1):78-83.
178. Zackrisson AL, Lindblom B, Ahlner J. High frequency of occurrence of CYP2D6 gene duplication/multiduplication indicating ultrarapid metabolism among suicide cases. Clinical pharmacology and therapeutics. 2010;88(3):354-9.
179. Dean L. Codeine Therapy and CYP2D6 Genotype. In: Pratt VM, Scott SA, Pirmohamed M, Esquivel B, Kane MS, Kattman BL, et al., editors. Medical Genetics Summaries. Bethesda (MD): National Center for Biotechnology Information (US); 2012.
180. Smith DM, Weitzel KW, Elsey AR, Langaee T, Gong Y, Wake DT, et al. CYP2D6-guided opioid therapy improves pain control in CYP2D6 intermediate and poor metabolizers: a pragmatic clinical trial. Genet Med. 2019;21(8):1842-50.
181. Laika B, Leucht S, Heres S, Steimer W. Intermediate metabolizer: increased side effects in psychoactive drug therapy. The key to cost-effectiveness of pretreatment CYP2D6 screening? Pharmacogenomics J. 2009;9(6):395-403.

182. US National Library of Medicine, ClinicalTrials.gov. Precision Medicine Guided Treatment for Cancer Pain 2019 [Available from: https://clinicaltrials.gov/ct2/show/study/NCT02664350.
183. Mendoza T, Mayne T, Rublee D, Cleeland C. Reliability and validity of a modified Brief Pain Inventory short form in patients with osteoarthritis. Eur J Pain. 2006;10(4):353-61.
184. Cavallari LH. Personal Communication. RE: Precision Medicine Guided Treatment for Cancer Pain. 25th September 2020.
185. Eckhardt K, Li S, Ammon S, Schanzle G, Mikus G, Eichelbaum M. Same incidence of adverse drug events after codeine administration irrespective of the genetically determined differences in morphine formation. Pain. 1998;76(1-2):27-33.
186. Angst MS, Phillips NG, Drover DR, Tingle M, Ray A, Swan GE, et al. Pain sensitivity and opioid analgesia: a pharmacogenomic twin study. Pain. 2012;153(7):1397-409.
187. Lotsch J, Rohrbacher M, Schmidt H, Doehring A, Brockmoller J, Geisslinger G. Can extremely low or high morphine formation from codeine be predicted prior to therapy initiation? Pain. 2009;144(1-2):119-24.
188. Samer CF, Daali Y, Wagner M, Hopfgartner G, Eap CB, Rebsamen MC, et al. The effects of CYP2D6 and CYP3A activities on the pharmacokinetics of immediate release oxycodone. Br J Pharmacol. 2010;160(4):907-18.
189. Zwisler ST, Enggaard TP, Noehr-Jensen L, Pedersen RS, Mikkelsen S, Nielsen F, et al. The hypoalgesic effect of oxycodone in human experimental pain models in relation to the CYP2D6 oxidation polymorphism. Basic Clin Pharmacol Toxicol. 2009;104(4):335-44.
190. Temel JS, Greer JA, Muzikansky A, Gallagher ER, Admane S, Jackson VA, et al. Early palliative care for patients with metastatic non-small-cell lung cancer. The New England journal of medicine. 2010;363(8):733-42.
191. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: Adult cancer pain. Version 2.2016 https://www.nccn.org/professionals/physician_gls/default.aspx2016 [
192. Susce MT, Murray-Carmichael E, de Leon J. Response to hydrocodone, codeine and oxycodone in a CYP2D6 poor metabolizer. Prog Neuropsychopharmacol Biol Psychiatry. 2006;30(7):1356-8.
193. Ciszkowski C, Madadi P, Phillips MS, Lauwers AE, Koren G. Codeine, ultrarapid-metabolism genotype, and postoperative death. The New England journal of medicine. 2009;361(8):827-8.
194. Gasche Y, Daali Y, Fathi M, Chiappe A, Cottini S, Dayer P, et al. Codeine intoxication associated with ultrarapid CYP2D6 metabolism. The New England journal of medicine. 2004;351(27):2827-31.
195. Owen RP, Sangkuhl K, Klein TE, Altman RB. Cytochrome P450 2D6. Pharmacogenetics and genomics. 2009;19(7):559-62.
196. Amur S, LaVange L, Zineh I, Buckman-Garner S, Woodcock J. Biomarker Qualification: Toward a Multiple Stakeholder Framework for Biomarker Development, Regulatory Acceptance, and Utilization. Clinical pharmacology and therapeutics. 2015;98(1):34-46.
197. Altar CA, Amakye D, Bounos D, Bloom J, Clack G, Dean R, et al. A prototypical process for creating evidentiary standards for biomarkers and diagnostics. Clinical pharmacology and therapeutics. 2008;83(2):368-71.
198. Thorn CF, Klein TE, Altman RB. PharmGKB: the Pharmacogenomics Knowledge Base. Methods in molecular biology (Clifton, NJ). 2013;1015:311-20.
199. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. The New England journal of medicine. 2000;342(25):1887-92.

200. Thomas P. Guide to Clinical Preventive Services: Report of the US Preventive Services Task Force. JAMA: The Journal of the American Medical Association. 1996;276(11):923-4.
201. Rawlins M. De Testimonio: on the evidence for decisions about the use of therapeutic interventions. Clinical medicine (London, England). 2008;8(6):579-88.
202. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016;21(4):125-7.
203. Chan AW, Hrobjartsson A. Promoting public access to clinical trial protocols: challenges and recommendations. Trials. 2018;19(1):116.
204. Chan AW, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ. 2013;346:e7586.
205. Pirmohamed M, Hughes DA. Pharmacogenetic tests: the need for a level playing field. Nature reviews Drug discovery. 2013;12(1):3-4.
206. Gammie T, Lu CY, Babar ZU. Access to Orphan Drugs: A Comprehensive Review of Legislations, Regulations and Policies in 35 Countries. PLoS One. 2015;10(10):e0140002.
207. Hughes DA, Plumpton CO. Rare disease prevention and treatment: the need for a level playing field. Pharmacogenomics. 2018;19(3):243-7.
208. Huddart R, Sangkuhl K, Whirl-Carrillo M, Klein TE. Are Randomized Controlled Trials Necessary to Establish the Value of Implementing Pharmacogenomics in the Clinic? Clinical pharmacology and therapeutics. 2019;106(2):284-6.
209. Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. Statistics in medicine. 2014;33(24):4186-201.
210. Green MJ, Botkin JR. "Genetic exceptionalism" in medicine: clarifying the differences between genetic and nongenetic tests. Annals of internal medicine. 2003;138(7):571-5.
211. Pirmohamed M. Acceptance of biomarker-based tests for application in clinical practice: criteria and obstacles. Clinical pharmacology and therapeutics. 2010;88(6):862-6.
212. Clarke M, Hopewell S, Chalmers I. Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. The Lancet. 2010;376(9734):20-1.
213. Chalmers I, Fox DM. Increasing the Incidence and Influence of Systematic Reviews on Health Policy and Practice. Am J Public Health. 2016;106(1):11-3.
214. The Lancet. Information for Authors 2018 [Available from: https://els-jbs-prod-cdn.literatumonline.com/pb/assets/raw/Lancet/authors/tl-info-for-authors-1530878364923.pdf.
215. The Lancet. Randomised trials in The Lancet: formatting guidelines 2015 [Available from: https://els-jbs-prod-cdn.jbs.elsevierhealth.com/pb/assets/raw/Lancet/authors/RCTguidelines.pdf.
216. PLOS One. Submission Guidelines [Available from: https://journals.plos.org/plosone/s/submission-guidelines#loc-guidelines-for-specific-study-types.
217. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. Journal of clinical epidemiology. 2010;63(8):e1-37.
218. The New England Journal of Medicine. New Manuscripts [Available from: https://www.nejm.org/author-center/new-manuscripts.
219. Annals of Internal Medicine. Information for Authors [Available from: https://www.acpjournals.org/journal/aim/authors2.

220. BMJ. BMJ Guidance for Authors 2018 [Available from: https://drive.google.com/file/d/1lh9zzpFUUoOuLVszt4kB8jMv097nfJTn/view.
221. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014;348:g1687.
222. JAMA Network. Instructions for Authors 2020 [Available from: https://jamanetwork.com/journals/jama/pages/instructions-for-authors.
223. Nature Research. Ethics and biosecurity [Available from: https://www.nature.com/nature-research/editorial-policies/ethics-and-biosecurity#bioethics-policy.
224. Neurology Journals. Research Policies and Guidelines [Available from: https://www.neurology.org/research-policies-and-guidelines.
225. Proceedings of the National Academy of Sciences of the United States of America. PNAS Information for Authors 2020 [Available from: http://blog.pnas.org/iforc.pdf.
226. Science. Science Journals: editorial policies [Available from: https://www.sciencemag.org/authors/science-journals-editorial-policies.
227. Jemielniak D, Masukume G, Wilamowski M. The Most Influential Medical Journals According to Wikipedia: Quantitative Analysis. J Med Internet Res. 2019;21(1):e11429.
228. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002.
229. Genin E, Chen DP, Hung SI, Sekula P, Schumacher M, Chang PY, et al. HLA-A*31:01 and different types of carbamazepine-induced severe cutaneous adverse reactions: an international study and meta-analysis. Pharmacogenomics J. 2014;14(3):281-8.
230. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
231. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-6.
232. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Annals of internal medicine. 2007;147(8):573-7.
233. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.
234. Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J Evid Based Med. 2015;8(1):2-10.
235. Wang B, Canestaro WJ, Choudhry NK. Clinical evidence supporting pharmacogenomic biomarker testing provided in US Food and Drug Administration drug labels. JAMA internal medicine. 2014;174(12):1938-44.
236. Ward J, Graham N, Strawbridge RJ, Ferguson A, Jenkins G, Chen W, et al. Polygenic risk scores for major depressive disorder and neuroticism as predictors of antidepressant response: Meta-analysis of three treatment cohorts. PloS one [Internet]. 2018 2018; 13(9):[e0203896 p.]. Available from: http://europepmc.org/abstract/MED/30240446

https://doi.org/10.1371/journal.pone.0203896

https://europepmc.org/articles/PMC6150505

https://europepmc.org/articles/PMC6150505?pdf=render.
237. Tantipong H, Morkchareonpong C, Jaiyindee S, Thamlikitkul V. Randomized controlled trial and meta-analysis of oral decontamination with 2% chlorhexidine solution for the prevention of ventilator-associated pneumonia. Infect Control Hosp Epidemiol. 2008;29(2):131-6.
238. Ellis P, Barrett-Lee P, Johnson L, Cameron D, Wardley A, O'Reilly S, et al. Sequential docetaxel as adjuvant chemotherapy for early breast cancer (TACT): an open-label, phase III, randomised controlled trial. The Lancet. 2009;373(9676):1681-92.
239. Clinical Pharmacogenetics Implementation Consortium. Levels of Evidence 2016 [Available from: https://cpicpgx.org/levels-of-evidence/.
240. Barbarino JM, Whirl-Carrillo M, Altman RB, Klein TE. PharmGKB: A worldwide resource for pharmacogenomic information. Wiley Interdiscip Rev Syst Biol Med. 2018;10(4):e1417.
241. Lever J, Barbarino JM, Gong L, Huddart R, Sangkuhl K, Whaley R, et al., editors. PGxMine: Text mining for curation of PharmGKB. Pacific Symposium on Biocomputing; 2020.
242. Bentley AR, Callier SL, Rotimi CN. Evaluating the promise of inclusion of African ancestry populations in genomics. NPJ Genom Med. 2020;5(1):5.
243. Rotimi CN, Jorde LB. Ancestry and disease in the age of genomic medicine. The New England journal of medicine. 2010;363(16):1551-8.
244. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584-91.
245. Oh SS, White MJ, Gignoux CR, Burchard EG. Making Precision Medicine Socially Precise. Take a Deep Breath. Am J Respir Crit Care Med. 2016;193(4):348-50.
246. Johnson D, Wilke MA, Lyle SM, Kowalec K, Jorgensen A, Wright GE, et al. A systematic review and analysis of the use of polygenic scores in pharmacogenomics. Clinical pharmacology and therapeutics. 2021;n/a(n/a).
247. Bentley AR, Callier S, Rotimi CN. Diversity and inclusion in genomic research: why the uneven progress? J Community Genet. 2017;8(4):255-66.
248. Shin J, Cao D. Comparison of warfarin pharmacogenetic dosing algorithms in a racially diverse large cohort. Pharmacogenomics. 2011;12(1):125-34.
249. Drozda K, Wong S, Patel SR, Bress AP, Nutescu EA, Kittles RA, et al. Poor warfarin dose prediction with pharmacogenetic algorithms that exclude genotypes important for African Americans. Pharmacogenetics and genomics. 2015;25(2):73-81.
250. Tan GM, Wu E, Lam YY, Yan BP. Role of warfarin pharmacogenetic testing in clinical practice. Pharmacogenomics. 2010;11(3):439-48.
251. Kudzi W, Ahorhorlu SY, Dzudzor B, Olayemi E, Nartey ET, Asmah RH. Genetic polymorphisms of patients on stable warfarin maintenance therapy in a Ghanaian population. BMC Res Notes. 2016;9(1):507.
252. PharmGKB. Warfarin 2020 [Available from: https://www.pharmgkb.org/chemical/PA451906.
253. Claudio-Campos K, Labastida A, Ramos A, Gaedigk A, Renta-Torres J, Padilla D, et al. Warfarin Anticoagulation Therapy in Caribbean Hispanics of Puerto Rico: A Candidate Gene Association Study. Frontiers in pharmacology. 2017;8:347.
254. US National Library of Medicine, Clinicaltrials.gov. Genetic Testing to Understand and Address Renal Disease Disparities Across the United States (GUARDD-US) 2020 [Available from: https://clinicaltrials.gov/ct2/show/study/NCT04191824?titles=GUARDD-US&cntry=US&draw=2&rank=1.

255. Tindana P, Yakubu A, Staunton C, Matimba A, Littler K, Madden E, et al. Engaging research ethics committees to develop an ethics and governance framework for best practices in genomic research and biobanking in Africa: the H3Africa model. BMC Med Ethics. 2019;20(1):69.
256. Asiimwe IG, Waitt C, Sekaggya-Wiltshire C, Hutchinson C, Okello E, Zhang EJ, et al. Developing and Validating a Clinical Warfarin Dose-Initiation Model for Black-African Patients in South Africa and Uganda. Clinical pharmacology and therapeutics. 2021;109(6):1564-74.
257. Antoniou M, Kolamunnage-Dona R, Wason J, Bathia R, Billingham C, Bliss JM, et al. Biomarker-guided trials: Challenges in practice. Contemp Clin Trials Commun. 2019;16:100493.
258. Rigter T, Jansen ME, de Groot JM, Janssen SWJ, Rodenburg W, Cornel MC. Implementation of Pharmacogenetics in Primary Care: A Multi-Stakeholder Perspective. Front Genet. 2020;11:10.
259. Slob EMA, Vijverberg SJH, Pijnenburg MW, Koppelman GH, Maitland-van der Zee AH. What do we need to transfer pharmacogenetics findings into the clinic? Pharmacogenomics. 2018;19(7):589-92.
260. Hey SP, London AJ, Weijer C, Rid A, Miller F. Is the concept of clinical equipoise still relevant to research? BMJ. 2017;359:j5787.
261. Stewart DJ, Whitney SN, Kurzrock R. Equipoise lost: ethics, costs, and the regulation of cancer clinical research. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(17):2925-35.
262. Miller FG, Brody H. A critique of clinical equipoise. Therapeutic misconception in the ethics of clinical trials. Hastings Cent Rep. 2003;33(3):19-28.
263. London AJ. Two dogmas of research ethics and the integrative approach to human-subjects research. J Med Philos. 2007;32(2):99-116.

Chapter 4: Systematic reviews and a simulated prospective trial

4.1 Introduction

4.1.1 SJS/TEN, HLA genotypes, and carbamazepine

Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are cutaneous adverse drug reactions (cADRs) that consist of rash, fever, and epidermal detachment from the dermis (1-6). SJS is the diagnosis when less than 10% of the body surface area is affected, TEN when more than 30% is affected (1, 7, 8). Within this range, the condition is known as SJS/TEN. Mortality for these conditions is high, ranging from 1-5% in SJS up to 25-40% in TEN, with a high risk of associated morbidities (1, 5, 6, 9, 10). Around 30% of patients with severe reactions die (11, 12). Related hypersensitivity reactions include drug reaction with eosinophilia and systemic symptoms (DRESS) and maculopapular exanthema (MPE) (2, 13-15).

These reactions are associated with many drugs, the main causative agents being allopurinol (16, 17), lamotrigine (15, 18), and carbamazepine. The link between carbamazepine and SJS/TEN was first quantified in a 1995 international case-control study (19). Over a 4-year period, 245 cases of SJS/TEN were recruited and compared to 1147 controls (patients admitted to hospital for acute conditions or a procedure not related to medication use). The relative risk of SJS/TEN in patients exposed to carbamazepine compared to control patients was calculated to be 12 (95% CI 3.5 – 38), with the highest risk within the first 2 months of treatment (19). A further review of Canadian health records in 1997 found a risk of 6.2 per 10,000 new users of carbamazepine (95% CI 2.5 – 14.1 per 10,000) (21). A 2015 analysis in a UK setting (based on UK and wider Northern European datasets) estimated a risk of SJS/TEN of 1.18 per 10,000 carbamazepine patients (24).

In 2004, the first link between carbamazepine-induced SJS/TEN and the *HLA-B*15:02* allele was quantified (27). *HLA-B*15:02* is a human leukocyte antigen (HLA) of the class I major histocompatibility complex (MHC) gene cluster (28). MHC class I molecules are expressed in all nucleated cells (29). They present intracellular proteins to T-cells via CD8 for antigen processing, and effector cell activation (29-31) (Figure 4.1A). This can be useful for pathogen and abnormal cell detection, but can be deleterious, causing autoimmune and hypersensitivity

reactions (29, 31, 32). The antigen presented by cells in response to carbamazepine is unknown (33, 34). However, cross-sensitivity to other drugs in the same class has been reported (33, 35), suggesting a common mechanism (33).

HLA alleles have a designated syntax to describe their specific types (Figure 4.1B).

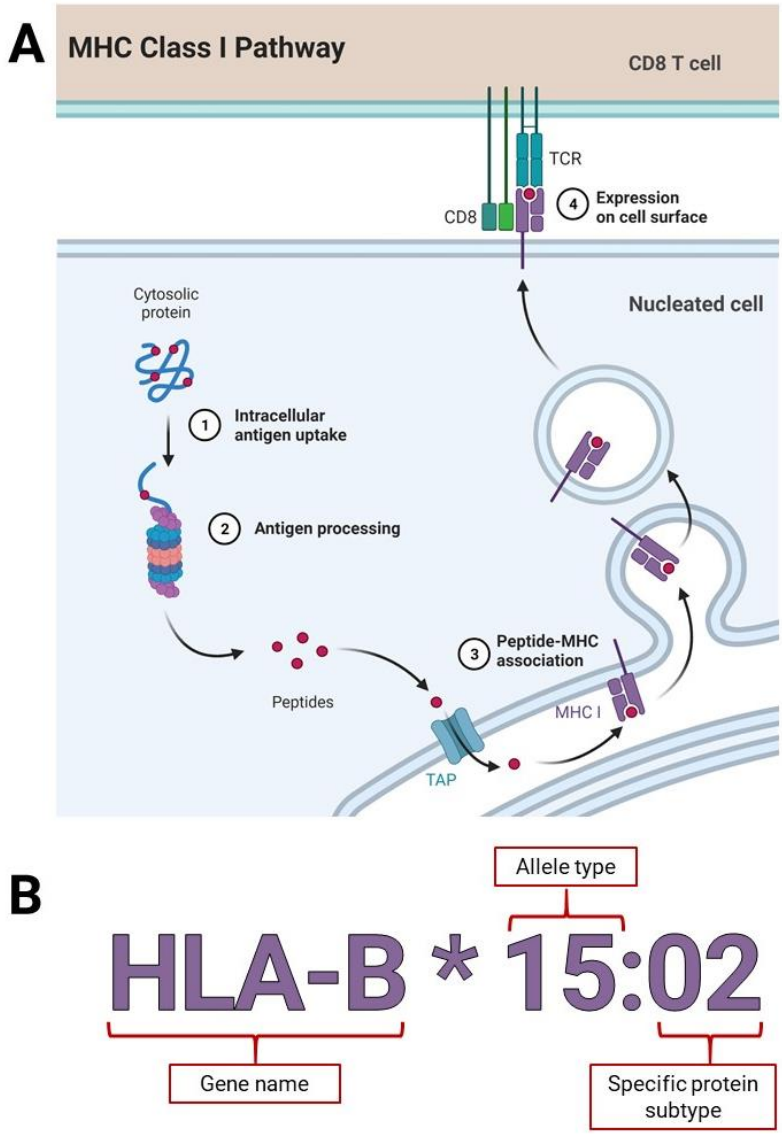


Figure 4.1 - HLA explainer. A) The process of antigen presentation by the MHC Class I molecule. Antigens from intracellular sources are taken up and processed into peptides. They then enter the endoplasmic reticulum through the TAP and are loaded on to MHC class I molecules. These molecules migrate to the plasma membrane and present the antigen to CD8+ T-cells via the TCR (29, 30, 32, 36, 37). HLA = human leukocyte antigen. MHC = major histocompatibility complex. TAP = transporter associated with antigen presentation. TCR = T-cell receptor. Created with BioRender.com.

As of 2020, >6000 HLA-A and >7000 HLA-B alleles have been identified (38). While many of these alleles have been linked with disease or ADRs, the link between *HLA-B*15:02* and SJS/TEN is one of the strongest associations seen between an allele and an ADR (2, 3, 39). Presence of the *HLA-B*15:02* allele is associated with carbamazepine-induced SJS/TEN and odds ratios of between 47.67 (95% CI 2.55 – 890.45) (18) and 1357.00 (159.84 – 11520.40) (40) have been found in various populations (2).

4.1.1.1 Chen et al 2011 prospective interventional study

The benefit of screening for *HLA-B*15:02* in an Asian population was proven with a 2011 prospective study in Taiwan (3). A total of 4877 patients requiring carbamazepine were recruited and genotyped for *HLA-B*15:02*. Of these, 372 (7.7%) tested positive for the allele and were advised not to take carbamazepine and to instead take an alternative medication. Those testing negative continued with their carbamazepine prescription. The authors calculated, based on the historical incidence of SJS/TEN in Taiwan, that in the absence of genetic testing 10 cases of SJS/TEN would have been expected in a group of patients of this size being treated with carbamazepine. The benefits of genotyping were clear, with no cases of SJS/TEN among any participants in the study ($p<0.001$ when comparing to assumed incidence based on historical records).

*4.1.1.2 HLA-A*31:01*

There is a large amount of evidence for the association between *HLA-B*15:02* and carbamazepine (CBZ)-induced SJS/TEN (2, 41). However, this allele is much more common in Asian (and mainly Han Chinese) populations compared to the rest of the world (42). A UK study of Caucasian SJS/TEN patients found none were positive for *HLA-B*15:02* (43).

While *HLA-B*15:02* is the most common causative risk allele for CBZ-induced SJS/TEN in Asia (Figure 4.2A) (44), the most common causative gene in the rest of the world is *HLA-A*31:01* (Figure 4.2B) (45, 46).

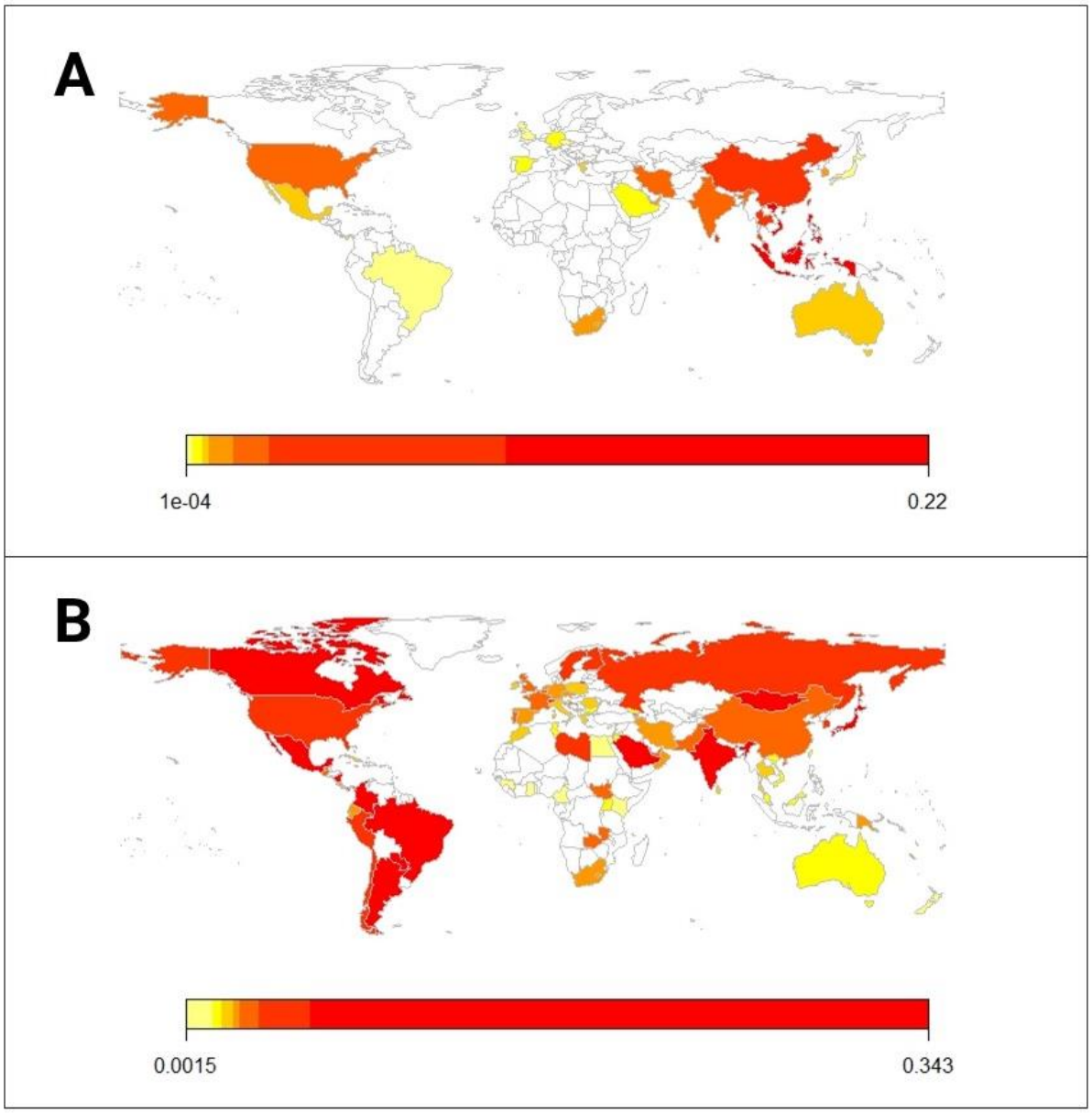


*Figure 4.2 - Frequency of the A) HLA-B*15:02 and B) HLA-A*31:01 alleles worldwide. Areas without colour indicate no data available. Created based on data from allelefrequencies.net (47).*

However, whilst meta-analyses have found odds ratios (ORs) of 3.9 (95% CI 1.4 – 11.5) (48) to 9.45 (95% CI 6.41 – 13.93) (2) for the *HLA-A*31:01*-carbamazepine-SJS/TEN link in mixed Asian/Caucasian cohorts, the association is much less well-studied (2, 24, 45) than that with *HLA-B*15:02*.

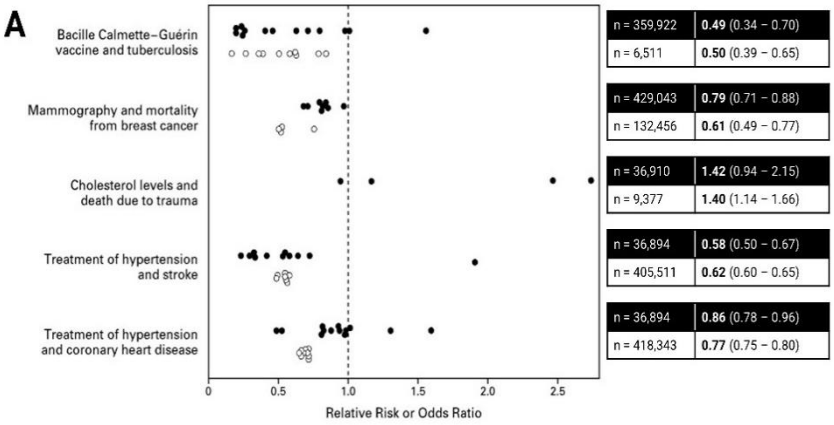
4.1.2 Prospective studies in rare conditions

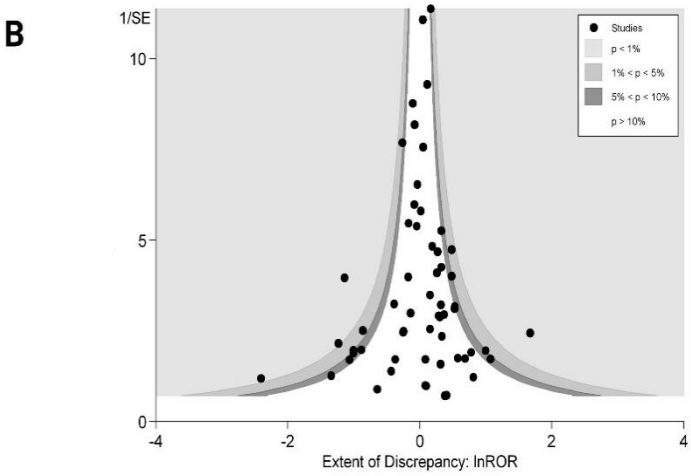
As previously discussed, the randomised controlled trial (RCT) is considered the gold standard for establishing the clinical validity and utility (Chapter 3) of a genotype-guided approach to treatment (49-51). Such trials have demonstrated the benefits of genotyping to guide treatment with drugs such as abacavir (52), and warfarin (53). However, it can be difficult to perform well-controlled RCTs in very rare conditions (54-56) such as SJS/TEN, due to the large sample size required to ensure sufficient statistical power (54, 56-59). In addition, there are circumstances where conducting an RCT is not appropriate. For example, there are ethical issues associated with randomising patients to a drug if it is known that they carry an allele known to confer a heightened risk of ADR on that drug (60-65).

In these cases, data from sources other than RCTs is required to demonstrate clinical validity and utility. Data from observational studies such as case-control studies, disease registries, and n-of-1 trials can be useful tools (54, 56, 63, 64, 66, 67). These are becoming more accepted as sources of evidence by regulatory agencies (see Chapter 2) (68, 69). Further, in their study of the effectiveness of screening for *HLA-B*15:02* prior to carbamazepine treatment, whilst opting for a prospective, interventional study design, Chen et al. did not conduct an RCT, but rather compared results from their prospectively recruited cohort to historical records. Nonetheless, even with this non-randomised design, it was necessary to recruit a large sample size of almost 5000 patients.

It is often overlooked that several well-designed observational studies can provide strong evidence when their data are combined (65, 70, 71). Concato, *et al.* (2000) identified five interventions where separate meta-analyses had been completed first including only RCT data, and second including only observational data (70). For all five interventions, the summary estimates calculated from RCT data were very similar to those calculated from observational data (Figure 4.3A). More recent work by Golder, *et al* (2011) examined the differences in ADR risk reported in observational vs RCT data (65). When comparing RCTs to cohort studies, the reported confidence intervals overlapped in 100% of cases. When comparing RCTs to case-control studies, they overlapped in 90% of cases. The discrepancies between observational and RCT data followed a symmetrical distribution, providing evidence against any systematic bias (Figure 4.3B). Similar work by Benson & Hartz (2000) used individual study data, instead of meta-analysis data (71). In 17 out of 19 studied treatment areas, the effect estimated from observational studies was very

similar to those calculated from RCTs, with overlapping confidence intervals (Figure 4.3C).


A
Bacille Calmette–Guérin vaccine and tuberculosis
Mammography and mortality from breast cancer
Cholesterol levels and death due to trauma
Treatment of hypertension and stroke
Treatment of hypertension and coronary heart disease
n = 359,922 | 0.49 (0.34 – 0.70)
n = 6,511 | 0.50 (0.39 – 0.65)
n = 429,043 | 0.79 (0.71 – 0.88)
n = 132,456 | 0.61 (0.49 – 0.77)
n = 36,910 | 1.42 (0.94 – 2.15)
n = 9,377 | 1.40 (1.14 – 1.66)
n = 36,894 | 0.58 (0.50 – 0.67)
n = 405,511 | 0.62 (0.60 – 0.65)
n = 36,894 | 0.86 (0.78 – 0.96)
n = 418,343 | 0.77 (0.75 – 0.80)
0
0.5
1.0
1.5
2.0
2.5
Relative Risk or Odds Ratio


B
1/SE
10
5
0
-4
-2
0
2
4
Extent of Discrepancy: lnROR
Studies
p < 1%
1% < p < 5%
5% < p < 10%
p > 10%

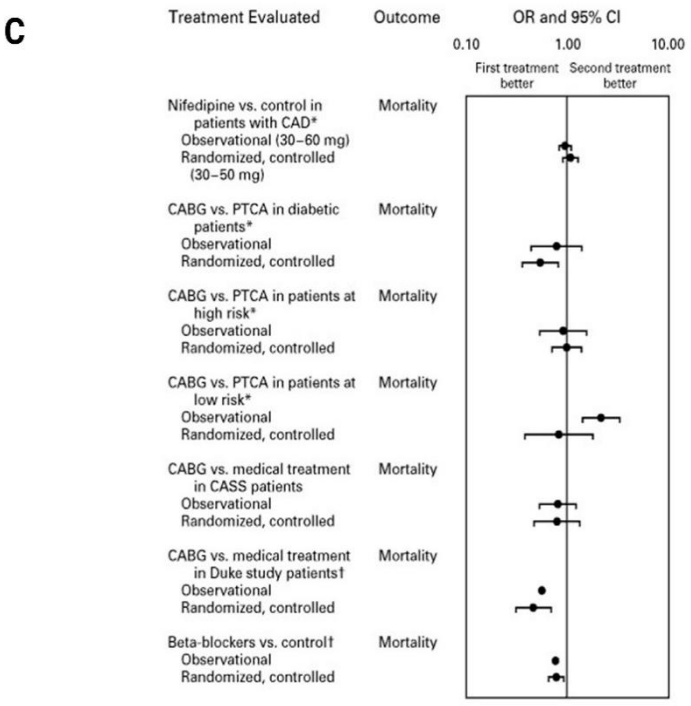

C
Treatment Evaluated
Outcome
OR and 95% CI
0.10
1.00
10.00
First treatment better
Second treatment better
Nifedipine vs. control in patients with CAD*
Observational (30–60 mg)
Randomized, controlled (30–50 mg)
Mortality
CABG vs. PTCA in diabetic patients*
Observational
Randomized, controlled
Mortality
CABG vs. PTCA in patients at high risk*
Observational
Randomized, controlled
Mortality
CABG vs. PTCA in patients at low risk*
Observational
Randomized, controlled
Mortality
CABG vs. medical treatment in CASS patients
Observational
Randomized, controlled
Mortality
CABG vs. medical treatment in Duke study patients†
Observational
Randomized, controlled
Mortality
Beta-blockers vs. control†
Observational
Randomized, controlled
Mortality

Figure 4.3 - meta-analyses including only RCTs produced similar estimates of relative risk or odds ratios to those including only observational data. A) Data from observational (open circles) and RCTs plotted by Concato et al (2000) (70). Solid circles are RCTs and open circles are observational studies. The adjacent table shows sample sizes and effect sizes (with 95% confidence intervals) for each intervention. The black rows are RCTs and the white rows are observational studies. The results of the observational studies in cholesterol levels and death due to trauma were not individually reported so could not be plotted. B) shows a funnel plot of the log discrepancies between meta-analyses of RCT and observational data, plotted by Golder et al (2011) (65). Each dot represents one study. The y-axis (1/standard error) illustrates precision of the estimate. Studies lower down the y-axis have less precision. C) Confidence intervals of odds ratios of RCT and observational data results as plotted by Benson and Hartz (2000). Cardiac interventions only. CABG = coronary artery bypass graft surgery. CAD = coronary artery disease. CASS = Coronary Artery Surgery Study. lnROR = log ratio of odds ratios. OR = odds ratio. PTCA = percutaneous transluminal coronary angioplasty. RCT = randomised controlled trial. SE = Standard error.

These studies show that effect estimates obtained from observational data are often comparable to those obtained from RCT data, and for this reason, it is reasonable to consider using observational data as an alternative to prospective, interventional studies when exploring the benefits of a genotype-guided approach to treatment where outcome is a rare event such as SJS/TEN.

However, effect estimates (for example in the form of odds ratios, ORs) alone are not sufficient evidence for a test's utility (67). The clinical validity also needs to be considered. Clinical validity is determined by the discriminative accuracy and predictive value of a test. Discriminative accuracy consists of sensitivity and specificity, and is defined as the ability of a test to discriminate between the presence and absence of an outcome (67). Predictive value is the ability of a genetic test to predict an outcome from the presence or absence of a variant (67). It consists of positive predictive value (PPV) and negative predictive value (NPV). The PPV of a genetic test is the probability of an outcome (e.g., an ADR) when a genetic variant is present. NPV is the probability of the outcome not occurring when the genetic variant is not present.

It is important to estimate these measures of clinical validity, when considering the true impact of genetic testing in the wider population. The measures are important to allow translation of pharmacogenetics knowledge into practice and to allow for the measurement of clinical utility, the ability of the test to improve health outcomes (67, 72, 73).

Effect estimates (such as OR) for gene-ADR associations are often calculated from case-control studies. Measures of clinical validity generally improve when OR is

higher. However, a higher OR does not automatically translate into higher values for all measures of clinical validity. These values also depend on the frequency of the genetic variant and of the ADR of interest.

Tonk, *et al.* (2016) showed that sensitivity, specificity, PPV and NPV of a genetic test can be estimated from case-control studies if estimates of the frequency of the genetic variant and the frequency of the ADR are incorporated into the calculations (67) (Figure 4.4).

		Adverse drug reaction		
		+	-	Total
Variant	+	a	b	p
	-	c	d	$(1-p)$
	Total	q	$(1-q)$	1

$p = Variant\ frequency$

$q = ADR\ frequency$

$$OR = \frac{ad}{bc}$$

$$PPV = \frac{a}{p}$$

$$Sensitivity = \frac{a}{q}$$

$$NPV = \frac{d}{(1-p)}$$

$$Specificity = \frac{d}{(1-q)}$$

Figure 4.4 - Method for calculating the effect estimates (OR), discriminative accuracy (sensitivity and specificity), and predictive value (PPV and NPV) of a genetic test, given the variant frequency and the ADR frequency are known. Adapted from Tonk et al 2016 (67). ADR = adverse drug reaction. NPV = negative predictive value. OR = odds ratio. PPV = positive predictive value.

Figure 4.4 depicts the 2x2 table from a typical case-control study of a genetic association between a genetic variant and adverse event, with a = proportion of

patients who are variant carriers and have the ADR; b = proportion of patients who are variant carriers but do not have the ADR; c = proportion of patients who are variant non-carriers but have the ADR; d = proportion of patients who are variant non-carriers and do not have the ADR. According to Tonk, et al., if the genetic variant frequency (p) and the ADR frequency (q) are known, the values of a-d can be calculated as follows (67):

$$a = \frac{\left(p*OR+(1-p)+q*(OR-1)\right)-\sqrt{\left(\left(-p*OR-(1-p)-q*(OR-1)\right)^2-4*(OR-1)*p*q*OR\right)}}{2*(OR-1)}$$

$$b = p - a$$

$$c = q - a$$

$$d = 1 - a - b - c$$

Given that it is possible to estimate measures of clinical validity from observational case-control studies, here I propose to answer the question: will a prospective interventional clinical trial allow the estimation of the clinical validity of a genetic marker with better precision than observational case-control data, if that genetic marker already has good evidence of association in those observational studies?

In order to explore this, I chose a rare ADR, SJS/TEN, well known to be associated with 2 genetic variants (*HLA-B*15:02* and *HLA-A*31:01*). I wished to compare the precision of clinical validity measures and ORs obtained from prospective, interventional studies to those obtained from observational studies of the associations. For the *HLA-B*15:02* variant, I compared the precision of effect estimates and clinical validity obtained from the previously conducted prospective study of Chen, *et al.* (2011) to data from observational studies, synthesised within a random-effects meta-analysis. For the *HLA-A*31:01* variant, there were no previous prospective studies testing its clinical validity and therefore I simulated a prospective study with similar characteristics to Chen, *et al.* but assuming *HLA-A*31:01* genotyping instead of *HLA-B*15:02*. The effect estimates and clinical validity from the simulated study were then compared to those obtained from data on the same association from previous observational studies, again synthesised within a random-effects meta-analysis. I hypothesise that with full and effective use of observational

data, along with estimates of allele frequency and ADR frequency, prospective, interventional studies may not be required to demonstrate the benefits of testing for the *HLA-A*31:01* allele prior to commencing carbamazepine. Since prospective studies are generally expensive (in both time and money) and may be difficult to generalise to wider populations (70, 74, 75), observational studies provide valuable and easier to collect data that can be used to determine clinical validity.

4.2 Methods

4.2.1 Previous meta-analyses of observational studies

Before undertaking these systematic reviews and meta-analyses, I examined the literature for previous reviews in this field in order to inform my methods.

Four previous reviews were located (2, 48, 76, 77) (Table 4.1). One of these only included *HLA-B*15:02* (76), another only *HLA-A*31:01* (48). The others contained analyses for both alleles (2, 77). I referred to these previous reviews to help refine the search terms and outcome measures prior to conducting my own reviews. The ancestries of participants in these reviews were majority Asian, mostly Han Chinese and Japanese.

Reference	No. included studies	No. included participants	Allele	Allele frequency	Odds ratio for ADR (95% CI)	Ancestries	Assessed study quality?	Primary outcome(s)
Yip, et al. 2012 (20)	23	690 (cases) 1585 (CBZ-tolerant controls) 7056 (population controls)	HLA-B*15:02	0.96 (CBZ cases) 0.11 (CBZ controls)	113.39 (51.24 – 250.97) *	1332 Han Chinese 346 Malay 215 Thai 111 Caucasian 18 Indian 7 Japanese	Reported in methods but not results – Jorgensen and Williamson method (22)	Hypersensitivity reaction to carbamazepine
			HLA-A*31:01	n/a	9.45 (6.41 – 13.93)	497 Japanese 402 Caucasian 162 Han Chinese 74 Korean		
Tangamornsuksan, et al. 2013 (23)	16	227 (SJS/TEN cases) 602 (matched controls) 2949 (population controls)	HLA-B*15:02	0.79 [174/220] (cases) ▲ 0.033 [102/3066] (controls) ▲	79.84 (28.45 – 224.06)	909 Japanese 542 Korean 493 Han Chinese 132 Thai 45 European 18 Indian 1632 Mixed populations	Yes – Newcastle-Ottawa scale	Carbamazepine induced SJS/TEN
Grover, et al. 2014 (25)	20	336 (cases) 692 (CBZ-tolerant controls)	HLA-B*15:02	0.60 [203/336] (cases) ▲ 0.10 [70/692] (CBZ controls) ▲	19.33 (8.51 – 43.91)	Not broken down by ancestry	No	Carbamazepine-induced cutaneous ADRs
		413 (cases) 1020 (CBZ-tolerant controls)	HLA-A*31:01	0.27 [111/413] (cases) ▲ 0.08 [78/1020] CBZ (controls) ▲	7.75 (5.34 – 11.25)	Not broken down by ancestry		
Genin, et al. 2013 (26)	5	167 (CBZ cases) 1291 (CBZ-tolerant controls)	HLA-A*31:01	0.11 [18/167] cases ▲ 0.07 [91/1291] controls ▲	3.9 (1.4 – 11.5)	615 European 426 Japanese 329 Han Chinese 57 Korean 31 Asian other	No	Carbamazepine induced SJS/TEN

*Table 4.1 - previous systematic reviews and meta-analyses of carbamazepine-induced SJS/TEN including HLA-B*15:02 or HLA-A*31:01. CBZ cases are those that developed the ADR when exposed to carbamazepine. CBZ controls did not develop the ADR when exposed (CBZ-tolerant). Numbers in ancestries only include the number of patients for each allele, so may not sum to the total number of participants where other alleles were investigated. ADR = adverse drug reaction. CBZ = carbamazepine. CI = confidence interval. SJS/TEN = Stevens-Johnson syndrome/toxic epidermal necrolysis. *In Asian (Han Chinese/Thai/Malaysian patients only. ▲ number of risk allele positive patients who experienced the ADR/total number of patients who experienced the ADR.*

4.2.2 Protocol development

A protocol was developed to guide the systematic reviews and meta-analyses for both *HLA-B*15:02* and *HLA-A*31:01*. The protocol (see Appendix 4.1) was developed in accordance with the 2015 Preferred Reporting Items for Systematic reviews and Meta-Analyses for Protocols (PRISMA-P) statement (78). Both reviews were conducted in the same manner. Reporting of the methods for these reviews follows the PRISMA 2020 reporting guidelines for systematic reviews (79). This protocol was registered on The International Prospective Register of Systematic Reviews, PROSPERO, on 9th December 2019 (CRD42019161000) (80).

4.2.3 Search strategies

The first systematic review examines the association between *HLA-B*15:02* and carbamazepine-induced adverse drug reactions (including SJS/TEN). I aimed to include all studies to date that examined this association, including retrospective, prospective, case-control, and RCT designs. To be included, studies had to include participants exposed to carbamazepine, assessed for *HLA-B*15:02*, and with cases of any hypersensitivity reaction (including SJS, TEN, SJS/TEN, DRESS, MPE, etc). Studies also had to include a comparator group (of carbamazepine-tolerant controls, or healthy volunteers) also genotyped for *HLA-B*15:02*. Literature reviews, case studies, non-human studies, and any papers that were only an abstract were excluded.

The second systematic review was conducted in a similar manner. It examined the association between *HLA-A*31:01* and carbamazepine-induced adverse drug reactions. The same inclusion and exclusion criteria as above were applied, but with *HLA-A*31:01* in place of *HLA-B*15:02.*

The primary outcome of interest was the development of SJS/TEN in response to carbamazepine. Secondary outcomes were development of (in response to carbamazepine): any hypersensitivity reaction, SJS, TEN, DRESS, or MPE. Details of any outcomes defined by papers' authors as a carbamazepine-related ADR were also extracted.

Searches were not limited to English results only, but found that any non-English papers did not pass initial screening processes. Conference abstracts meeting eligibility criteria were included, and used to locate full relevant journal articles. If a full article could not be located, the conference abstracts were excluded as they did not contain enough information to inform the systematic review and meta-analyses.

For the *HLA-B*15:02* review the Medline database was searched on 7th January 2020 (Table 4.2). No filters or limits on dates or publication type were used when searching.

#	Search term	Notes
1	HLA-B Antigens	MeSH term
2	"HLA-B*15:02"	Free text (.mp)
3	Carbamazepine	MeSH term
4	"tegretol"	Free text (.mp)
5	Stevens-Johnson Syndrome	MeSH term
6	"toxic epidermal necrolysis"	Free text (.mp)
7	"SJS"	Free text (.mp)
8	"TEN"	Free text (.mp)
9	"SJS/TEN"	Free text (.mp)
10	"drug reaction with eosinophilia and systemic symptoms"	Free text (.mp)
11	"DRESS"	Free text (.mp)
12	"maculopapular exanthema"	Free text (.mp)
13	"MPE"	Free text (.mp)
14	Drug Hypersensitivity OR "hypersensitivity reaction"	MeSH term/Free text (.mp)
15	"cutaneous adverse drug reaction"	Free text (.mp)
16	Drug-Related Side Effects and Adverse Reactions OR "cutaneous ADR"	MeSH term/Free text (.mp)
17	1 OR 2	Combining
18	3 OR 4	Combining
19	5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16	Combining
20	17 AND 18 AND 19	Combining

*Table 4.2- HLA-B*15:02 review search terms Medline. '.mp' denotes a search of the Medline fields: title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms. MeSH = Medical Subject Headings, used to index articles.*

For the *HLA-A*31:01* review, the Medline database was searched on 25th February 2020, using very similar search terms to the *HLA-B*15:02* search (Table 4.3). No filters or limits were used when searching.

#	Search term	Notes
1	HLA-A Antigens	MeSH term
2	"HLA-A*31:01"	Free text (.mp)
3	Carbamazepine	MeSH term
4	"tegretol"	Free text (.mp)
5	Stevens-Johnson Syndrome	MeSH term
6	"toxic epidermal necrolysis"	Free text (.mp)
7	"SJS"	Free text (.mp)
8	"TEN"	Free text (.mp)
9	"SJS/TEN"	Free text (.mp)
10	"drug reaction with eosinophilia and systemic symptoms"	Free text (.mp)
11	"DRESS"	Free text (.mp)
12	"maculopapular exanthema"	Free text (.mp)
13	"MPE"	Free text (.mp)
14	Drug Hypersensitivity OR "hypersensitivity reaction"	MeSH term/Free text (.mp)
15	"cutaneous adverse drug reaction"	Free text (.mp)
16	Drug-Related Side Effects and Adverse Reactions OR "cutaneous ADR"	MeSH term/Free text (.mp)
17	1 OR 2	Combining
18	3 OR 4	Combining
19	5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16	Combining
20	17 AND 18 AND 19	Combining

*Table 4.3 - HLA-A*31:01 review search terms Medline. '.mp' denotes a search of the Medline fields: title, abstract, original title, name of substance word, subject heading word, floating sub-heading word,*

keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms. MeSH = Medical Subject Headings, used to index articles.

4.2.4 Screening

Search results were initially screened by title and abstract by 2 authors. Differences were resolved by discussion. Full text screening was then conducted on the remaining papers in the same manner. Papers were classified by the type of control group: either carbamazepine-tolerant controls or healthy volunteer controls (or both).

4.2.5 Data extraction

Data was extracted using a standard data extraction form (Appendix 4.2). The form was piloted by extracting data from 5 studies, and amending the form as required. The form contained sections for study design, patient demographics, as well as genotyping results. The number of SJS/TEN cases, controls, and healthy volunteers (if applicable) and their genotyping status (*HLA-B*15:02* or *HLA-A*31:01* positivity) were collected. Details of other outcomes were also collected where provided (SJS and TEN individually, DRESS, MPE, and any other cutaneous ADRs).

Data was extracted by DJ, with the extracted data from 10% of studies checked by AJ.

4.2.6 Meta-analysis

Papers located in the systematic review phase were then used in meta-analyses of the association between *HLA-B*15:02* and *HLA-A*31:01* genotype and carbamazepine-induced SJS/TEN. Separate comparisons of cases to carbamazepine-tolerant controls, and cases to healthy volunteers was undertaken, where the data were available.

The text of papers included in the systematic review was examined to identify if there was an overlap of participants between papers (e.g. “some of these cases were previously reported in Smith, *et al.*”). In these cases, authors were contacted to identify details of the overlap. If there was no reply, only the paper with the larger number of participants was included in the meta-analysis.

Three different sets of meta-analyses were undertaken. In each set, one meta-analysis compared cases to carbamazepine-tolerant controls whilst the other compared cases to healthy volunteers.

The first set of meta-analyses investigated the association between the *HLA-B*15:02* genotype and susceptibility to carbamazepine-induced SJS/TEN, including all identified papers.

Following this, meta-analyses were conducted where only papers available before the publication of Chen, *et al.* (2011) (3) were included. The rationale for this was in order for me to consider whether a prospective study was justified, given the already accrued evidence to date from observational studies.

The final meta-analyses investigated the association between *HLA-A*31:01* genotype and susceptibility to carbamazepine-induced cADRs, including SJS, TEN, SJS/TEN, DRESS, and MPE phenotypes (where data available).

Meta-analyses were performed in R, using the 'meta' and 'forestplot' packages (81, 82). Forest plots, odds ratios, and 95% confidence intervals were generated. The I^2 statistic was used to assess heterogeneity. The 'meta' package allows for the meta-analysis of binary outcome data with the 'metabin' call. By default, this uses the Mantel-Haenzel method to calculate effect estimates. This method has better statistical properties for rare events, and is the preferred method used by Cochrane reviewers (83). A random effects approach to calculating the pooled effect estimate (OR) was chosen since the included studies were heterogeneous (in study design and included ethnicities) and so it was assumed that effect estimates would be similar but not identical across studies (84).

Results were also compared to existing meta-analyses, including an assessment of any differences in included papers and in effect sizes. Some studies were included in the systematic reviews but not in the meta-analyses. Reasoning for these decisions was undertaken based on the availability of summary statistics and lack of overlap with other included papers. Full details are provided in the full list of studies, included in Appendix 4.3.

4.2.7 Quality assessment

Each included paper, in both reviews, was also assessed for quality using the criteria of Jorgensen and Williamson (2008) (85). These criteria have been used in other meta-analyses to assess study quality (2, 86, 87). Study quality was assessed by DJ, with AJ checking 10% of papers. Differences were resolved by discussion. Full details of study quality assessment are available in Appendix 4.4.

Results were represented diagrammatically with a heat map so that the general quality of included studies could be visualised, however the results of quality assessment were not used to inform any subgroup analyses of the meta-analyses.

4.2.8 Simulation

I aimed to simulate a prospective study with the same study design as Chen *et al.* 2011 (3), but assuming that treatment was guided in accordance with *HLA-A*31:01* allele carrier status instead of *HLA-B*15:02*. The details of this trial were analysed in the PICO format (Table 4.4) (88, 89). These details were then used to guide the design of the simulated prospective study.

Patients/Population	4877 Han Chinese patients requiring carbamazepine treatment (indications include epilepsy, neuralgia, neuropathic pain, tinnitus, psychiatric disorders). Ages 6 months to 99 years. Excluding: carbamazepine allergy, patients who had undergone bone marrow transplant, patients not of Han Chinese descent
Intervention	Genotyping – all patients genotyped at first clinic visit. Those who were ***HLA-B*15:02*** **positive** were given information about SJS/TEN and recommended alternative drugs. ***HLA-B*15:02*** **negative** patients also received information about SJS/TEN but were started on carbamazepine.
Comparison	Compared to historical incidence of SJS/TEN in Taiwan. This was based on records of patients with ICD9 diagnostic code 695.1 (erythema multiforme) (90). This number was modified to calculate the number of patients with carbamazepine-induced SJS/TEN according to methodology from a previous study in China (91).
Outcome	Rates of SJS/TEN and other cADRs. Only SJS/TEN was compared to historical data

Table 4.4 - PICO analysis of Chen et al 2011 (3). ICD9 = International Classification of Diseases 9. cADRs = cutaneous adverse drug reactions. SJS/TEN = Stevens-Johnson syndrome/toxic epidermal necrolysis

It is assumed that the simulated prospective cohort would include patients without the *HLA-A*31:01* risk allele (non-carriers), who are treated with carbamazepine, as well as patients with the *HLA-A*31:01* allele (carriers) who are not treated with carbamazepine.

Since all previous studies identified were case-control studies, none provided an estimate for the probability of developing SJS/TEN in non-carriers of *HLA-A*31:01*. Bayes' theory (92) was therefore used to estimate the risk:

$$P(A|B) = \frac{P(B\ |A) * P(A)}{P(B)}$$

where $P(A)$ is the probability of SJS/TEN when taking carbamazepine and $P(B)$ is the probability of *HLA-A*31:01* not being present. $P(B\ |A)$ is the probability of not being a *HLA-A*31:01* allele carrier given that they have developed SJS/TEN and $P(A\ |B)$ is the risk of interest (risk of developing SJS/TEN in non-carriers of *HLA-A*31:01*).

$P(A)$ was estimated to be 0.23% (or 0.0023). This was the same historical incidence assumed by Chen, *et al* (3). $P(B)$ was estimated from allelefrequencies.net (47) (Appendix 4.5), to be 0.950411 (this is the weighted mean of all Chinese population estimates listed on allelefrequencies.net).

$P(B\ |A)$ was estimated from case-control studies included in the meta-analysis, as the proportion of non-carriers amongst cases. This equated to 0.265. This was calculated using data from all patients regardless of ethnicity, since there was an insufficient amount of data in Han Chinese participants to provide a reliable estimate.

Therefore, using Bayes' theory (92) there is:

$$P(A|B) = (0.264535 * 0.0023) \,/\, 0.950411 \ = 0.0006401762$$

Therefore, the probability of developing SJS/TEN in carbamazepine patients not carrying *HLA-A*31:01* is estimated as $6.40 * 10^{-4}$.

These and the other estimates used as parameters for the simulation of the prospective cohort are summarised in Table 4.5. The 'event' in this simulation is SJS/TEN.

Parameter	Description	Value	Justification
`nPatients`	Number of patients in each simulated prospective cohort	5,000	Chen et al recruited around 5000 participants (3)
`nSims`	Number of simulations	10,000	This number is sufficient to allow for calculation of confidence intervals using bootstrapping (93, 94)
`prospective.nc`	Event rate for non-carriers	0.0006401762	As per calculation above
`prospective.c`	Event rate for carriers	0	As per justification above
`histinc`	Event rate for patients who do not receive genotyping	0.0023	This is equal to the risk of SJS/TEN in the general, ungenotyped patient population, taken from Chen, *et al* (3), as explained above
`allele.freq`	Frequency of being a non-carrier for the *HLA-A*31:01* allele in the general Han Chinese population	0.9606	Taken from allelefrequencies.net, as explained above (47)

*Table 4.5 – assumptions used in the R code to simulated a prospective interventional study of HLA-A*31:01 genotyping for the prevention of SJS/TEN.*

Data was simulated for n=5000 patients (`nPatients`), undertaking 10,000 simulations each time. To do this, carrier status for each simulated participant in the cohort was simulated using the '`rbinom`' function in R, assuming a probability of 0.9606 of being a non-*HLA-A*31:01* carrier. The outcome for each participant (ADR or no ADR) was then simulated, conditional on their allele carrier status. For non-

carriers, this was again done using the '`rbinom`' function in R assuming a probability of 0.0006401762 of developing the ADR. An event rate of 0 was assumed in carriers, since these patients would test positive if genotyped in a prospective trial and would not receive carbamazepine.

Code used for these analyses can be found in Appendix 4.6.

4.2.9 Comparison of effect estimates and measures of clinical validity

As I wanted to compare the precision of effect estimates and measures of clinical validity (in particular PPV and NPV) between those obtained from observational data and a) the prospective trial (Chen, *et al.* (2011) in the case of *HLA-B*15:02*; and b) a simulated prospective trial in the case of *HLA-A*31:01*, I first considered how the precision of the estimates derived from observational data varied with the accrual of cases. To do this, the following steps were taken:

i) First, the effect estimate (OR) was assumed to be equal to that observed in the meta-analysis for *HLA-B*15:02*/*HLA-A*31:01* for the Han Chinese population (the same population as that studied in Chen et al. (3));

ii) estimating the frequency, *p*, of *HLA-B*15:02* (0.0496) and *HLA-A*31:01* (0.0207) in Han Chinese from allelefrequencies.net and assuming the same incidence, *q*, of SJS/TEN (0.0023) as that assumed in Chen et al. (3) I used the approach suggested in Tonk et al. (67) (see Figure 4.4) to estimate PPV and NPV;

iii) since the precision of these estimates will vary with the number of SJS/TEN cases, the 95% confidence interval for the OR for 1- 100 cases was calculated using the approach outlined below. A total of 500 controls were assumed, but I also ran the calculations for 250 and 1000 controls to check that this did not impact the conclusions, and no significant difference in results was seen.

4.2.9.1 Approach for calculating 95% confidence interval for the OR at various numbers of cases

If we assume the data can be presented in a 2x2 table as follows:

		cases	Controls	Total
Variant	+	a	b	$a+b$
	-	c	d	$c+d$
	Total	$a+c$	$b+d$	$a+b+c+d$

Where:

$$a = Number\ of\ cases\ with\ the\ variant$$

$$b = Number\ of\ controls\ with\ the\ variant$$

$$c = Number\ of\ cases\ without\ the\ variant$$

$$d = Number\ of\ controls\ without\ the\ variant$$

then, since we already have an estimate for OR, and know that:

$$OR = {}^{ad}/_{bc}$$

we can create the following equations, assuming an allele frequency of p and solve the following equations to obtain a 95% confidence interval for a particular number of cases:

$$b = p * controls$$

$$d = (1-p) * controls$$

$$c = \frac{d * cases}{(OR * b) + d}$$

$$a = cases - c$$

$$SE = \sqrt{\left(\left(\frac{1}{a}\right) + \left(\frac{1}{b}\right) + \left(\frac{1}{c}\right) + \left(\frac{1}{d}\right)\right)}$$

$$Lower\ limit\ of\ 95\%\ confidence\ interval = \exp\left(\log(OR) - (1.96 * SE)\right)$$

$$Upper\ limit\ of\ 95\%\ confidence\ interval = \exp\left(\log(OR) + (1.96 * SE)\right)$$

Once estimates for the 95% confidence intervals for the OR at various numbers of cases were obtained, we could input the lower limit and upper limit, to replace OR in

the earlier equations, to obtain 95% confidence intervals for PPV and NPV at various number of cases. The 95% confidence intervals for OR, PPV and NPV, were then plotted against number of cases.

4.3 Results

4.3.1 *HLA-B*15:02* systematic review and meta-analysis

The search of the Medline database yielded 146 results. The search strategy was validated by checking that these results included some key references identified from initial literature searching. All the key references were included.

After screening these results by title and abstract, 51 results remained for full text screening. Further evaluation by full text reduced this to 47 papers for inclusion in the systematic review (11, 15-17, 40, 43, 48, 95-133) and 19 of these reported data in sufficient detail that they could be included in the carbamazepine tolerant controls meta-analysis. A total of 12 were included in the healthy volunteer controls meta-analysis (Figure 4.5).

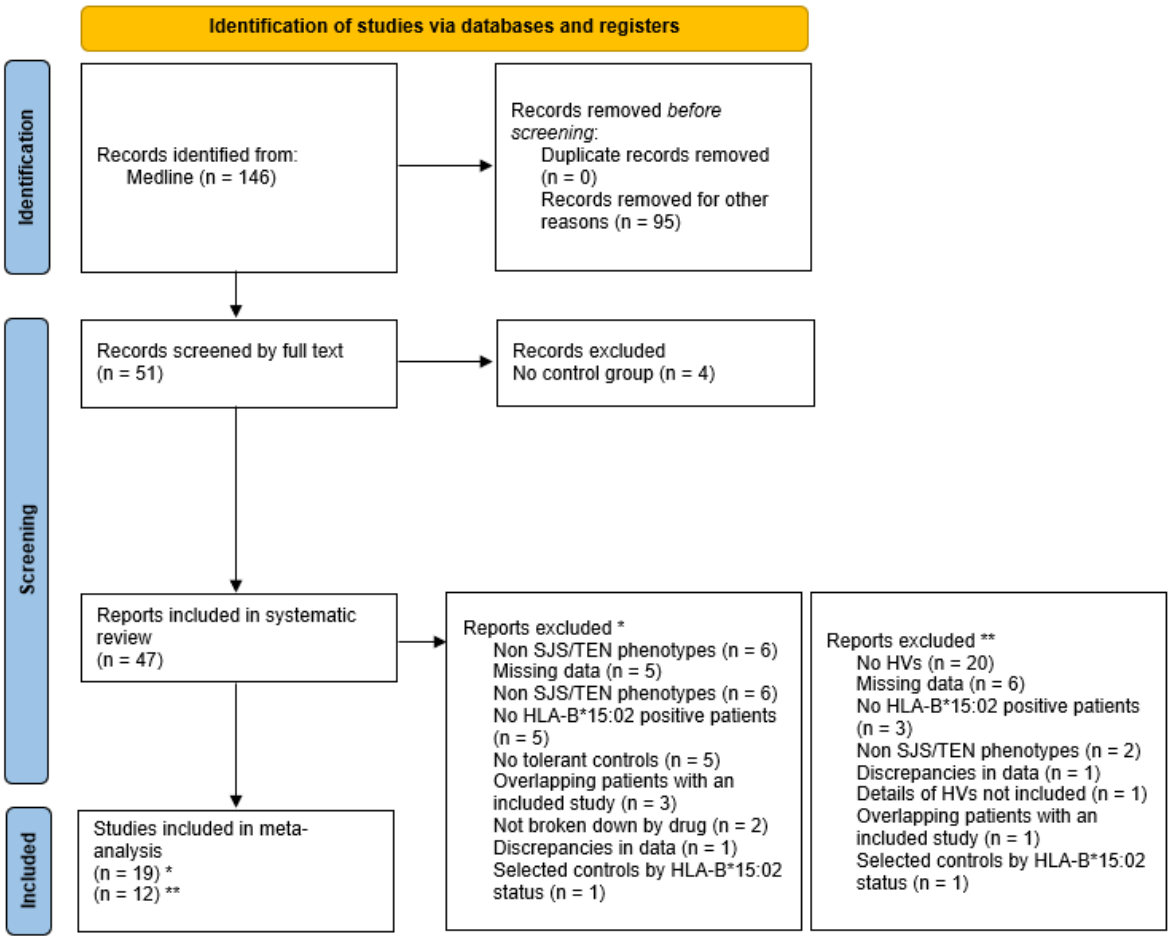
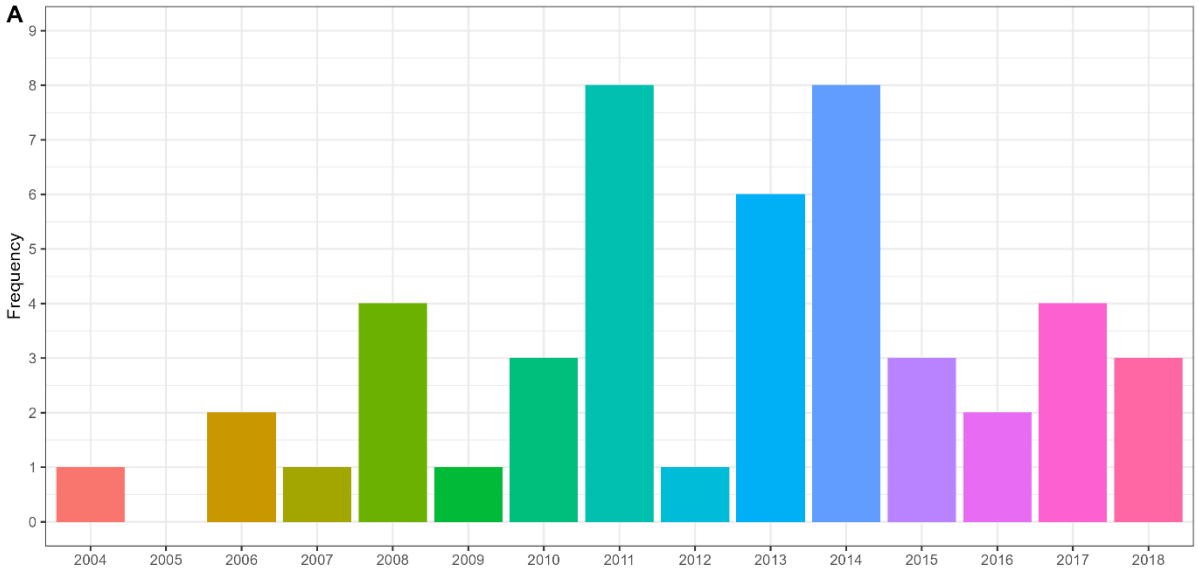
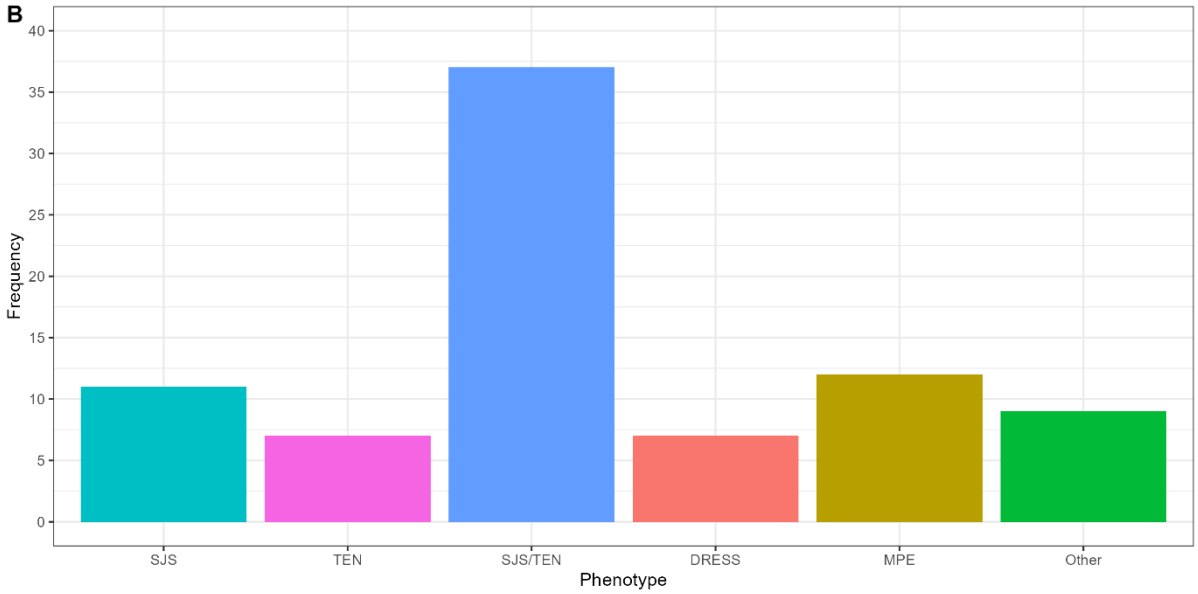


*Figure 4.5 – flow of papers included in HLA-B*15:02 systematic review and meta-analysis. * denotes meta-analysis with cases and tolerant controls. ** denotes meta-analysis with cases and healthy volunteer controls. HVs = healthy volunteers. SJS/TEN = Stevens Johnson syndrome/toxic epidermal necrolysis. Flow chart adapted from PRISMA 2020 (79).*

Full details of included papers are shown in Appendix 4.3. Papers were published between 2004 and 2018, with peaks in 2011 and 2014 (Figure 4.6A). The most common phenotype included was SJS/TEN, with SJS and TEN also being included as separate outcomes (Figure 4.6B). The most common design was case control, with carbamazepine-tolerant patients as controls (Figure 4.6C). One paper was a prospective design, but without any intervention (126).


A
Frequency
0
1
2
3
4
5
6
7
8
9
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018


B
Frequency
0
5
10
15
20
25
30
35
40
SJS
TEN
SJS/TEN
DRESS
MPE
Other
Phenotype

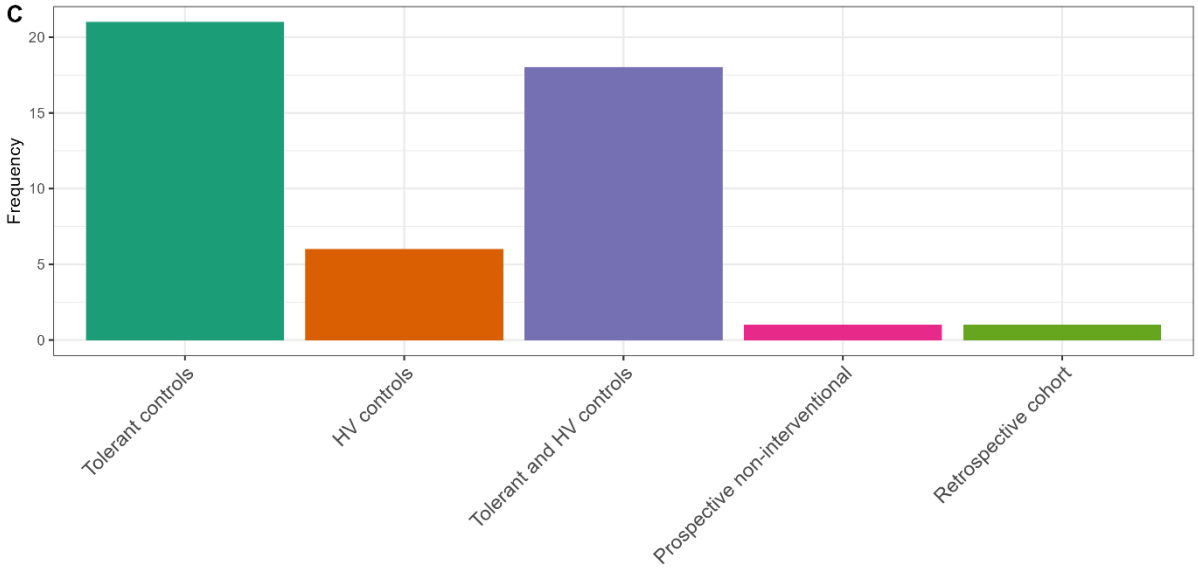

C
Frequency
0
5
10
15
20
Tolerant controls
HV controls
Tolerant and HV controls
Prospective non-interventional
Retrospective cohort

*Figure 4.6 - A) Number of papers in HLA-B*15:02 systematic review by year of publication. B) Phenotypes included in papers in systematic review. Some papers included more than one phenotype. C) Designs of papers included in the systematic review. Case control refers to papers that used carbamazepine-tolerant patients as controls. HV = healthy volunteers.*

The mean total sample size was 647.5 (SD 1678.32) and the median total sample size was 190.0 (IQR 74.0 – 340.0). A total of 29,137 participants were included in total across all papers, of which 2,560 were cases, 10,545 were carbamazepine-tolerant controls, and 16,032 were healthy volunteers (although there is some overlap in participants between papers).

The mean age of cases and drug-tolerant controls was similar (cases: 36, SD 12.7, controls: 35, SD 12.0). The mean age for healthy volunteers was older (42, SD 10.3), although there was more missing data for this variable. Cases, controls, and healthy volunteers were gender balanced (48.8%, 51.5%, and 42.7% male respectively).

The majority of papers performed well on reporting the way genes were chosen for genotyping and details of sample size and study design. However, very few papers included results of testing for Hardy-Weinberg equilibrium, consideration of missing data and how it was dealt with, and how adherence with treatment was assessed and adjusted for in the analyses. A diagrammatic overview of study methodological quality is shown in Figure 4.7, and full analyses and the criteria used are located in Appendix 4.4.

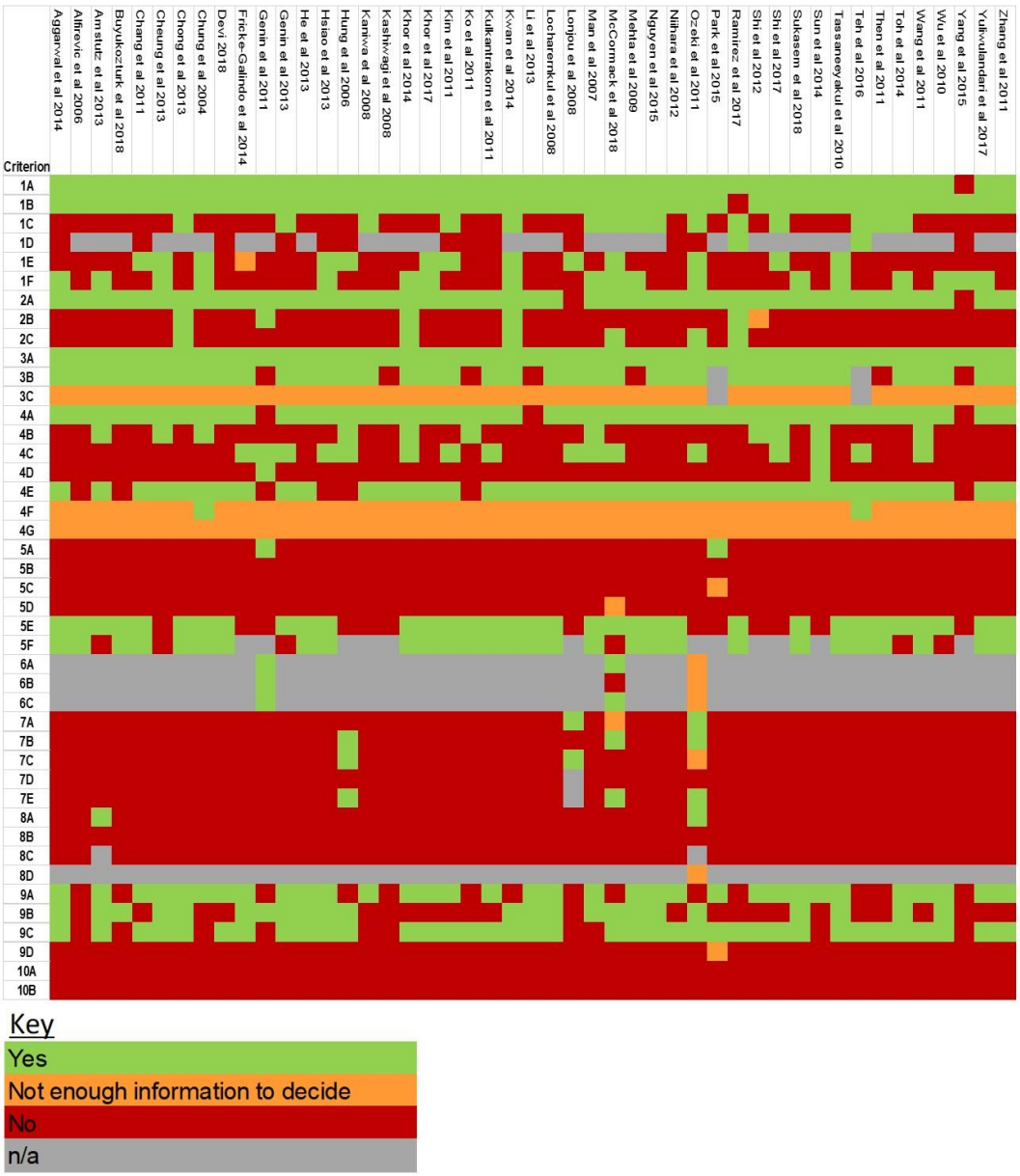


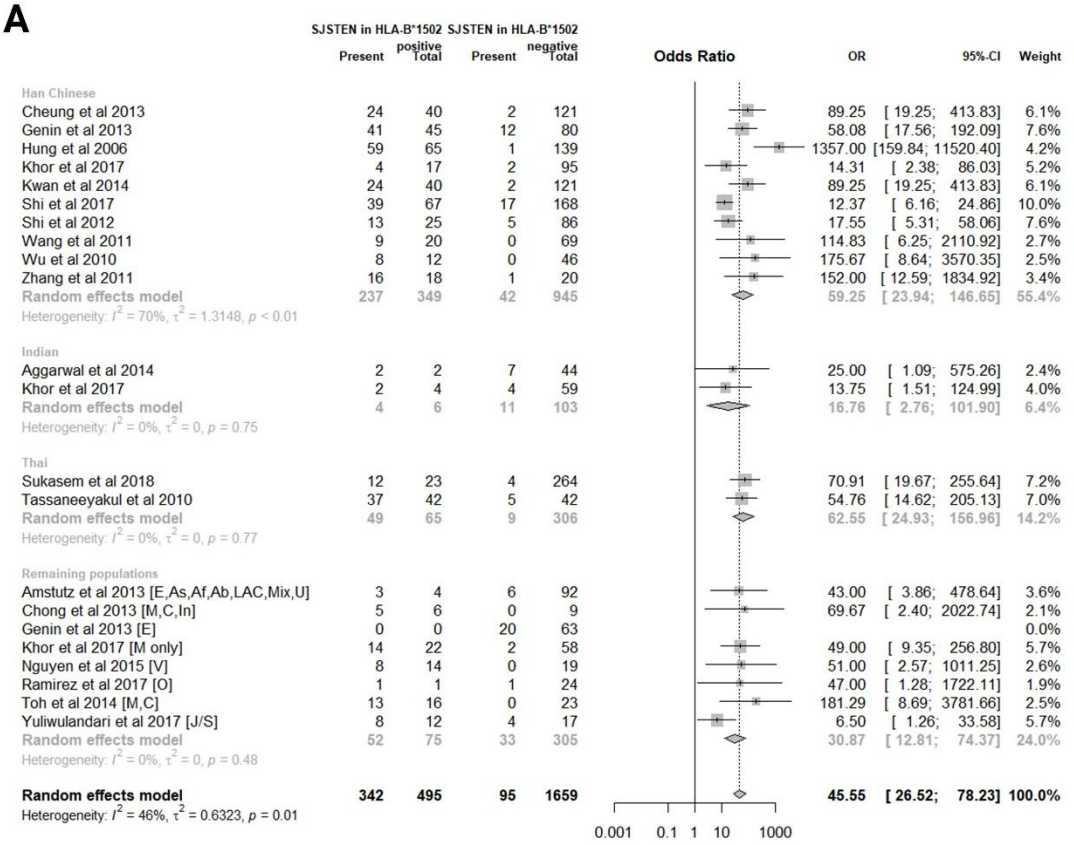
*Figure 4.7 - quality of studies included in HLA-B*15:02 systematic review, according to criteria of Jorgensen and Williamson (2008) (85). Full criteria and data are included in Appendix 4.4.1.*

*4.3.1.1 HLA-B*15:02 meta-analysis*

The first meta-analysis investigated the association between the *HLA-B*15:02* genotype and susceptibility to carbamazepine-induced SJS/TEN, compared to carbamazepine-tolerant controls. A total of 19 papers were included, comprising 495 cases and 1659 controls (Figure 4.8A). The second analysis was the same, but used healthy volunteers as controls. A total of 12 papers were included in this analysis, comprising 467 cases and 11336 controls (Figure 4.8B).

There was overlap in the patients included in Hung, *et al.* 2006 and Hsiao, *et al*. 2013. While Hsiao, *et al*. had a larger overall sample size, Hung, *et al.* had more SJS/TEN patients. Hung *et al.* was therefore included in the SJS/TEN meta-analysis and Hsiao *et al* excluded.

A


SJSTEN in HLA-B*1502 positive
SJSTEN in HLA-B*1502 negative
Present Total Present Total
Odds Ratio
OR 95%-CI Weight
Han Chinese
Cheung et al 2013 24 40 2 121 89.25 [19.25; 413.83] 6.1%
Genin et al 2013 41 45 12 80 58.08 [17.56; 192.09] 7.6%
Hung et al 2006 59 65 1 139 1357.00 [159.84; 11520.40] 4.2%
Khor et al 2017 4 17 2 95 14.31 [2.38; 86.03] 5.2%
Kwan et al 2014 24 40 2 121 89.25 [19.25; 413.83] 6.1%
Shi et al 2017 39 67 17 168 12.37 [6.16; 24.86] 10.0%
Shi et al 2012 13 25 5 86 17.55 [5.31; 58.06] 7.6%
Wang et al 2011 9 20 0 69 114.83 [6.25; 2110.92] 2.7%
Wu et al 2010 8 12 0 46 175.67 [8.64; 3570.35] 2.5%
Zhang et al 2011 16 18 1 20 152.00 [12.59; 1834.92] 3.4%
Random effects model 237 349 42 945 59.25 [23.94; 146.65] 55.4%
Heterogeneity: I^2 = 70%, τ^2 = 1.3148, p < 0.01
Indian
Aggarwal et al 2014 2 2 7 44 25.00 [1.09; 575.26] 2.4%
Khor et al 2017 2 4 4 59 13.75 [1.51; 124.99] 4.0%
Random effects model 4 6 11 103 16.76 [2.76; 101.90] 6.4%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.75
Thai
Sukasem et al 2018 12 23 4 264 70.91 [19.67; 255.64] 7.2%
Tassaneeyakul et al 2010 37 42 5 42 54.76 [14.62; 205.13] 7.0%
Random effects model 49 65 9 306 62.55 [24.93; 156.96] 14.2%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.77
Remaining populations
Amstutz et al 2013 [E,As,Af,Ab,LAC,Mix,U] 3 4 6 92 43.00 [3.86; 478.64] 3.6%
Chong et al 2013 [M,C,In] 5 6 0 9 69.67 [2.40; 2022.74] 2.1%
Genin et al 2013 [E] 0 0 20 63 0.0%
Khor et al 2017 [M only] 14 22 2 58 49.00 [9.35; 256.80] 5.7%
Nguyen et al 2015 [V] 8 14 0 19 51.00 [2.57; 1011.25] 2.6%
Ramirez et al 2017 [O] 1 1 1 24 47.00 [1.28; 1722.11] 1.9%
Toh et al 2014 [M,C] 13 16 0 23 181.29 [8.69; 3781.66] 2.5%
Yuliwulandari et al 2017 [J/S] 8 12 4 17 6.50 [1.26; 33.58] 5.7%
Random effects model 52 75 33 305 30.87 [12.81; 74.37] 24.0%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.48
Random effects model 342 495 95 1659 45.55 [26.52; 78.23] 100.0%
Heterogeneity: I^2 = 46%, τ^2 = 0.6323, p = 0.01
0.001 0.1 1 10 1000

B

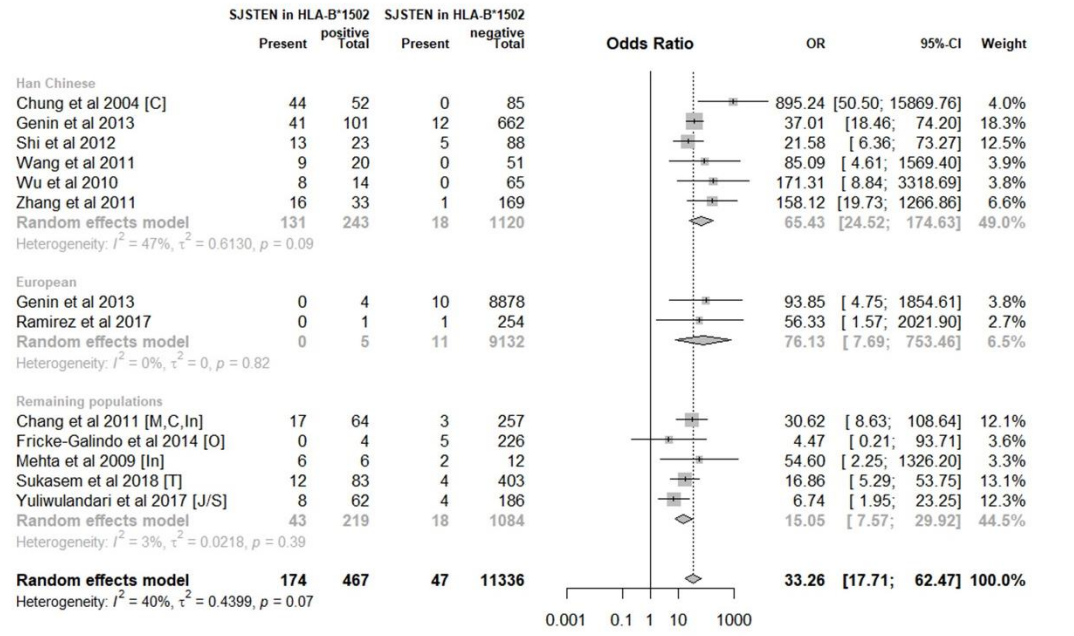

SJSTEN in HLA-B*1502 positive
SJSTEN in HLA-B*1502 negative
Present Total Present Total
Odds Ratio
OR 95%-CI Weight
Han Chinese
Chung et al 2004 [C] 44 52 0 85 895.24 [50.50; 15869.76] 4.0%
Genin et al 2013 41 101 12 662 37.01 [18.46; 74.20] 18.3%
Shi et al 2012 13 23 5 88 21.58 [6.36; 73.27] 12.5%
Wang et al 2011 9 20 0 51 85.09 [4.61; 1569.40] 3.9%
Wu et al 2010 8 14 0 65 171.31 [8.84; 3318.69] 3.8%
Zhang et al 2011 16 33 1 169 158.12 [19.73; 1266.86] 6.6%
Random effects model 131 243 18 1120 65.43 [24.52; 174.63] 49.0%
Heterogeneity: I^2 = 47%, τ^2 = 0.6130, p = 0.09
European
Genin et al 2013 0 4 10 8878 93.85 [4.75; 1854.61] 3.8%
Ramirez et al 2017 0 1 1 254 56.33 [1.57; 2021.90] 2.7%
Random effects model 0 5 11 9132 76.13 [7.69; 753.46] 6.5%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.82
Remaining populations
Chang et al 2011 [M,C,In] 17 64 3 257 30.62 [8.63; 108.64] 12.1%
Fricke-Galindo et al 2014 [O] 0 4 5 226 4.47 [0.21; 93.71] 3.6%
Mehta et al 2009 [In] 6 6 2 12 54.60 [2.25; 1326.20] 3.3%
Sukasem et al 2018 [T] 12 83 4 403 16.86 [5.29; 53.75] 13.1%
Yuliwulandari et al 2017 [J/S] 8 62 4 186 6.74 [1.95; 23.25] 12.3%
Random effects model 43 219 18 1084 15.05 [7.57; 29.92] 44.5%
Heterogeneity: I^2 = 3%, τ^2 = 0.0218, p = 0.39
Random effects model 174 467 47 11336 33.26 [17.71; 62.47] 100.0%
Heterogeneity: I^2 = 40%, τ^2 = 0.4399, p = 0.07
0.001 0.1 1 10 1000

*Figure 4.8 – Meta-analyses of the association between HLA-B*15:02 and carbamazepine-induced SJS/TEN. Studies are divided by ancestry group where >1 paper in that ancestry was located. Papers that reported participants of multiple ancestries, but did not break down results by ancestry, are reported under 'Remaining populations'. A) An meta-analysis comparing SJS/TEN cases and carbamazepine-tolerant controls. B) A meta-analysis comparing SJS/TEN cases and healthy controls (not exposed to carbamazepine). Ethnicity codes in remaining populations: As = Asian, Af = African, Ab = Aboriginal, C = Han Chinese, E = European, In = Indian, J/S = Javanese/Sundanese, LAC = Latin American/Caribbean, M = Malay, Mix = Mixed, O = Other, T = Thai, U = Unknown, V = Vietnamese.*

In the meta-analysis of cases compared to carbamazepine-tolerant controls, there were 10 studies that included Han Chinese patients (n=1294). The pooled OR for this group was 59.25 (95% CI 23.94 – 146.55), with high heterogeneity (I^2 = 70%). A further 2 studies included Indian patients (n=109). The pooled OR in this group was 16.76 (95% CI 2.76 – 101.90). Heterogeneity was 0%. Two studies included Thai patients (n=371). The OR in this group was 62.55 (95% CI 24.93 – 156.96), with I^2 = 0%. The final group of patients represented all remaining populations, and those where details of individual populations could not be separated from overall summary statistics (n=380). The pooled OR in this group was 30.87 (95% CI 12.81 – 74.37), with I^2 = 0%.

The overall OR for all populations in the carbamazepine-tolerant controls meta-analysis was 45.55 (95% CI 26.52 – 78.23) with low to moderate heterogeneity (I^2 = 46%).

In the meta-analysis of cases compared to healthy volunteer controls, there were 6 studies that included Han Chinese patients (n=1363). The pooled OR for this group was 65.43 (95% CI 24.52 – 174.63) with low to moderate heterogeneity (I^2 = 47%). Two further studies included European patients (n=9137). The pooled OR in this group was 76.13 (95% CI 7.69 – 753.46, and the I^2 was 0%. The remaining populations across several ancestry groups produced an OR of 15.05 (95% CI 7.57 – 29.92), with low heterogeneity (I^2 = 3%).

The overall OR for all populations in the healthy volunteer controls meta-analysis was 33.26 (95% CI 17.71 – 62.47), with low to moderate heterogeneity (I^2 = 40%).

4.3.1.1.1 Comparison to other meta-analyses

Results were compared to the three previous meta-analyses in *HLA-B*15:02* and carbamazepine-induced SJS/TEN, which I had used to help inform the search strategy (2, 76, 77). The comparison is only in carbamazepine-tolerant controls compared to cases.

The overall OR calculated in this meta-analysis is smaller than any of the previous meta-analyses of cases compared to carbamazepine-tolerant controls. The 95% CI is also narrower in this analysis (Figure 4.9). The confidence intervals of the estimate overlap with the confidence intervals of previously published works.

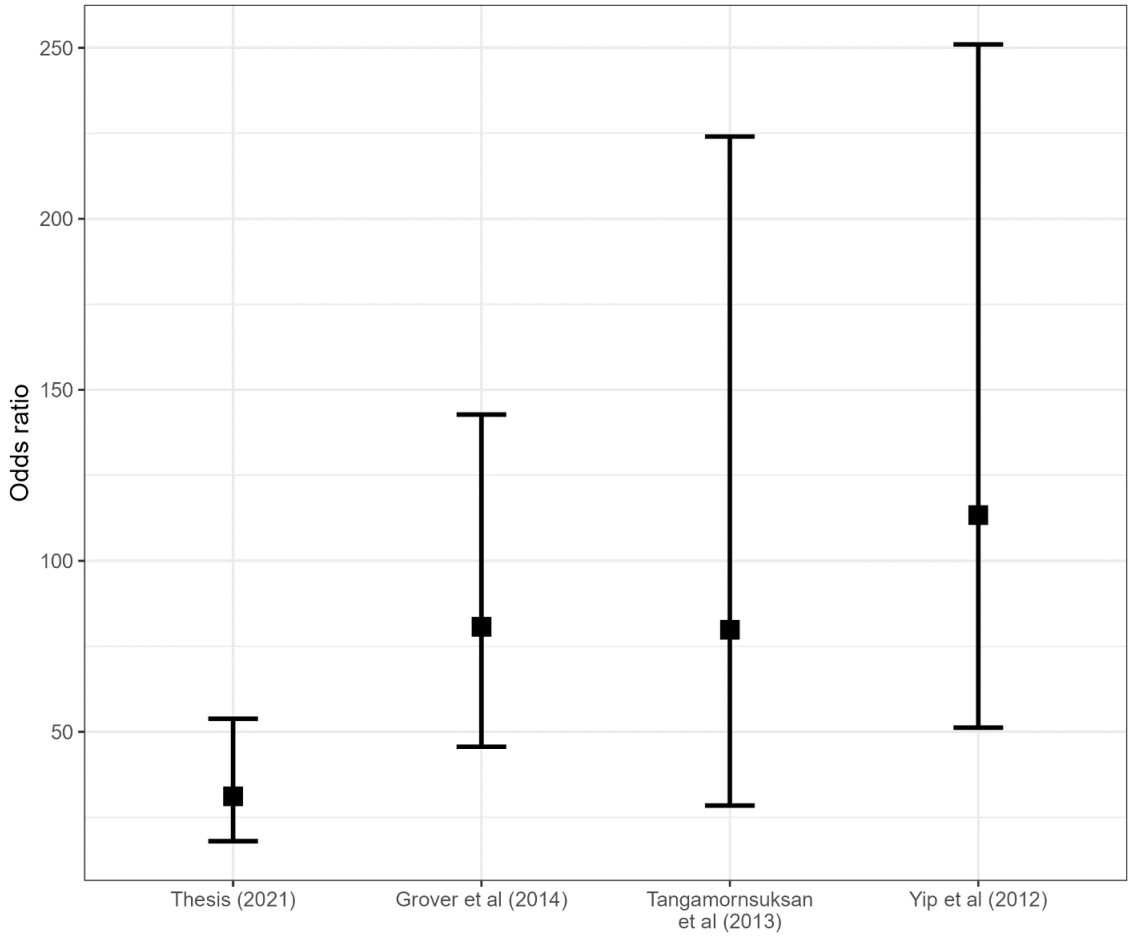


Figure 4.9 - comparison of odds ratios obtained in current meta-analysis of cases vs carbamazepine-tolerant controls in carbamazepine-induced SJS/TEN, compared to previously published meta-analyses.

Similarly to previous meta-analyses, the majority of my estimate (55.4% weighting) came from the Han Chinese population, however there was a smaller and narrower confidence interval of OR (31.15, 95% CI 18.03 – 53.82) compared to Grover et al. (OR 80.70, 95% CI 45.62 – 142.77), Tangamornsuksan et al. (OR 79.84, 95% CI 28.45 – 224.06) and Yip, et al. (OR 113.39, 95% CI 51.24 – 250.97). These differences can be explained by the inclusion of more and newer data in this meta-analysis, and some differences in the eligibility criteria that resulted in different papers being included in each meta-analysis. For example, Then, *et al.* (2011) was included in the Tangamornsuksan, *et al.* meta-analysis but excluded from ours as it

only reported SJS, not SJS/TEN. This study has a wide confidence interval around its estimate (OR 221.0, 95% CI 3.85 – 12694.65), with a 5% weighting on the final meta-analysed estimate of the OR. Other similar instances are seen across other excluded papers. A full comparison of the papers included in this meta-analysis compared to papers included in previous meta-analyses is shown in Appendix Table 8. One paper included in two meta-analyses was excluded in this meta-analysis at the screening stage, as per the inclusion and exclusion criteria, as it is a meeting abstract (18). This was felt to not have enough information to assess study quality and bias. Other papers were excluded for reporting SJS and TEN as separate outcomes since these are clinically recognised as different outcomes to SJS/TEN (107, 112, 127). for having no *HLA-B*15:02* positive patients (43, 118) (these were excluded prior to forest plot stage, while other meta-analyses included them in the forest plot but assigned them 0% weighting), and for having cases overlapping with another, included paper (109).

4.3.1.1.2 Comparison to only papers published prior to prospective study

Next, only papers that were available before recruitment to the prospective Chen, *et al.* 2011 paper began recruiting (July 2007) (3) were analysed. Four papers published before 1st July 2007 were located (27, 40, 43, 114). However, one of these did not break down their results by drug (114). The authors were contacted but no reply was received. Another had no *HLA-B*15:02* positive patients so was unable to be included in the meta-analysis (43). The remaining two papers had overlaps in their included patients (27, 40). Only one of these (the larger and more recent (40)) would therefore be included in the meta-analysis.

I therefore decided to analyse all papers published up to the time Chen et al 2011 was published (March 2011) (3), rather than the time recruitment started. There were seven unique papers for inclusion in this meta-analysis (27, 40, 116, 125, 129, 130, 133). Five of these included carbamazepine-tolerant patients as controls (Figure 4.10A) (40, 125, 129, 130, 133), and five included healthy volunteers as controls (Figure 4.10B) (27, 116, 129, 130, 133).

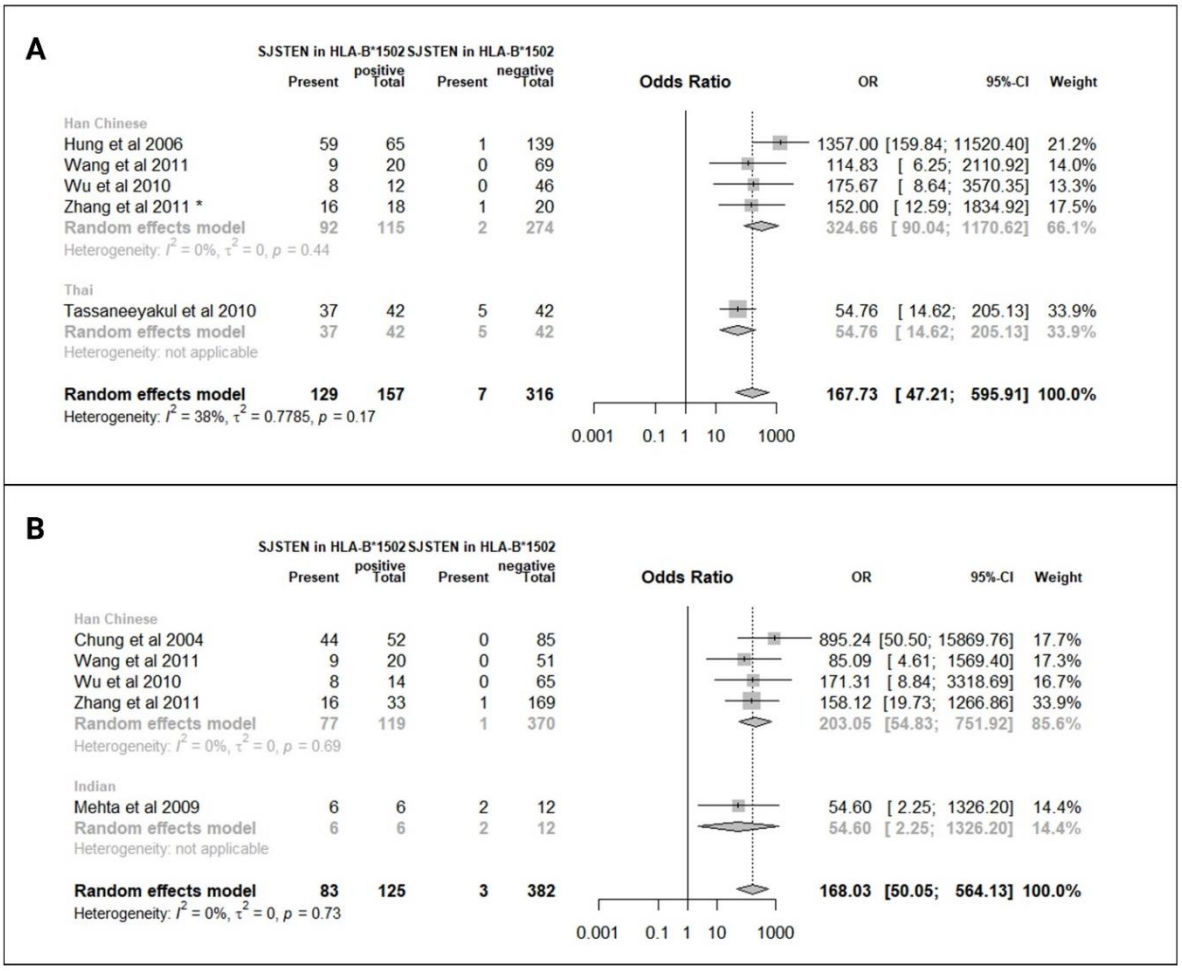


*Figure 4.10 - Meta-analysis of papers published prior to the March 2011 publication of Chen et al 2011 (3), with A) Carbamazepine-tolerant controls, B) healthy controls. * this paper was published during March 2011.*

In the meta-analysis of cases compared to carbamazepine-tolerant controls, there were 4 studies that included Han Chinese patients (n=473). The pooled OR for this group was 324.66 (95% CI 90.04– 1170.62), with I^2 = 0%. One further study included Thai patients (n=84). The OR in this study was 54.76 (95% CI 14.62 – 205.13).

The overall OR for all populations in the carbamazepine-tolerant controls meta-analysis was 167.73 (95% CI 47.21 – 595.91) with low to moderate heterogeneity (I^2 = 38%).

In the meta-analysis of cases compared to healthy volunteer controls, there were 4 studies that included Han Chinese patients (n=489). The pooled OR for this group

was 203.05 (95% CI 54.83 – 751.92) with I^2 = 0%. One further study included Indian patients (n=18). The OR in this study was 54.60 (95% CI 2.25 – 1326.20.

The overall OR for all populations in the healthy volunteer controls meta-analysis was 168.03 (95% CI 50.05 – 564.13), with I^2 = 0%.

4.3.1.1.3 Comparison to prospective study

Figure 11 shows a plot of the OR, PPV and NPV and estimated 95% confidence intervals against number of cases. Here, the point estimate of the OR is estimated from the meta-analysis of tolerant controls against cases in Han Chinese participants, only including data from prior to Chen, *et al.* (324.66). PPV and NPV were estimated as 0.0437 and 0.9999 respectively.

It is clear that the precision of the estimates where there are only 7 cases (solid orange line, the number of serious ADR cases in the Chen, *et al.* study) is much inferior to the precision gained from meta-analysis of case-control studies, where 94 cases were collected in Han Chinese participants (in meta-analysis of studies published prior to Chen, *et al.*) (Figure 4.10A). No SJS/TEN cases were collected by Chen, *et al.*, limiting the precision of their estimates. It can therefore be seen that a very large prospective study would be required to produce estimates as precise as data already collected from observational studies.

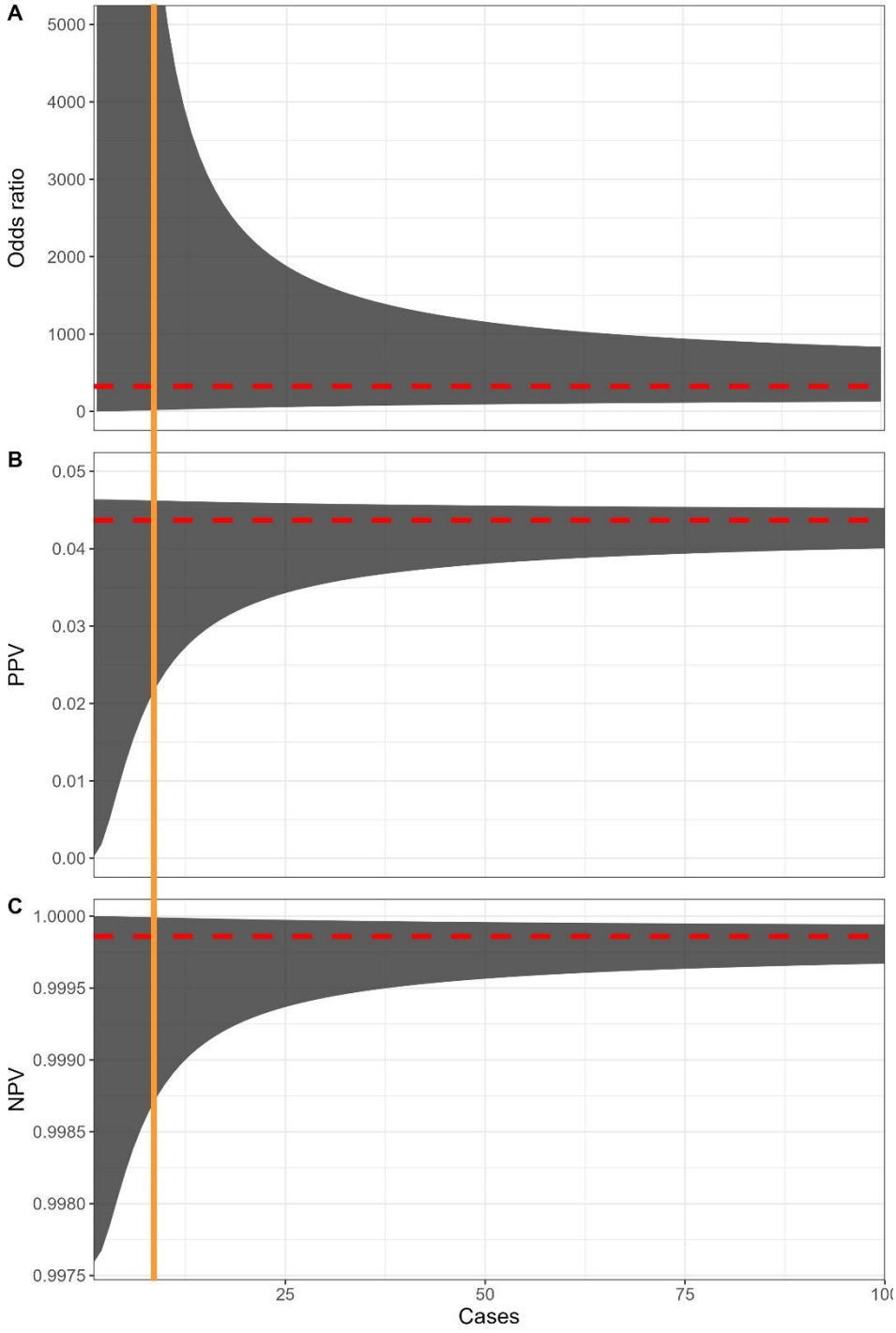


*Figure 4.11 – Change in the width of confidence intervals of the A) odds ratio, B) PPV and C) NPV (measures of clinical validity) of genetic testing for HLA-B*15:02 as the number of accrued cases of carbamazepine-induced SJS/TEN increases. The dashed red line is the point estimate of OR, PPV, and NPV calculated from the meta-analysis (OR = 167.76, PPV = 0.0437, NPV = 0.9998. The orange line is the number of cases of serious adverse drug reactions accrued in a previous prospective study (Chen, et al. 2011) (3). These diagrams adapted from work by Matt Nelson of GlaxoSmithKline (personal communication)*

4.3.2 *HLA-A*31:01* systematic review and meta-analysis

In this systematic review and meta-analysis, 83 results were obtained from searching the literature. As for the previous review, these results were checked if they contained previously-identified key references in order to validate the search strategy. After screening by title and abstract, 30 papers remained for full text screening. After this screening 24 papers remained for inclusion in the systematic review (15, 17, 40, 46, 48, 96-98, 101, 102, 104, 105, 107, 111, 115, 118-122, 128, 134-136), and 8 of these were included in the carbamazepine tolerant controls meta-analysis. A total of 4 were included in the healthy volunteers controls meta-analysis (Figure 4.12).

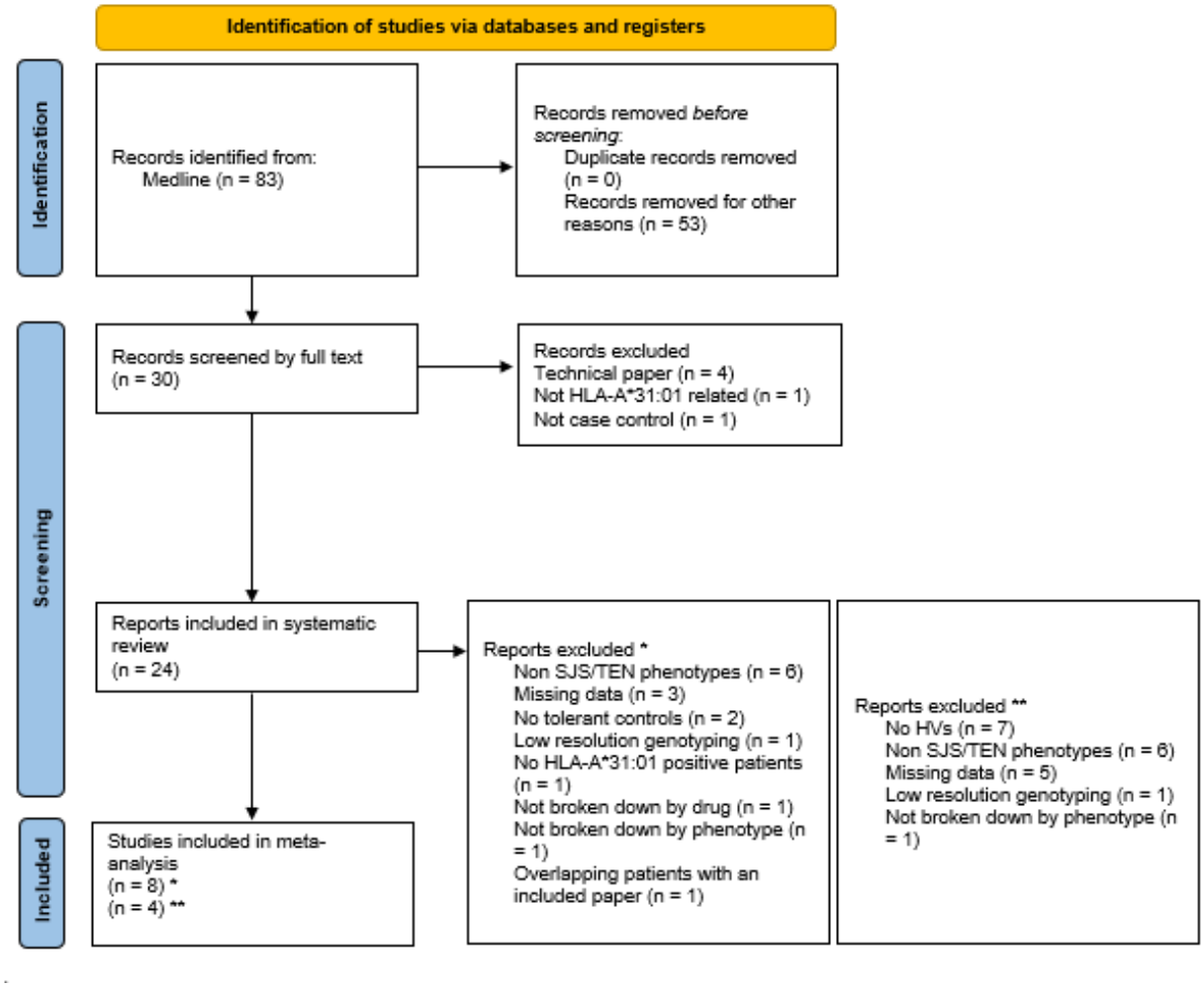
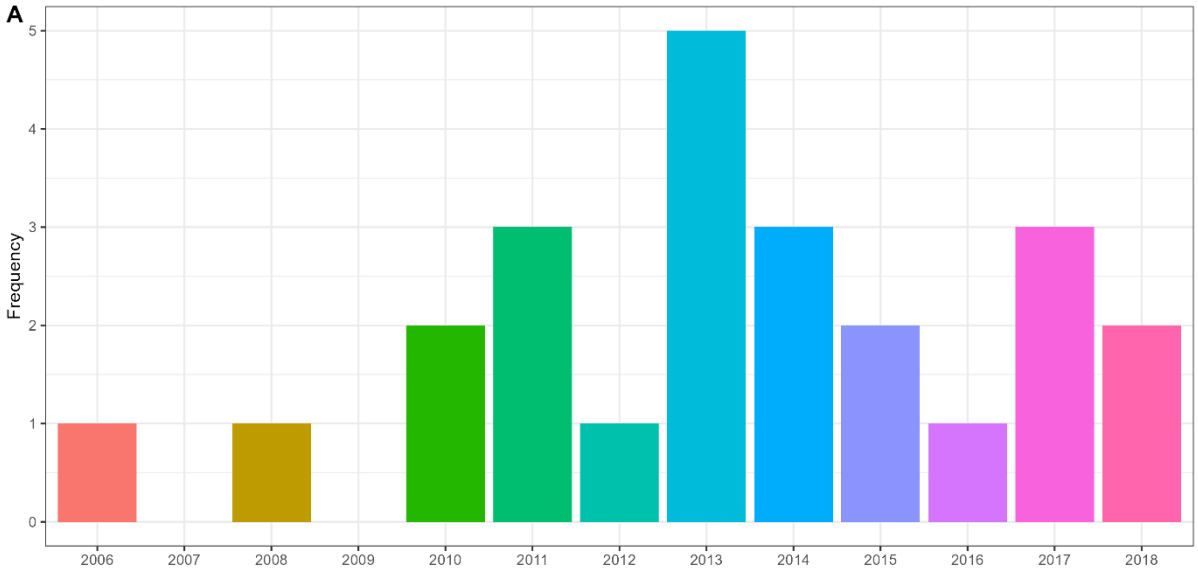
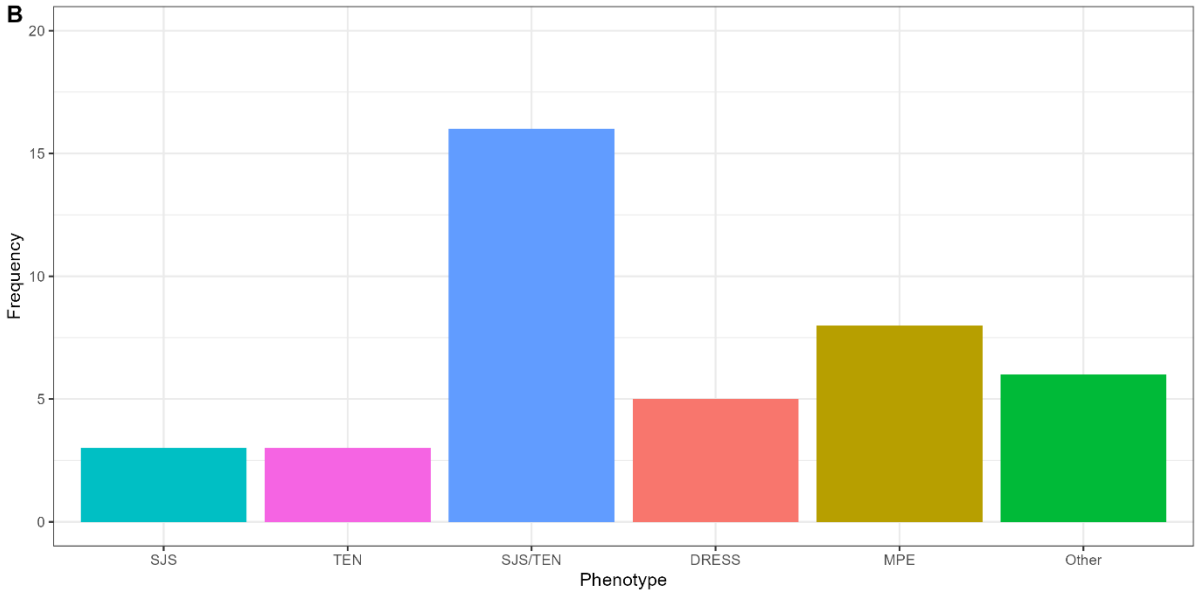


*Figure 4.12 - flow of papers included in HLA-A*31:01 systematic review and meta-analysis. * denotes meta-analysis with cases and tolerant controls. ** denotes meta-analysis with cases and healthy volunteer controls. HVs = healthy volunteers. SJS/TEN = Stevens Johnson syndrome/toxic epidermal necrolysis. Flow chart adapted from PRISMA 2020 (79).*

Full details of included papers are shown in Appendix 4.3.

Papers were published between 2006 and 2018, with most being published in 2013 (Figure 4.13A). The most common phenotype included was SJS/TEN (Figure 4.13B). The most common design was case control, tolerant controls and healthy volunteers as controls (Figure 4.13C).


A
Frequency
0
1
2
3
4
5
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018


B
Frequency
0
5
10
15
20
SJS
TEN
SJS/TEN
DRESS
MPE
Other
Phenotype

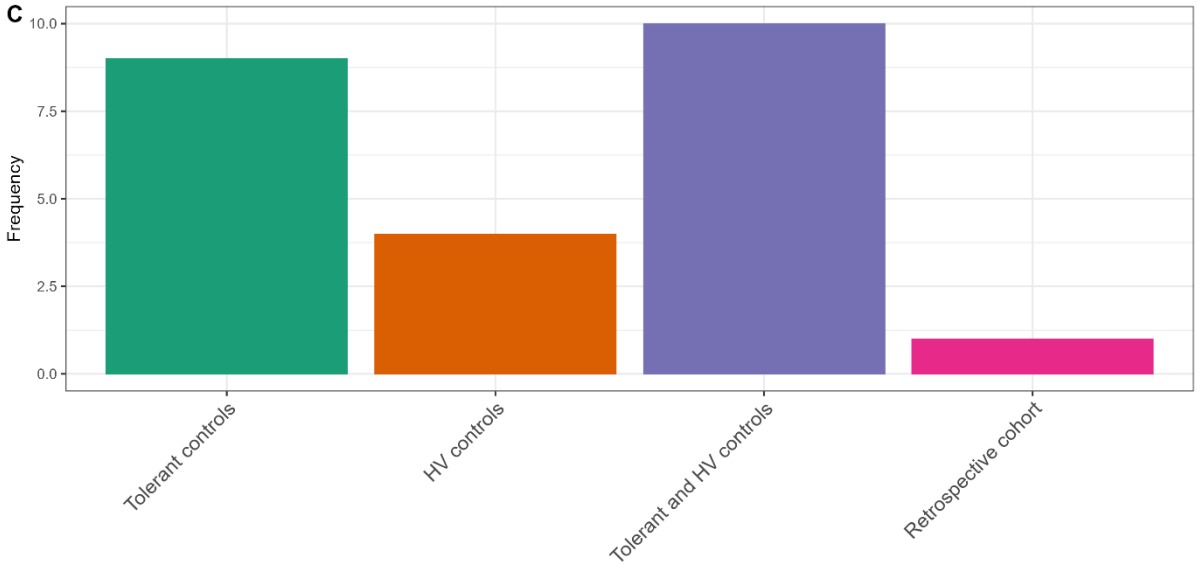

C
Frequency
0.0
2.5
5.0
7.5
10.0
Tolerant controls
HV controls
Tolerant and HV controls
Retrospective cohort

*Figure 4.13 - A) Number of papers in HLA-A*31:01 systematic review by year of publication. B) Phenotypes included in papers in systematic review. Some papers included more than one phenotype. C) Designs of papers included in the systematic review. Case control refers to papers that used carbamazepine-tolerant patients as controls. HV = healthy volunteers*

The mean total sample size was 1242 (SD 2302.4) and the median total sample size was 334 (IQR 194 – 764). A total of 29,805 participants were included, of which 1678 were cases, 9606 were controls, and 18,521 were healthy volunteers. The mean age of cases (39.3, SD 13.0), controls (36.8, SD 12.5), and healthy volunteers (43.8, SD 10.5) were similar. Cases and controls had a fairly even gender split (cases 47.4% male, controls 54.5% male), but healthy volunteers were less balanced (37.4% male).

The quality of included papers was assessed according to the criteria of Jorgensen and Williamson (2008) (85) (Figure 4.14). A pattern broadly similar to the papers of the *HLA-B*15:02* papers was observed. Very few papers included assessments of Hardy-Weinberg equilibrium, an explanation of how missing data was explored or

dealt with, and how adherence was assessed and adjusted for (Figure 4.14).

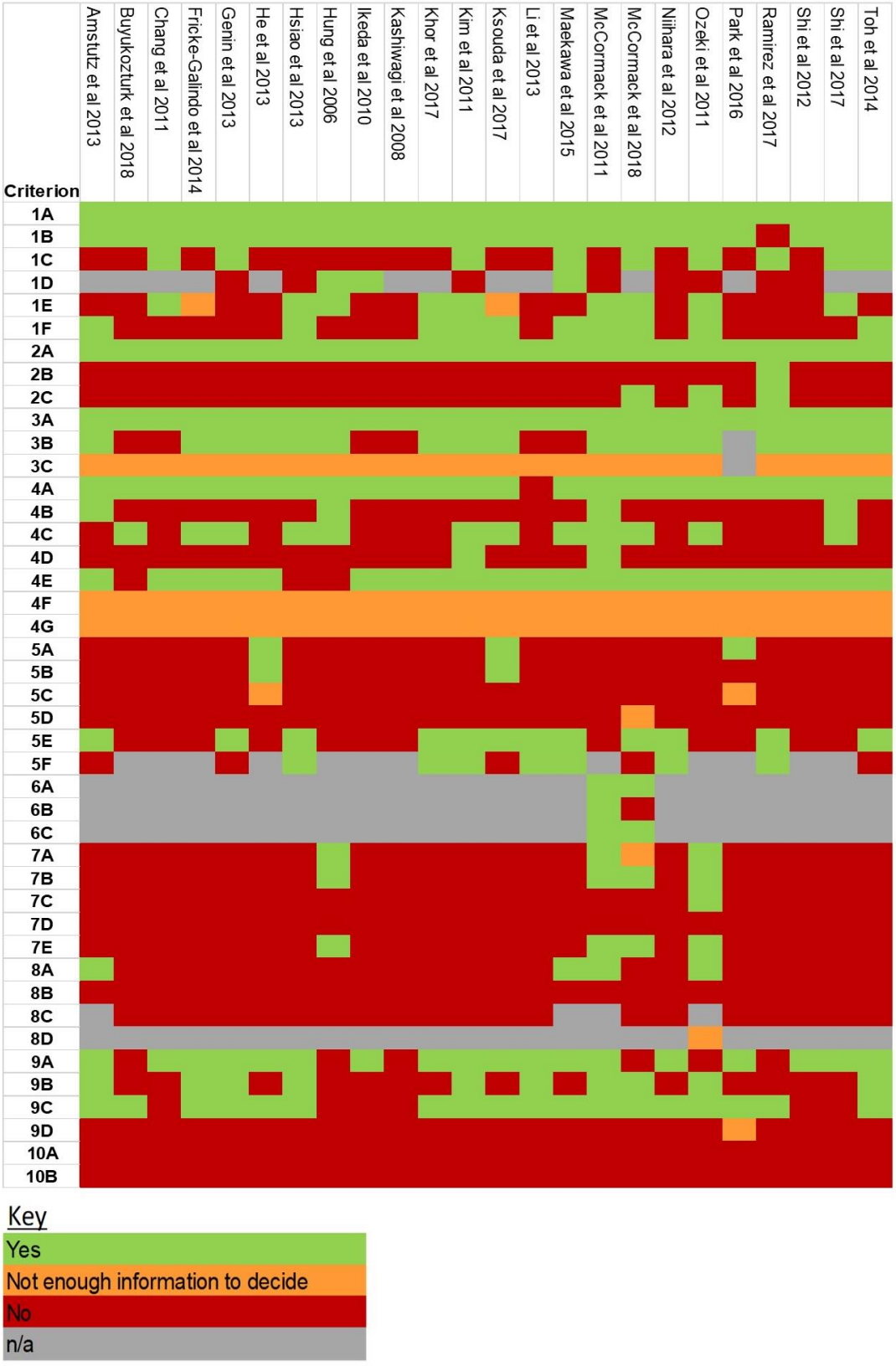


*Figure 4.14 - quality of studies included in HLA-A*31:01 systematic review, according to criteria of Jorgensen and Williamson (2008) (85). Full criteria and data are included in Appendix 4.4.2.*

*4.3.2.1 HLA-A*31:01 meta-analysis*

All papers up to February 2020 (date of the search) were included in the meta-analysis of *HLA-A*31:01* and CBZ-induced SJS/TEN.

A total of 8 papers that compared SJS/TEN cases to CBZ-tolerant controls were located (Figure 4.15A) (46, 48, 96, 102, 105, 119, 120, 122). Four papers that compared SJS/TEN cases to healthy controls were located (Figure 4.15B) (48, 98, 120, 136).

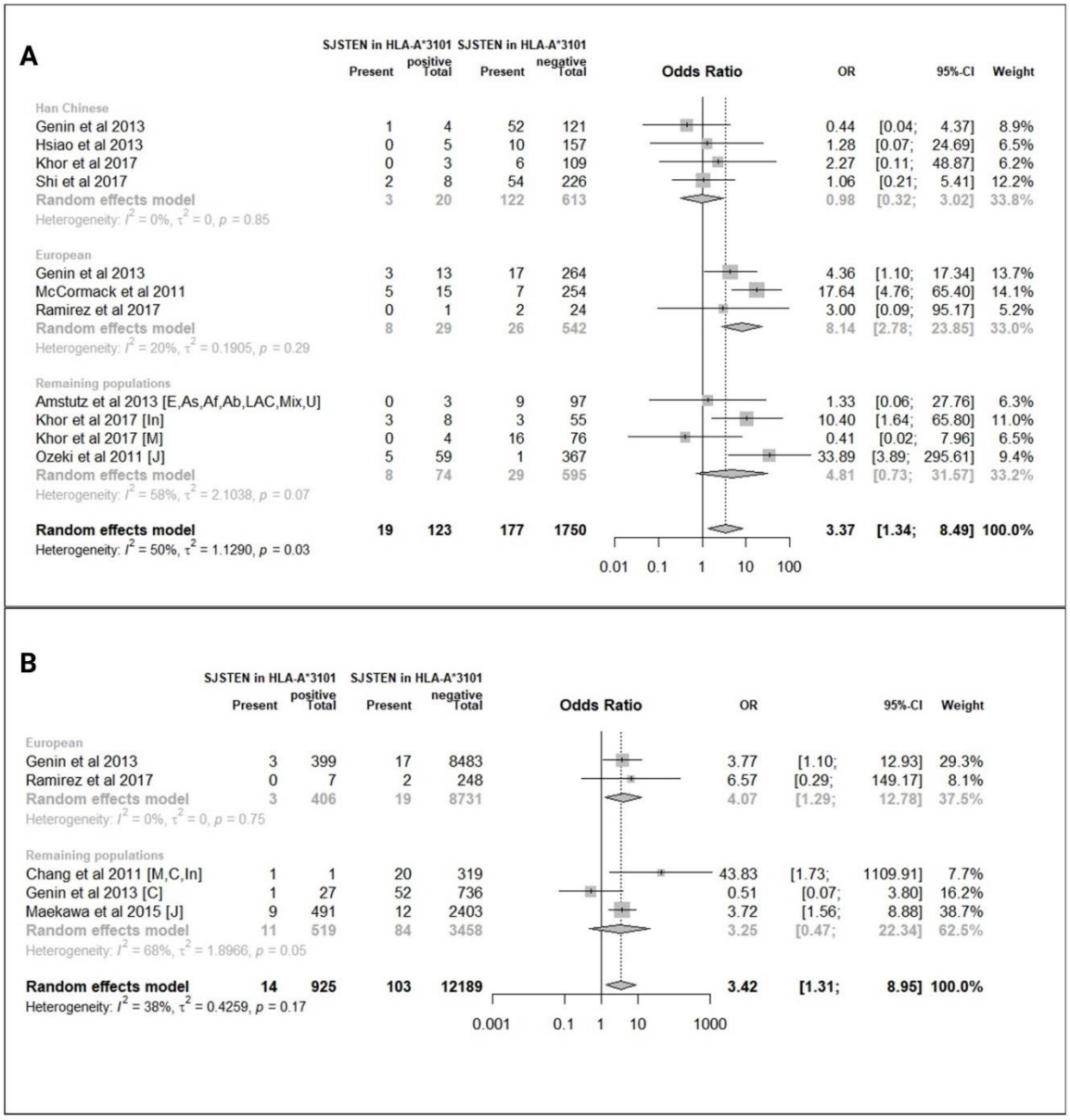


*Figure 4.15 - Meta-analysis of all papers that examined the risk of carbamazepine-induced SJS/TEN associated with HLA-A*31:01 compared to A) carbamazepine-tolerant controls and B) healthy controls (not taking carbamazepine). Ethnicity codes: As = Asian, Af = African, Ab = Aboriginal, C = Han*

Chinese, E = European, In = Indian, J = Japanese, LAC = Latin American/Caribbean, M = Malay, Mix = Mixed, U = Unknown.

In the meta-analysis of cases compared to carbamazepine-tolerant controls, there were 4 studies that included Han Chinese patients (n=633). The pooled OR for this group was 0.98 (95% CI 0.32 – 3.02), with I^2 = 0%. There were 3 studies that included European patients (n=571). The pooled OR in this group was 8.14 (95% CI 2.78 – 23.85) with low heterogeneity (I^2 = 20%). The final group of patients represented all remaining populations, and those where details of individual populations could not be separated from overall summary statistics (n=669). The pooled OR in this group was 4.81 (0.73 – 31.57) with moderate heterogeneity (I^2 = 58%).

The overall OR for all populations in the carbamazepine-tolerant controls meta-analysis was 3.37 (95% CI 1.34 – 8.49) with moderate heterogeneity (I^2 = 50%).

In the meta-analysis of cases compared to healthy volunteer controls, there were 2 studies that included European patients (n=9137). The pooled OR for this group was 4.07 (95% CI 1.29 – 12.78) with I^2 = 0%. Three further studies included patients from remaining populations (n=3977). The pooled OR in this group was 3.25 (95% CI 0.48 – 22.34) with moderate heterogeneity (I^2 = 68%).

The overall OR for all populations in the healthy volunteer controls meta-analysis was 3.42 (95% CI 1.31 – 8.95), with low to moderate heterogeneity (I^2 = 38%).

Outcomes other than SJS/TEN were also analysed for *HLA-A*31:01*. Only outcomes where more than one paper was located with sufficient data were analysed.

For SJS alone, comparing to carbamazepine-tolerant controls, there were 2 studies, in Han Chinese and Korean patients (n=305) (Figure 4.16A). The OR in this group was 1.68 (95% CI 0.23 – 12.15), with moderate heterogeneity (I^2 = 64%).

For DRESS, there were sufficient papers to compare cases to carbamazepine-tolerant controls (Figure 4.16B) and to healthy volunteers (Figure 4.16C). In the carbamazepine-tolerant control meta-analysis, there were 2 studies in European patients (n=294). The pooled OR for this group was 46.59 (95% CI 12.47 – 173.99), with I^2 = 0%. There were 2 further studies in Han Chinese patients (n=257). The pooled OR in this group was 15.81 (95% CI 5.76 – 43.43), with I^2 = 0%. There was one remaining paper in this group, in Tunisian patients (n=32). The OR in this study was 32.0 (95% CI 2.63 – 389.25). The overall OR for all populations within the

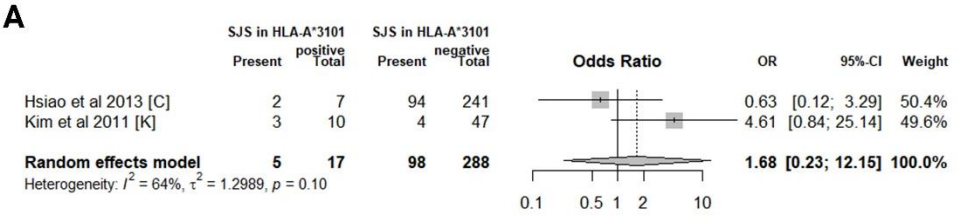
carbamazepine-tolerant controls meta-analysis was 24.27 (95% CI 11.31 – 52.08), with I^2 = 0%.

In the healthy volunteer controls meta-analysis there were 2 studies in European patients (n=9129). The pooled OR in this group was 44.99 (95% CI 14.40 – 140.57), with I^2 = 0%. There was 1 further study, in Han Chinese patients. The OR in this study was 26.31 (95% CI 7.17 – 96.53). The overall OR for all populations with healthy volunteer controls was 35.64 (95% CI 15.13 – 83.95), with I^2 = 0%.

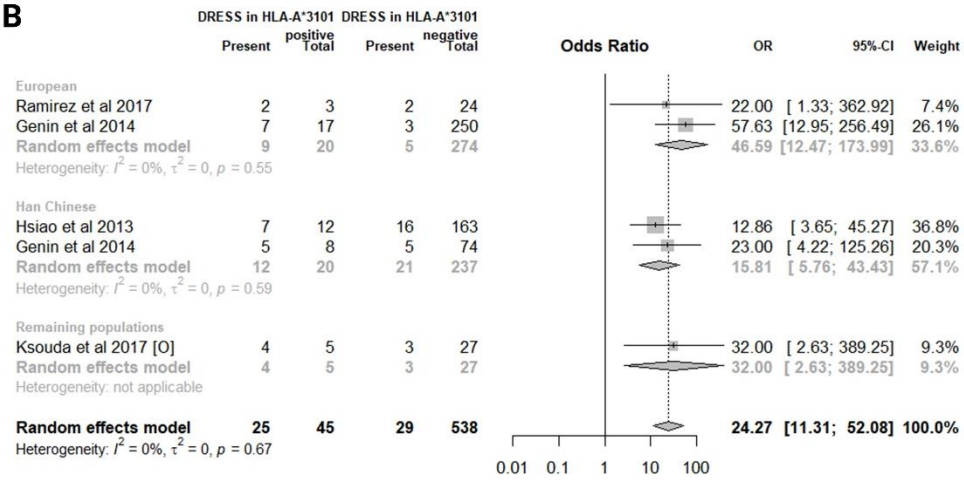
Finally, the MPE reaction was also evaluated, comparing cases to carbamazepine-tolerant controls (Figure 4.16D). There were 2 studies in Han Chinese patients (n=483). The pooled OR in this group was 2.78 (95% CI 0.86 – 9.03) with I^2 = 17%. There were 3 studies in remaining populations (n=1103). The pooled OR in this group was 6.98 (95% CI 3.86 – 12.62), with I^2 = 0%. The overall OR for all populations with carbamazepine-tolerant controls was 5.51 (95% CI 3.16 – 9.62), with low heterogeneity (I^2 = 6%).

There were insufficient papers to meta-analyse for other outcomes (TEN alone, and SJS and MPE compared to healthy volunteers).

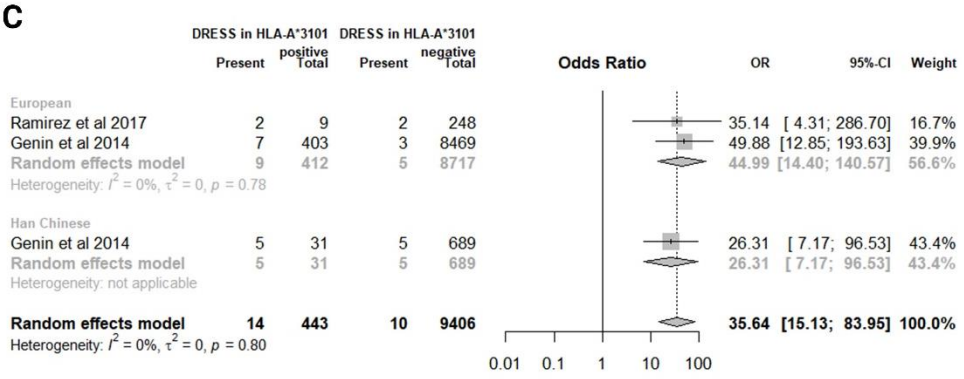
A


SJS in HLA-A*3101 positive
SJS in HLA-A*3101 negative
Present Total Present Total
Odds Ratio
OR 95%-CI Weight
Hsiao et al 2013 [C] 2 7 94 241 0.63 [0.12; 3.29] 50.4%
Kim et al 2011 [K] 3 10 4 47 4.61 [0.84; 25.14] 49.6%
Random effects model 5 17 98 288 1.68 [0.23; 12.15] 100.0%
Heterogeneity: I^2 = 64%, τ^2 = 1.2989, p = 0.10
0.1 0.5 1 2 10

B


DRESS in HLA-A*3101 positive
DRESS in HLA-A*3101 negative
Present Total Present Total
Odds Ratio
OR 95%-CI Weight
European
Ramirez et al 2017 2 3 2 24 22.00 [1.33; 362.92] 7.4%
Genin et al 2014 7 17 3 250 57.63 [12.95; 256.49] 26.1%
Random effects model 9 20 5 274 46.59 [12.47; 173.99] 33.6%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.55
Han Chinese
Hsiao et al 2013 7 12 16 163 12.86 [3.65; 45.27] 36.8%
Genin et al 2014 5 8 5 74 23.00 [4.22; 125.26] 20.3%
Random effects model 12 20 21 237 15.81 [5.76; 43.43] 57.1%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.59
Remaining populations
Ksouda et al 2017 [O] 4 5 3 27 32.00 [2.63; 389.25] 9.3%
Random effects model 4 5 3 27 32.00 [2.63; 389.25] 9.3%
Heterogeneity: not applicable
Random effects model 25 45 29 538 24.27 [11.31; 52.08] 100.0%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.67
0.01 0.1 1 10 100

C


DRESS in HLA-A*3101 positive
DRESS in HLA-A*3101 negative
Present Total Present Total
Odds Ratio
OR 95%-CI Weight
European
Ramirez et al 2017 2 9 2 248 35.14 [4.31; 286.70] 16.7%
Genin et al 2014 7 403 3 8469 49.88 [12.85; 193.63] 39.9%
Random effects model 9 412 5 8717 44.99 [14.40; 140.57] 56.6%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.78
Han Chinese
Genin et al 2014 5 31 5 689 26.31 [7.17; 96.53] 43.4%
Random effects model 5 31 5 689 26.31 [7.17; 96.53] 43.4%
Heterogeneity: not applicable
Random effects model 14 443 10 9406 35.64 [15.13; 83.95] 100.0%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.80
0.01 0.1 1 10 100

D

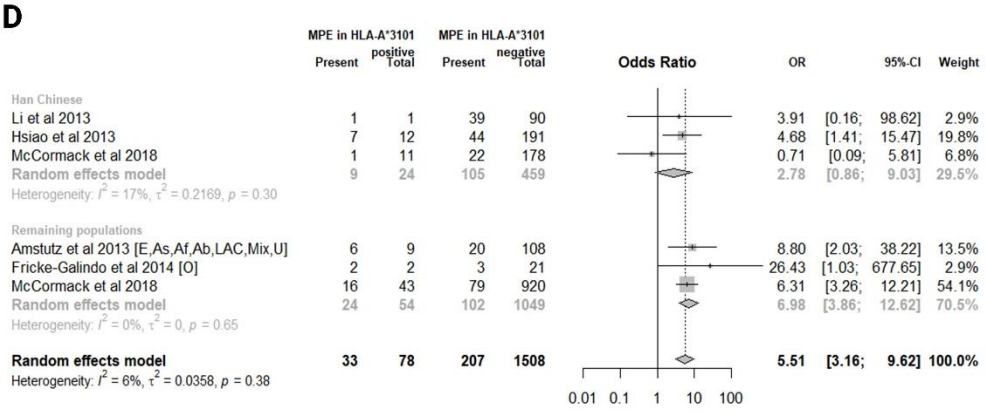

MPE in HLA-A*3101 positive
MPE in HLA-A*3101 negative
Present Total Present Total
Odds Ratio
OR 95%-CI Weight
Han Chinese
Li et al 2013 1 1 39 90 3.91 [0.16; 98.62] 2.9%
Hsiao et al 2013 7 12 44 191 4.68 [1.41; 15.47] 19.8%
McCormack et al 2018 1 11 22 178 0.71 [0.09; 5.81] 6.8%
Random effects model 9 24 105 459 2.78 [0.86; 9.03] 29.5%
Heterogeneity: I^2 = 17%, τ^2 = 0.2169, p = 0.30
Remaining populations
Amstutz et al 2013 [E,As,Af,Ab,LAC,Mix,U] 6 9 20 108 8.80 [2.03; 38.22] 13.5%
Fricke-Galindo et al 2014 [O] 2 2 3 21 26.43 [1.03; 677.65] 2.9%
McCormack et al 2018 16 43 79 920 6.31 [3.26; 12.21] 54.1%
Random effects model 24 54 102 1049 6.98 [3.86; 12.62] 70.5%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.65
Random effects model 33 78 207 1508 5.51 [3.16; 9.62] 100.0%
Heterogeneity: I^2 = 6%, τ^2 = 0.0358, p = 0.38
0.01 0.1 1 10 100

*Figure 4.16 - A) A) SJS in HLA-A*31:01 only, case control (there were no papers with SJS and HVs). B) DRESS in HLA-A*31:01 only, case control. C) DRESS in HLA-A*31:01 only, HVs. D) MPE in HLA-A*31:01 only, case control (no papers with MPE and HVs).*

4.3.2.1.1 Comparison to simulated prospective study

I simulated a prospective study that used similar estimates as Chen, *et al.* The simulation, run 10,000 times with n=5000 patients each time, estimated that a mean of 3 cases of SJS/TEN would be detected if pre-treatment genotyping was introduced.

Figure 4.17 shows a plot of the OR, PPV, and NPV and their estimated 95% confidence intervals against number of SJS/TEN cases. PPV and NPV were estimated as 0.00735 and 0.99781 respectively.

The precision of the estimates where there are only 3 cases (solid orange line, the number of SJS/TEN cases in the simulation study) is much inferior to the precision gained from meta-analysis of case-control studies, where 125 cases were collected in Han Chinese participants (Figure 4.16A). It can therefore be seen that a very large prospective study would be required to produce estimates as precise as the data already collected from observational studies.

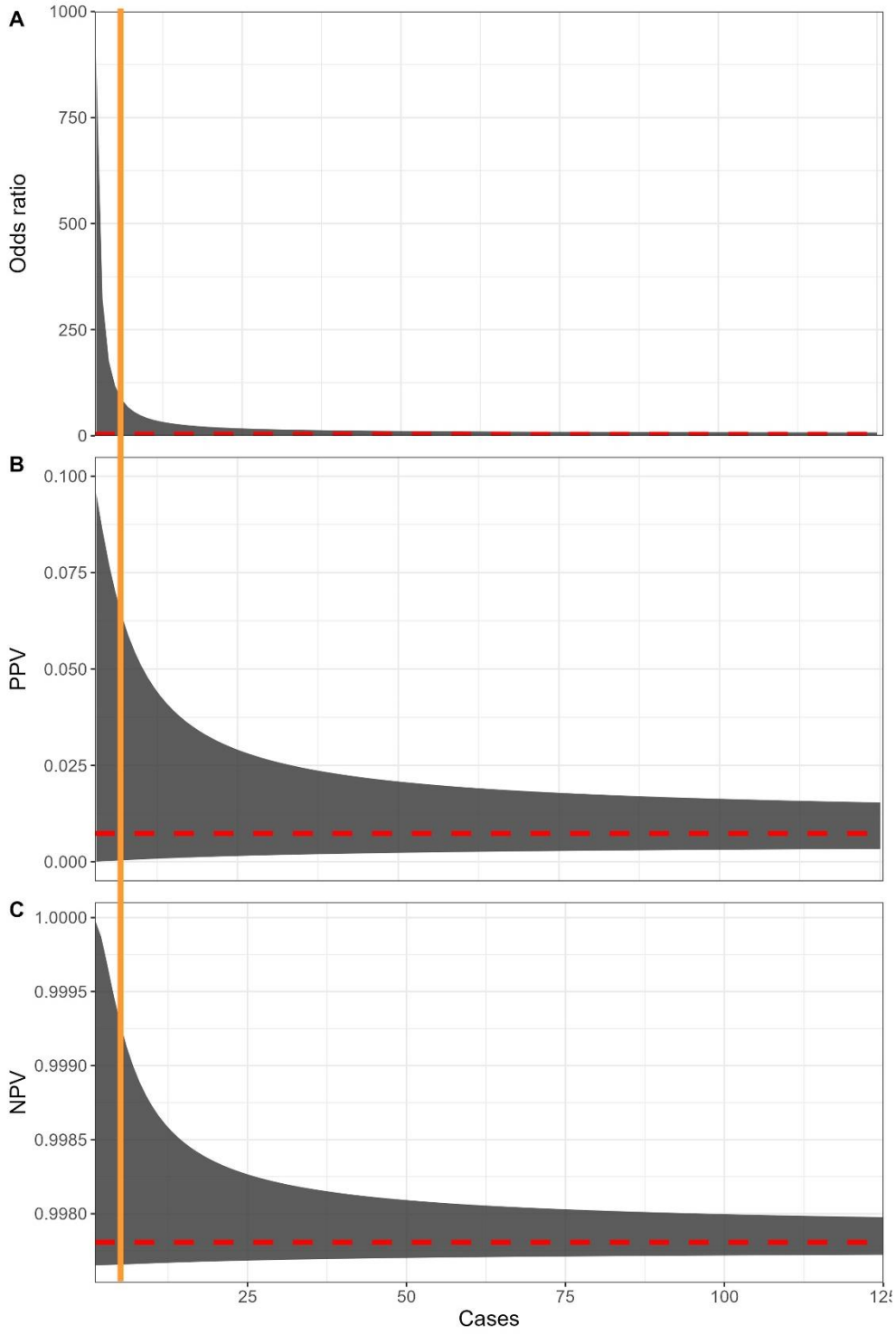


*Figure 4.17 – Change in the width of confidence intervals of the A) odds ratio, B) PPV and C) NPV (measures of clinical validity) of genetic testing for HLA-A*31:01 as the number of accrued cases of carbamazepine-induced SJS/TEN increases. The dashed red line is the point estimate of OR, PPV, and NPV calculated from the meta-analysis (OR = 3.37, PPV = 0.0074, NPV = 0.9978. The orange line is the number of cases of SJS/TEN accrued by a previous simulated prospective study, after recruiting 5000 participants total. These diagrams adapted from work by Matt Nelson of GlaxoSmithKline (personal communication*

4.4 Discussion

This chapter presents two novel systematic reviews and meta-analyses of rare ADRs, and demonstrates how meta-analysis of observational data can provide more precise estimates of clinical validity as compared to those obtained from prospective interventional studies.

Genotyping for *HLA-B*15:02* before carbamazepine prescription already has strong evidence for its clinical use, and is recommended by several national regulatory bodies (137). Genotyping has been shown in a prospective trial to significantly reduce the risk of SJS/TEN (3). However, evidence from case-control studies conducted prior to the trial can be used to estimate clinical validity of genotype-guided prescribing, with good precision (OR 324.66 [95% CI 90.04 to 1170.62], PPV 0.0437 [0.0399 to 0.0453], NPV 0.9998 [0.9997 to 0.9999]). Low PPV is characteristic of tests with very rare outcomes and remains low even when associations (ORs) are high (67).

The evidence available for genotyping for *HLA-A*31:01* is less strong, despite its higher frequency across more populations compared to *HLA-B*15:02* (see Figure 4.2). There is much less research into *HLA-A*31:01* and SJS/TEN. Meta-analyses confirm the increased risk of SJS/TEN with this allele (OR 3.37 in tolerant controls, OR 3.42 in healthy volunteer controls). These analyses consist of 1,873 and 13,114 participants, respectively. I have shown that these data can provide precise estimates of clinical validity (OR 3.37 [95% CI 1.34 to 8.49], PPV 0.0074 [0.0034 to 0.0154], NPV 0.9978 [0.9977 to 0.9980]).

This simulated prospective study of *HLA-A*31:01* genotyping shows that if 5000 participants were recruited to a prospective, interventional study with the same design as Chen et al. (3), approximately 3 cases of SJS/TEN would be observed. With this small number of cases, the estimates of clinical validity (OR, PPV, and NPV) are far less precise than those obtained from meta-analysis of observational data.

Testing for *HLA-A*31:01* prior to carbamazepine prescription was shown to be cost-effective in a UK setting in 2015 (24). Using a Markov model, the authors showed that the reduction in risk of ADRs (from 780 per 10,000 to 700 per 10,000), and subsequent effect on quality of life, meant that the initial cost of genotyping was more than recouped over a lifetime. This information, combined with my analysis,

provides quantitative evidence for a benefit of *HLA-A*31:01* screening prior to carbamazepine use.

There was some difficulty encountered in extracting data from studies. Where I attempted to contact authors for further data or clarification, I received very few responses. Wider engagement of authors with readers of publications would benefit meta-analyses, and is of even greater importance in individual patient data meta-analyses. Accounting for overlap between patients across studies was another challenge. Other meta-analyses have also dealt with this issue. For example, a meta-analysis by Tangarosan, *et al.* states that patients in Wang, *et al.* (2011) overlap with other papers. I found no mention of this in the Wang paper, or the others mentioned. There may be some, unmentioned overlap, but without clear reporting this cannot be picked up by systematic reviewers.

One prospective interventional study (similar to Chen, *et al.*) has been conducted into the association between *HLA-A*31:01* and SJS/TEN. This study, which only recruited Japanese patients, found a benefit of *HLA-A*31:01* screening in this population (138). A total of 1130 patients were genotyped for HLA-A*31:01 and those who tested positive (17.5%) were recommended non-carbamazepine drugs. No cases of SJS/TEN were observed in the cohort, which was compared to a historical incidence calculated from Japanese BioBank participants (OR 0.60, 95% CI 0.36 – 1.0, *p=0.048*). The authors predicted that 3 cases of SJS/TEN had been prevented by introducing pre-treatment genetic testing. This paper had not been published when this chapter was planned, hence why I chose to compare the meta-analysis results to a simulated study. The fact that no SJS/TEN cases were detected means that again any estimates of clinical validity obtained from this study would have been less precise than those obtained from a meta-analysis of observational studies.

My plots showing the estimates of OR, PPV and NPV from the information derived from meta-analyses show that the precision of the estimates peaks at around 50 cases. This shows that recruiting past this number does not improve the precision of estimates of clinical validity. For a rare ADR, a prospective intervention study would require a huge number of participants to achieve 50 cases, therefore utilising information from observational studies instead is more appropriate.

PPV was extremely low for both *HLA-B*15:02* and *HLA-A*31:01* genotyping. This was expected due to the rarity of SJS/TEN. PPV is low for rare ADRs, even where the OR is high (67). In contrast, NPV was high, showing that almost all of the

patients who do not carry the genotype will not experience SJS/TEN. Clinically, this is the preferred configuration for these measurements. A low PPV and high NPV ensures a conservative approach to prescribing carbamazepine that will minimise the risk of SJS/TEN for all patients. The fact that there are other effective drugs available for the treatment of epilepsy (and other carbamazepine indications, e.g., bipolar disorder) provides further confidence in this conservative prescribing strategy.

Ultimately, both prospective interventional studies and observational studies are needed to form a solid base of research. Each method has its own strengths and limitations (51, 74, 139). Prospective interventional studies, particularly RCTs, are the 'gold standard' of evidence (50), but can be expensive, time-consuming, and are near-impossible in very rare conditions (54, 56-59). Observational data is generally easier to collect, and is appropriate to use with very rare outcomes, or where an RCT would be unethical (61, 63, 64, 74). However, observational research can suffer from risk of bias and is sensitive to confounders (51). Nonetheless, I have shown that combining several observational studies can produce highly precise estimates of clinical validity, making this an attractive and feasible alternative to conducting a prospective, interventional study where this might be impractical, or impossible.

4.4.1 Limitations

Observational data can be an excellent alternative when it is not possible to perform a prospective interventional study. However, analyses in different disease areas have found that observational data can be associated with higher risk of bias and changes to the direction and size of effects (140, 141). The methodological quality and risk of bias of studies included in the meta-analyses were assessed using the criteria of Jorgensen and Williamson (85). This assessment was qualitative, not quantitative. The results of this assessment were therefore not used to perform sub-group analyses of meta-analyses. However, the results can be used to narratively describe overall study quality. No studies met all the criteria, indicating that better attention to study quality is required in future studies. This finding should be kept in mind when evaluating the results of the meta-analyses.

One paper was excluded in my analysis that other meta-analyses have included in their analyses. This paper is an abstract from a 2009 meeting of the American Epilepsy Society (18). Even though it is only a meeting abstract, other papers did

include Liao et al in their meta-analyses (2, 76). However, this would have involved changing this project's pre-determined inclusion and exclusion criteria.

One paper was excluded as there were discrepancies in numbers between their tables and text (124). I contacted the authors for clarity but did not receive a response. However, this paper performed reasonably well on the quality assessment. This suggests the checklist was not quite suited for my purposes. A stricter checklist, that incorporates measures such as completeness of reporting and data availability would be a useful addition to systematic reviewing in this area. This is an interesting future avenue of study.

While I tried to account for overlap of patients between studies by contacting authors for details, this was not always possible. This was mitigated by only including the study with the larger sample size when overlap was detected, but there is a chance that overlap of patients may have occurred. This could potentially impact estimates.

When analysing by type of ADR, SJS and TEN events were not combined into one 'SJS/TEN' event, as other meta-analyses have done (112). Combining them may overestimate the number of cases, and indeed SJS, SJS/TEN, and TEN are related but separate clinical entities (4, 7). However, these exclusions are important to note as a caveat of the final risk estimates.

My simulation made the assumption that patients who were *HLA-A*31:01* carriers would have an SJS/TEN rate of 0, since they would test positive if genotyped in a prospective trial and so would not receive carbamazepine. This may not hold in a clinical scenario. A patient would likely be prescribed an alternative drug for their condition. Certainly in epilepsy, many common anti-epileptics have their own associated risks of SJS/TEN (142, 143), although these are generally lower than the risk associated with carbamazepine (144). Phenytoin and lamotrigine are the only drugs with higher risks per 100,000 exposed patients (142). A patient with a positive genetic test for a risk allele should be prescribed a different drug (145), and presumably not one with cross-reactivity with carbamazepine. However, this assumption may still lead to an underestimate of the number of SJS/TEN cases expected in *HLA-A*31:01* carriers.

Finally, a further extension of this work would be a comparison of estimates of clinical validity from the *HLA-A*31:01* meta-analysis to those obtained from a published prospective interventional study (Mushiroda, *et al.* (138)). This is made difficult since these authors did not locate any SJS/TEN cases, although their total

number of ADRs (47 in total, causality relating to carbamazepine ranging from ‘definitely’ ‘unlikely’) could be used as an indication instead. This is an interesting avenue for future work.

4.5 Conclusion

I have shown that observational data can be used in place of prospective data in the case of very rare ADRs such as SJS/TEN to provide precise estimates of clinical validity. The two meta-analyses also provide a much-needed updated view of the field.

Testing for *HLA-A*31:01* has been proven cost-effective (24) but is not currently recommended in the UK (146, 147). My analysis provides further, quantitative evidence of the benefit of *HLA-A*31:01* genotyping prior to the prescription of carbamazepine.

4.6 References

1. Roujeau JC, Stern RS. Severe adverse cutaneous reactions to drugs. The New England journal of medicine. 1994;331(19):1272-85.
2. Yip VL, Marson AG, Jorgensen AL, Pirmohamed M, Alfirevic A. HLA genotype and carbamazepine-induced cutaneous adverse drug reactions: a systematic review. Clinical pharmacology and therapeutics. 2012;92(6):757-65.
3. Chen P, Lin JJ, Lu CS, Ong CT, Hsieh PF, Yang CC, et al. Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan. The New England journal of medicine. 2011;364(12):1126-33.
4. Harr T, French LE. Toxic epidermal necrolysis and Stevens-Johnson syndrome. Orphanet J Rare Dis. 2010;5(1):39.
5. Langley A, Worley B, Pardo Pardo J, Beecker J, Ramsay T, Saavedra A, et al. Systemic interventions for treatment of Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN), and SJS/TEN overlap syndrome. Cochrane Database of Systematic Reviews. 2018(9).
6. Patel TK, Barvaliya MJ, Sharma D, Tripathi C. A systematic review of the drug-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in Indian population. Indian journal of dermatology, venereology and leprology. 2013;79(3):389-98.

7. Bastuji-Garin S, Rzany B, Stern RS, Shear NH, Naldi L, Roujeau JC. Clinical classification of cases of toxic epidermal necrolysis, Stevens-Johnson syndrome, and erythema multiforme. Arch Dermatol. 1993;129(1):92-6.
8. Roujeau JC. Clinical heterogeneity of drug hypersensitivity. Toxicology. 2005;209(2):123-9.
9. Phillips EJ, Sukasem C, Whirl-Carrillo M, Muller DJ, Dunnenberger HM, Chantratita W, et al. Clinical Pharmacogenetics Implementation Consortium Guideline for HLA Genotype and Use of Carbamazepine and Oxcarbazepine: 2017 Update. Clinical pharmacology and therapeutics. 2018;103(4):574-81.
10. Sekula P, Dunant A, Mockenhaupt M, Naldi L, Bouwes Bavinck JN, Halevy S, et al. Comprehensive survival analysis of a cohort of patients with Stevens-Johnson syndrome and toxic epidermal necrolysis. J Invest Dermatol. 2013;133(5):1197-204.
11. Cheung YK, Cheng SH, Chan EJ, Lo SV, Ng MH, Kwan P. HLA-B alleles associated with severe cutaneous reactions to antiepileptic drugs in Han Chinese. Epilepsia. 2013;54(7):1307-14.
12. Svensson CK, Cowen EW, Gaspari AA. Cutaneous drug reactions. Pharmacol Rev. 2001;53(3):357-79.
13. Phillips KA, Veenstra DL, Oren E, Lee JK, Sadee W. Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. JAMA. 2001;286(18):2270-9.
14. Cacoub P, Musette P, Descamps V, Meyer O, Speirs C, Finzi L, et al. The DRESS syndrome: a literature review. Am J Med. 2011;124(7):588-97.
15. Fricke-Galindo I, Martinez-Juarez IE, Monroy-Jaramillo N, Jung-Cook H, Falfan-Valencia R, Ortega-Vazquez A, et al. HLA-A*02:01:01/-B*35:01:01/-C*04:01:01 haplotype associated with lamotrigine-induced maculopapular exanthema in Mexican Mestizo patients. Pharmacogenomics. 2014;15(15):1881-91.
16. Genin E, Schumacher M, Roujeau JC, Naldi L, Liss Y, Kazma R, et al. Genome-wide association study of Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis in Europe. Orphanet J Rare Dis. 2011;6:52.
17. Park HJ, Kim YJ, Kim DH, Kim J, Park KH, Park JW, et al. HLA Allele Frequencies in 5802 Koreans: Varied Allele Types Associated with SJS/TEN According to Culprit Drugs. Yonsei Med J. 2016;57(1):118-26.
18. Liao WP, Shi YW, Cheng SH, Ng MH, Kwan P. Association between Hla-B*1502 Allele and Cutaneous Reactions Induced by Carbamazepine or Lamotrigine in Han Chinese. Epilepsia. 2009;50:252-3.

19. Roujeau JC, Kelly JP, Naldi L, Rzany B, Stern RS, Anderson T, et al. Medication use and the risk of Stevens-Johnson syndrome or toxic epidermal necrolysis. The New England journal of medicine. 1995;333(24):1600-7.
20. Delco F, Tchambaz L, Schlienger R, Drewe J, Krahenbuhl S. Dose adjustment in patients with liver disease. Drug Saf. 2005;28(6):529-45.
21. Tennis P, Stern RS. Risk of serious cutaneous disorders after initiation of use of phenytoin, carbamazepine, or sodium valproate: a record linkage study. Neurology. 1997;49(2):542-6.
22. Collins DC, Sundar R, Lim JSJ, Yap TA. Towards Precision Medicine in the Clinic: From Biomarker Discovery to Novel Therapeutics. Trends Pharmacol Sci. 2017;38(1):25-40.
23. Food and Drug Administration (US). Guidance for industry: E15 definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories 2008 [Available from: https://www.fda.gov/media/71389/download.
24. Plumpton CO, Yip VL, Alfirevic A, Marson AG, Pirmohamed M, Hughes DA. Cost-effectiveness of screening for HLA-A*31:01 prior to initiation of carbamazepine in epilepsy. Epilepsia. 2015;56(4):556-63.
25. British National Formulary. ABACAVIR [Available from: https://bnf.nice.org.uk/drug/abacavir.html#preTreatmentScreeningInformation.
26. van Schie RM, Wadelius MI, Kamali F, Daly AK, Manolopoulos VG, de Boer A, et al. Genotype-guided dosing of coumarin derivatives: the European pharmacogenetics of anticoagulant therapy (EU-PACT) trial design. Pharmacogenomics. 2009;10(10):1687-95.
27. Chung WH, Hung SI, Hong HS, Hsih MS, Yang LC, Ho HC, et al. Medical genetics: a marker for Stevens-Johnson syndrome. Nature. 2004;428(6982):486.
28. Monos DS, Winchester RJ. The Major Histocompatibility Complex. In: Rich RR, Fleisher TA, Shearer WT, Schroeder HW, Frew AJ, Weyand CM, editors. Clinical Immunology. London: Content Repository Only!; 2019. p. 79-92.e1.
29. Neefjes J, Jongsma ML, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol. 2011;11(12):823-36.
30. Wieczorek M, Abualrous ET, Sticht J, Alvaro-Benito M, Stolzenberg S, Noe F, et al. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. Front Immunol. 2017;8(292):292.

31. Michels AW, Ostrov DA. New approaches for predicting T cell-mediated drug reactions: A role for inducible and potentially preventable autoimmunity. J Allergy Clin Immunol. 2015;136(2):252-7.
32. Phillips EJ, Mallal SA. HLA and drug-induced toxicity. Curr Opin Mol Ther. 2009;11(3):231-42.
33. Pearce RE, Lu W, Wang Y, Uetrecht JP, Correia MA, Leeder JS. Pathways of carbamazepine bioactivation in vitro. III. The role of human cytochrome P450 enzymes in the formation of 2,3-dihydroxycarbamazepine. Drug Metab Dispos. 2008;36(8):1637-49.
34. Thorn CF, Leckband SG, Kelsoe J, Leeder JS, Muller DJ, Klein TE, et al. PharmGKB summary: carbamazepine pathway. Pharmacogenetics and genomics. 2011;21(12):906-10.
35. Shear NH, Spielberg SP. Anticonvulsant hypersensitivity syndrome. In vitro assessment of risk. J Clin Invest. 1988;82(6):1826-32.
36. Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, et al. Nomenclature for factors of the HLA system, 2010. Tissue Antigens. 2010;75(4):291-455.
37. Marsh SGE, Parham P, Barber LD. The HLA FactsBook: Elsevier Science; 1999.
38. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 2015;43(Database issue):D423-31.
39. Zhou P, Zhang S, Wang Y, Yang C, Huang J. Structural modeling of HLA-B*1502/peptide/carbamazepine/T-cell receptor complex architecture: implication for the molecular mechanism of carbamazepine-induced Stevens-Johnson syndrome/toxic epidermal necrolysis. J Biomol Struct Dyn. 2016;34(8):1806-17.
40. Hung SI, Chung WH, Jee SH, Chen WC, Chang YT, Lee WR, et al. Genetic susceptibility to carbamazepine-induced cutaneous adverse drug reactions. Pharmacogenetics and genomics. 2006;16(4):297-306.
41. Guerrini R, Perucca E. Genetic testing to prevent adverse reactions to antiepileptic drugs: Primum non nocere. Neurology. 2018;90(4):155-6.
42. Lonjou C, Thomas L, Borot N, Ledger N, de Toma C, LeLouet H, et al. A marker for Stevens-Johnson syndrome ...: ethnicity matters. Pharmacogenomics J. 2006;6(4):265-8.
43. Alfirevic A, Jorgensen AL, Williamson PR, Chadwick DW, Park BK, Pirmohamed M. HLA-B locus in Caucasian patients with carbamazepine hypersensitivity. Pharmacogenomics. 2006;7(6):813-8.

44. Fernando SL, Broadfoot AJ. Prevention of severe cutaneous adverse drug reactions: the emerging value of pharmacogenetic screening. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2010;182(5):476-80.
45. Yip VL, Pirmohamed M. The HLA-A*31:01 allele: influence on carbamazepine treatment. Pharmgenomics Pers Med. 2017;10:29-38.
46. McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperaviciute D, Carrington M, et al. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. The New England journal of medicine. 2011;364(12):1134-43.
47. Gonzalez-Galarza FF, McCabe A, Santos E, Jones J, Takeshita L, Ortega-Rivera ND, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res. 2020;48(D1):D783-D8.
48. Genin E, Chen DP, Hung SI, Sekula P, Schumacher M, Chang PY, et al. HLA-A*31:01 and different types of carbamazepine-induced severe cutaneous adverse reactions: an international study and meta-analysis. Pharmacogenomics J. 2014;14(3):281-8.
49. Hyde P. Fool's Gold: Examining the Use of Gold Standards in the Production of Research Evidence. British Journal of Occupational Therapy. 2016;67(2):89-94.
50. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the Gold Standard--Lessons from the History of RCTs. The New England journal of medicine. 2016;374(22):2175-81.
51. Barton S. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. BMJ. 2000;321(7256):255-6.
52. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, et al. HLA-B*5701 screening for hypersensitivity to abacavir. The New England journal of medicine. 2008;358(6):568-79.
53. Pirmohamed M, Burnside G, Eriksson N, Jorgensen AL, Toh CH, Nicholson T, et al. A randomized trial of genotype-guided dosing of warfarin. The New England journal of medicine. 2013;369(24):2294-303.
54. Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. Statistics in medicine. 2014;33(24):4186-201.
55. Antoniou M, Kolamunnage-Dona R, Wason J, Bathia R, Billingham C, Bliss JM, et al. Biomarker-guided trials: Challenges in practice. Contemp Clin Trials Commun. 2019;16:100493.

56. Billingham L, Malottki K, Steven N. Small sample sizes in clinical trials: a statistician's perspective. Clinical Investigation. 2012;2(7):655-7.
57. Miller F, Zohar S, Stallard N, Madan J, Posch M, Hee SW, et al. Approaches to sample size calculation for clinical trials in rare diseases. Pharm Stat. 2018;17(3):214-30.
58. Shirts BH, Jacobson A, Jarvik GP, Browning BL. Large numbers of individuals are required to classify and define risk for rare variants in known cancer risk genes. Genet Med. 2014;16(7):529-34.
59. Zakrzewska JM, Palmer J, Bendtsen L, Di Stefano G, Ettlin DA, Maarbjerg S, et al. Challenges recruiting to a proof-of-concept pharmaceutical trial for a rare disease: the trigeminal neuralgia experience. Trials. 2018;19(1):704.
60. London AJ. Two dogmas of research ethics and the integrative approach to human-subjects research. J Med Philos. 2007;32(2):99-116.
61. Stewart DJ, Whitney SN, Kurzrock R. Equipoise lost: ethics, costs, and the regulation of cancer clinical research. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(17):2925-35.
62. Webster F, Weijer C, Todd L, Grimshaw JM, Marshall AP, Cook D, et al. The ethics of future trials: qualitative analysis of physicians' decision making. Trials. 2016;17:12.
63. Augustine EF, Adams HR, Mink JW. Clinical trials in rare disease: challenges and opportunities. J Child Neurol. 2013;28(9):1142-50.
64. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J, et al. Clinical research for rare disease: opportunities, challenges, and solutions. Mol Genet Metab. 2009;96(1):20-6.
65. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. PLoS Med. 2011;8(5):e1001026.
66. Huddart R, Sangkuhl K, Whirl-Carrillo M, Klein TE. Are Randomized Controlled Trials Necessary to Establish the Value of Implementing Pharmacogenomics in the Clinic? Clinical pharmacology and therapeutics. 2019;106(2):284-6.
67. Tonk ECM, Gurwitz D, Maitland-van der Zee AH, Janssens A. Assessment of pharmacogenetic tests: presenting measures of clinical validity and potential population impact in association studies. Pharmacogenomics J. 2017;17(4):386-92.
68. Food and Drug Administration (US). Biomarker Qualification—collaborative effort to Qualify a Drug Development tool 2017 [Available from:

https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/BiomarkerQualificationProgram/UCM552881.pdf.
69. Blagden SP, Billingham L, Brown LC, Buckland SW, Cooper AM, Ellis S, et al. Effective delivery of Complex Innovative Design (CID) cancer trials-A consensus statement. British journal of cancer. 2020;122(4):473-82.
70. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. The New England journal of medicine. 2000;342(25):1887-92.
71. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. The New England journal of medicine. 2000;342(25):1878-86.
72. Gillis NK, Innocenti F. Evidence required to demonstrate clinical utility of pharmacogenetic testing: the debate continues. Clinical pharmacology and therapeutics. 2014;96(6):655-7.
73. Jansen ME, Rigter T, Rodenburg W, Fleur TMC, Houwink EJF, Weda M, et al. Review of the Reported Measures of Clinical Validity and Clinical Utility as Arguments for the Implementation of Pharmacogenetic Testing: A Case Study of Statin-Induced Muscle Toxicity. Frontiers in pharmacology. 2017;8(555):555.
74. Boyko EJ. Observational research--opportunities and limitations. J Diabetes Complications. 2013;27(6):642-8.
75. Concato J, Lawler EV, Lew RA, Gaziano JM, Aslan M, Huang GD. Observational methods in comparative effectiveness research. Am J Med. 2010;123(12 Suppl 1):e16-23.
76. Tangamornsuksan W, Chaiyakunapruk N, Somkrua R, Lohitnavy M, Tassaneeyakul W. Relationship between the HLA-B*1502 allele and carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis: a systematic review and meta-analysis. JAMA Dermatol. 2013;149(9):1025-32.
77. Grover S, Kukreti R. HLA alleles and hypersensitivity to carbamazepine: an updated systematic review with meta-analysis. Pharmacogenetics and genomics. 2014;24(2):94-112.
78. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1.
79. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

80. Johnson D, Jorgensen AL. The influence of HLA-B*15:02 and HLA-A*31:01 on the risk of developing adverse skin reactions to carbamazepine: protocol for two systematic reviews. CRD42019161000. 2019 [Available from: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=161000.
81. Schwarzer G. Package 'meta' 2020 [updated 3rd July 2020. Available from: https://cran.r-project.org/web//packages/meta/meta.pdf.
82. Gordon M, Lumley T. Package 'forestplot' 2020 [updated 16th July 2020. Available from: https://cran.r-project.org/web/packages/forestplot/forestplot.pdf.
83. Deeks JJ, Higgins JP, Altman DG, Group CSM. Analysing data and undertaking meta-analyses. Cochrane handbook for systematic reviews of interventions. 2019:241-84.
84. Rosenblad A. Introduction to Meta-Analysis by Michael Borenstein, Larry V. Hedges, Julian PT Higgins, Hannah R. Rothstein. Wiley Online Library; 2009.
85. Jorgensen AL, Williamson PR. Methodological quality of pharmacogenetic studies: issues of concern. Statistics in medicine. 2008;27(30):6547-69.
86. Jorgensen AL, FitzGerald RJ, Oyee J, Pirmohamed M, Williamson PR. Influence of CYP2C9 and VKORC1 on patient response to warfarin: a systematic review and meta-analysis. PLoS One. 2012;7(8):e44064.
87. Richardson M, Kirkham J, Dwan K, Sloan D, Davies G, Jorgensen A. Influence of genetic variants on toxicity to anti-tubercular agents: a systematic review and meta-analysis (protocol). Syst Rev. 2017;6(1):142.
88. El-Maraghi RH, Kielar AZ. CT colonography versus optical colonoscopy for screening asymptomatic patients for colorectal cancer a patient, intervention, comparison, outcome (PICO) analysis. Acad Radiol. 2009;16(5):564-71.
89. Staunton M. Evidence-based Radiology: Steps 1 and 2—Asking Answerable Questions and Searching for Evidence. Radiology. 2007;242(1):23-31.
90. World Health Organization. International classification of diseases—Ninth revision (ICD-9). Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire. 1988;63(45):343-4.
91. Hung SI, Chung WH, Chen YT. HLA-B genotyping to detect carbamazepine-induced Stevens-Johnson syndrome: implications for personalizing medicine. Per Med. 2005;2(3):225-37.
92. Downey AB. Think Bayes: " O'Reilly Media, Inc."; 2021.
93. Kulesa A, Krzywinski M, Blainey P, Altman N. Sampling distributions and the bootstrap. Nat Methods. 2015;12(6):477-8.

94. Wolfe DA, Schneider G. Statistical Inference for the Center of a Population. In: Wolfe DA, Schneider G, editors. Intuitive Introductory Statistics. Springer Texts in Statistics. Cham: Springer International Publishing; 2017. p. 537-631.
95. Aggarwal R, Sharma M, Modi M, Garg VK, Salaria M. HLA-B * 1502 is associated with carbamazepine induced Stevens-Johnson syndrome in North Indian population. Hum Immunol. 2014;75(11):1120-2.
96. Amstutz U, Ross CJ, Castro-Pastrana LI, Rieder MJ, Shear NH, Hayden MR, et al. HLA-A 31:01 and HLA-B 15:02 as genetic markers for carbamazepine hypersensitivity in children. Clinical pharmacology and therapeutics. 2013;94(1):142-9.
97. Buyukozturk S, Kekik C, Gokyigit AZ, Tezer Filik FI, Karakaya G, Saygi S, et al. Cutaneous drug reactions to antiepileptic drugs and relation with HLA alleles in the Turkish population. Eur Ann Allergy Clin Immunol. 2018;50(1):36-41.
98. Chang CC, Too CL, Murad S, Hussein SH. Association of HLA-B*1502 allele with carbamazepine-induced toxic epidermal necrolysis and Stevens-Johnson syndrome in the multi-ethnic Malaysian population. International journal of dermatology. 2011;50(2):221-4.
99. Chong KW, Chan DW, Cheung YB, Ching LK, Hie SL, Thomas T, et al. Association of carbamazepine-induced severe cutaneous drug reactions and HLA-B*1502 allele status, and dose and treatment duration in paediatric neurology patients in Singapore. Arch Dis Child. 2014;99(6):581-4.
100. Devi K. The association of HLA B*15:02 allele and Stevens-Johnson syndrome/toxic epidermal necrolysis induced by aromatic anticonvulsant drugs in a South Indian population. International journal of dermatology. 2018;57(1):70-3.
101. He X-J, Jian L-Y, He X-L, Wu Y, Xu Y-Y, Sun X-J, et al. Association between the HLA-B*15:02 allele and carbamazepine-induced Stevens-Johnson syndrome/toxic epidermal necrolysis in Han individuals of northeastern China. Pharmacological Reports. 2013;65(5):1256-62.
102. Hsiao YH, Hui RC, Wu T, Chang WC, Hsih MS, Yang CH, et al. Genotype-phenotype association between HLA and carbamazepine-induced hypersensitivity reactions: strength and clinical correlations. J Dermatol Sci. 2014;73(2):101-9.
103. Kaniwa N, Saito Y, Aihara M, Matsunaga K, Tohkin M, Kurose K, et al. HLA-B locus in Japanese patients with anti-epileptics and allopurinol-related Stevens-Johnson syndrome and toxic epidermal necrolysis. Pharmacogenomics. 2008;9(11):1617-22.
104. Kashiwagi M, Aihara M, Takahashi Y, Yamazaki E, Yamane Y, Song Y, et al. Human leukocyte antigen genotypes in carbamazepine-induced severe

cutaneous adverse drug response in Japanese patients. J Dermatol. 2008;35(10):683-5.
105. Khor AH, Lim KS, Tan CT, Kwan Z, Tan WC, Wu DB, et al. HLA-A*31: 01 and HLA-B*15:02 association with Stevens-Johnson syndrome and toxic epidermal necrolysis to carbamazepine in a multiethnic Malaysian population. Pharmacogenetics and genomics. 2017;27(7):275-8.
106. Khor AH, Lim KS, Tan CT, Wong SM, Ng CC. HLA-B*15:02 association with carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in an Indian population: a pooled-data analysis and meta-analysis. Epilepsia. 2014;55(11):e120-4.
107. Kim SH, Lee KW, Song WJ, Kim SH, Jee YK, Lee SM, et al. Carbamazepine-induced severe cutaneous adverse reactions and HLA genotypes in Koreans. Epilepsy Res. 2011;97(1-2):190-7.
108. Ko TM, Chung WH, Wei CY, Shih HY, Chen JK, Lin CH, et al. Shared and restricted T-cell receptor use is crucial for carbamazepine-induced Stevens-Johnson syndrome. J Allergy Clin Immunol. 2011;128(6):1266-76 e11.
109. Kulkantrakorn K, Tassaneeyakul W, Tiamkao S, Jantararoungtong T, Prabmechai N, Vannaprasaht S, et al. HLA-B*1502 strongly predicts carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in Thai patients with neuropathic pain. Pain Pract. 2012;12(3):202-8.
110. Kwan PKL, Ng MHL, Lo SV. Association between HLA–B*15:02 allele and antiepileptic drug–induced severe cutaneous reactions in Hong Kong Chinese: A population-based study. Hong Kong Medical Journal. 2014;20(6):16-8.
111. Li LJ, Hu FY, Wu XT, An DM, Yan B, Zhou D. Predictive markers for carbamazepine and lamotrigine-induced maculopapular exanthema in Han Chinese. Epilepsy Res. 2013;106(1-2):296-300.
112. Locharernkul C, Loplumlert J, Limotai C, Korkij W, Desudchit T, Tongkobpetch S, et al. Carbamazepine and phenytoin induced Stevens-Johnson syndrome is associated with HLA-B*1502 allele in Thai population. Epilepsia. 2008;49(12):2087-91.
113. Lonjou C, Borot N, Sekula P, Ledger N, Thomas L, Halevy S, et al. A European study of HLA-B in Stevens-Johnson syndrome and toxic epidermal necrolysis related to five high-risk drugs. Pharmacogenetics and genomics. 2008;18(2):99-107.
114. Man CB, Kwan P, Baum L, Yu E, Lau KM, Cheng AS, et al. Association between HLA-B*1502 allele and antiepileptic drug-induced cutaneous reactions in Han Chinese. Epilepsia. 2007;48(5):1015-8.

115. McCormack M, Gui H, Ingason A, Speed D, Wright GEB, Zhang EJ, et al. Genetic variation in CFH predicts phenytoin-induced maculopapular exanthema in European-descent patients. Neurology. 2018;90(4):e332-e41.
116. Mehta TY, Prajapati LM, Mittal B, Joshi CG, Sheth JJ, Patel DB, et al. Association of HLA-B*1502 allele and carbamazepine-induced Stevens-Johnson syndrome among Indians. Indian journal of dermatology, venereology and leprology. 2009;75(6):579-82.
117. Nguyen DV, Chu HC, Nguyen DV, Phan MH, Craig T, Baumgart K, et al. HLA-B*1502 and carbamazepine-induced severe cutaneous adverse drug reactions in Vietnamese. Asia Pac Allergy. 2015;5(2):68-77.
118. Niihara H, Kakamu T, Fujita Y, Kaneko S, Morita E. HLA-A31 strongly associates with carbamazepine-induced adverse drug reactions but not with carbamazepine-induced lymphocyte proliferation in a Japanese population. J Dermatol. 2012;39(7):594-601.
119. Ozeki T, Mushiroda T, Yowang A, Takahashi A, Kubo M, Shirakata Y, et al. Genome-wide association study identifies HLA-A*3101 allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. Hum Mol Genet. 2011;20(5):1034-41.
120. Ramirez E, Bellon T, Tong HY, Borobia AM, de Abajo FJ, Lerma V, et al. Significant HLA class I type associations with aromatic antiepileptic drug (AED)-induced SJS/TEN are different from those found for the same AED-induced DRESS in the Spanish population. Pharmacol Res. 2017;115:168-78.
121. Shi YW, Min FL, Qin B, Zou X, Liu XR, Gao MM, et al. Association between HLA and Stevens-Johnson syndrome induced by carbamazepine in Southern Han Chinese: genetic markers besides B*1502? Basic Clin Pharmacol Toxicol. 2012;111(1):58-64.
122. Shi YW, Min FL, Zhou D, Qin B, Wang J, Hu FY, et al. HLA-A*24:02 as a common risk factor for antiepileptic drug-induced cutaneous adverse reactions. Neurology. 2017;88(23):2183-91.
123. Sukasem C, Chaichan C, Nakkrut T, Satapornpong P, Jaruthamsophon K, Jantararoungtong T, et al. Association between HLA-B Alleles and Carbamazepine-Induced Maculopapular Exanthema and Severe Cutaneous Reactions in Thai Patients. J Immunol Res. 2018;2018:2780272.
124. Sun D, Yu CH, Liu ZS, He XL, Hu JS, Wu GF, et al. Association of HLA-B*1502 and *1511 allele with antiepileptic drug-induced Stevens-Johnson syndrome in central China. J Huazhong Univ Sci Technolog Med Sci. 2014;34(1):146-50.

125. Tassaneeyakul W, Tiamkao S, Jantararoungtong T, Chen P, Lin SY, Chen WH, et al. Association between HLA-B*1502 and carbamazepine-induced severe cutaneous adverse drug reactions in a Thai population. Epilepsia. 2010;51(5):926-30.
126. Teh LK, Selvaraj M, Bannur Z, Ismail MI, Rafia H, Law WC, et al. Coupling Genotyping and Computational Modeling in Prediction of Anti-epileptic Drugs that cause Stevens Johnson Syndrome and Toxic Epidermal Necrolysis for Carrier of HLA-B*15:02. J Pharm Pharm Sci. 2016;19(1):147-60.
127. Then SM, Rani ZZ, Raymond AA, Ratnaningrum S, Jamal R. Frequency of the HLA-B*1502 allele contributing to carbamazepine-induced hypersensitivity reactions in a cohort of Malaysian epilepsy patients. Asian Pac J Allergy Immunol. 2011;29(3):290-3.
128. Toh DS, Tan LL, Aw DC, Pang SM, Lim SH, Thirumoorthy T, et al. Building pharmacogenetics into a pharmacovigilance program in Singapore: using serious skin rash as a pilot study. Pharmacogenomics J. 2014;14(4):316-21.
129. Wang Q, Zhou JQ, Zhou LM, Chen ZY, Fang ZY, Chen SD, et al. Association between HLA-B*1502 allele and carbamazepine-induced severe cutaneous adverse reactions in Han people of southern China mainland. Seizure. 2011;20(6):446-8.
130. Wu XT, Hu FY, An DM, Yan B, Jiang X, Kwan P, et al. Association between carbamazepine-induced cutaneous adverse drug reactions and the HLA-B*1502 allele among patients in central China. Epilepsy & behavior : E&B. 2010;19(3):405-8.
131. Yang F, Yang Y, Zhu Q, Chen SA, Fu X, Yan S, et al. Research on Susceptible Genes and Immunological Pathogenesis of Cutaneous Adverse Drug Reactions in Chinese Hans. J Investig Dermatol Symp Proc. 2015;17(1):29-31.
132. Yuliwulandari R, Kristin E, Prayuni K, Sachrowardi Q, Suyatna FD, Menaldi SL, et al. Association of the HLA-B alleles with carbamazepine-induced Stevens-Johnson syndrome/toxic epidermal necrolysis in the Javanese and Sundanese population of Indonesia: the important role of the HLA-B75 serotype. Pharmacogenomics. 2017;18(18):1643-8.
133. Zhang Y, Wang J, Zhao LM, Peng W, Shen GQ, Xue L, et al. Strong association between HLA-B*1502 and carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in mainland Han Chinese patients. Eur J Clin Pharmacol. 2011;67(9):885-7.

134. Ikeda H, Takahashi Y, Yamazaki E, Fujiwara T, Kaniwa N, Saito Y, et al. HLA class I markers in Japanese patients with carbamazepine-induced cutaneous adverse reactions. Epilepsia. 2010;51(2):297-300.
135. Ksouda K, Affes H, Mahfoudh N, Chtourou L, Kammoun A, Charfi A, et al. HLA-A*31:01 and carbamazepine-induced DRESS syndrom in a sample of North African population. Seizure. 2017;53:42-6.
136. Maekawa K, Nakamura R, Kaniwa N, Mizusawa S, Kitamoto A, Kitamoto T, et al. Development of a simple genotyping method for the HLA-A*31:01-tagging SNP in Japanese. Pharmacogenomics. 2015;16(15):1689-99.
137. PharmGKB. HLA-B Drug Label Annotations 2021 [Available from: https://www.pharmgkb.org/gene/PA35056/labelAnnotation.
138. Mushiroda T, Takahashi Y, Onuma T, Yamamoto Y, Kamei T, Hoshida T, et al. Association of HLA-A*31:01 Screening With the Incidence of Carbamazepine-Induced Cutaneous Adverse Reactions in a Japanese Population. JAMA Neurol. 2018;75(7):842-9.
139. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. British journal of cancer. 2014;110(3):551-5.
140. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ. 1998;317(7167):1185-90.
141. Rush CJ, Campbell RT, Jhund PS, Petrie MC, McMurray JJV. Association is not causation: treatment effects cannot be estimated from observational data in heart failure. Eur Heart J. 2018;39(37):3417-38.
142. Frey N, Bodmer M, Bircher A, Ruegg S, Jick SS, Meier CR, et al. The risk of Stevens-Johnson syndrome and toxic epidermal necrolysis in new users of antiepileptic drugs. Epilepsia. 2017;58(12):2178-85.
143. Borrelli EP, Lee EY, Descoteaux AM, Kogut SJ, Caffrey AR. Stevens-Johnson syndrome and toxic epidermal necrolysis with antiepileptic drugs: An analysis of the US Food and Drug Administration Adverse Event Reporting System. Epilepsia. 2018;59(12):2318-24.
144. Devi K, George S, Criton S, Suja V, Sridevi P. Carbamazepine--the commonest cause of toxic epidermal necrolysis and Stevens-Johnson syndrome: a study of 7 years. Indian Journal of Dermatology, Venereology & Leprology. 2005;71(5).
145. British National Formulary. CARBAMAZEPINE [Available from: https://bnf.nice.org.uk/drug/carbamazepine.html.

146. Medicines and Healthcare products Regulatory Agency. Carbamazepine, oxcarbazepine and eslicarbazepine: potential risk of serious skin reactions 2014 [Available from: https://www.gov.uk/drug-safety-update/carbamazepine-oxcarbazepine-and-eslicarbazepine-potential-risk-of-serious-skin-reactions.
147. PharmGKB. HLA-A*31:01:02 Drug Label Annotations 2021 [Available from: https://www.pharmgkb.org/haplotype/PA165954030/labelAnnotation.

Chapter 5: Discrete choice experiments in pharmacogenetics and adverse drug reactions: a systematic review

5.1 Introduction

Pharmacogenetics uses genetic biomarkers to predict treatment response or adverse events (1). Pharmacogenetics also has applications in improving patient prognosis, improving the cost-effectiveness of medicines, and in drug development (2). Further advancement in pharmacogenetics and personalised medicine depends on the acceptance of the technology by healthcare professionals, patients, and the general public (3). Individual patient preferences are key for successful implementations of new healthcare interventions (4-6). Holding accurate data on patient preferences for pharmacogenetic testing and aligning with them can aid in optimally configuring genetic testing services and associated treatments, and can also be used to increase uptake of such tests (7-10).

Preferences can broadly be split into categories of revealed and stated. Revealed preferences infer participant's preferences indirectly by using observations from real-life situations (11). For example, this might be a study that infers how often people prefer to give blood from a database of blood donations (12). In contrast, stated preference methods ask participants directly about their preferences. In this same blood donation example, this might be a survey asking how often people would prefer to give blood (12). One method of quantifying stated preferences is a discrete choice experiment (DCE). DCEs are an efficient and scientifically rigorous way to quantify patient stated preferences. The output can be used to estimate uptake rates, maximum acceptable risk, and ideal test characteristics. These outputs are often used to inform health policy (4, 7, 10, 13-18). Understanding patients' views is essential for the advancement of pharmacogenetics, and DCEs are an ideal method for collecting and interpreting these views.

DCEs are based on random utility theory (RUT), first proposed in 1927 (19), and further work by Lancaster (1966) (20) and McFadden (1986) (21-23). RUT assumes that participants' utilities can be summarised in research by systematic (explainable) and random (unexplainable) components. This is expressed as:

$$U_{in} = V_{in} + \varepsilon_{in}$$

where U_{in} is the utility that individual n associates with choice i, V_{in} is the systematic component of utility that individual n associates with choice i, and ε_{in} is the random component of utility individual n associates with choice i (21, 23, 24).

DCEs are an attribute-based approach to collect stated preference data (16, 21, 25-29). DCEs assume that an individual's choices are rational and can be used to reveal their preferences (utilities), and that participants seek to maximise their utilities [20, 23, 30]. DCEs also assume that interventions can be described by their attributes, and that valuation of these attributes depends on their levels [13, 29, 30].

In a DCE, participants choose between hypothetical scenarios that differ in terms of specified attributes and levels. Attributes are characteristics of the scenario, while levels are functions of each attribute (29, 31). For example, a DCE asking about an ideal disease screening service might assign cost of the test and time to receive results as attributes. Levels are then assigned for each of these attributes (e.g. £10, £100, £1000 for cost, 1 day, 1 week, 1 month for time). Participants 'trade-off' between attributes, allowing the DCE to measure the relative importance of each attribute (16, 28) in order to estimate the strength of the preferences (32, 33) (Figure 5.1). Money (or cost), risk, and time are common attribute domains (32). DCEs are an in demand method that allow quantification of preferences, willingness-to-pay (WTP), and of predicted uptake rates (14). Hall, *et al.* (2004) recognised the method as particularly useful in the evaluation of genetic screening, disease screening (such as breast cancer screening) and immunisation (23). Louviere & Hensher (1982) published the first paper relating to the theory and use of DCEs (34). The paper provides equations for the development of DCEs as well as several examples of choices modelled in this way.

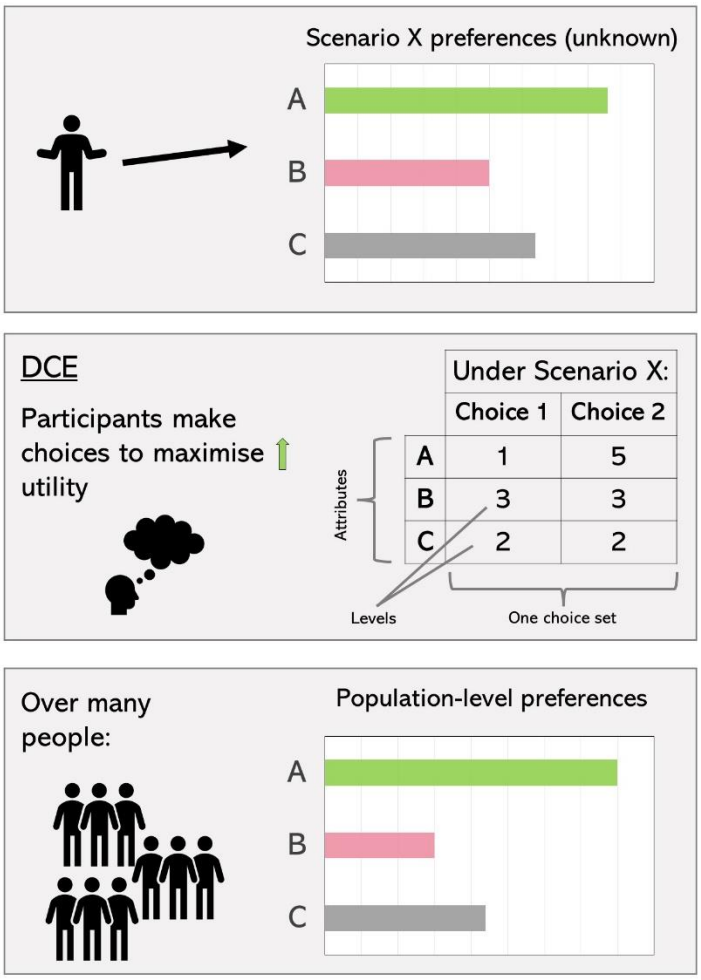


Figure 5.1 - explanation of the process underlying discrete choice experiments. Given scenario X, a person has unknown preferred values for attributes A, B, and C. When they complete the DCE, they make choices to maximise their utility for scenario X. They choose within choice sets with different combinations of attributes and their levels. When the DCE is completed by many people, an estimate of population-level preferences for each attribute can be calculated.

Although often used interchangeably, DCEs should be distinguished from conjoint analysis and conjoint measurement (21, 35). Conjoint analysis is a generic term that describes several ways of eliciting preferences. Conjoint measurement is a mathematical theory concerned with the behaviour of number systems (21). In contrast, DCEs are grounded in utility maximisation and have well-tested links with real behaviour (21).

DCEs are often the best way to ascertain the utility of a service to patients (29, 36-38) and their use in the published literature has been increasing year on year (14, 28, 30, 39, 40) (Figure 5.2).

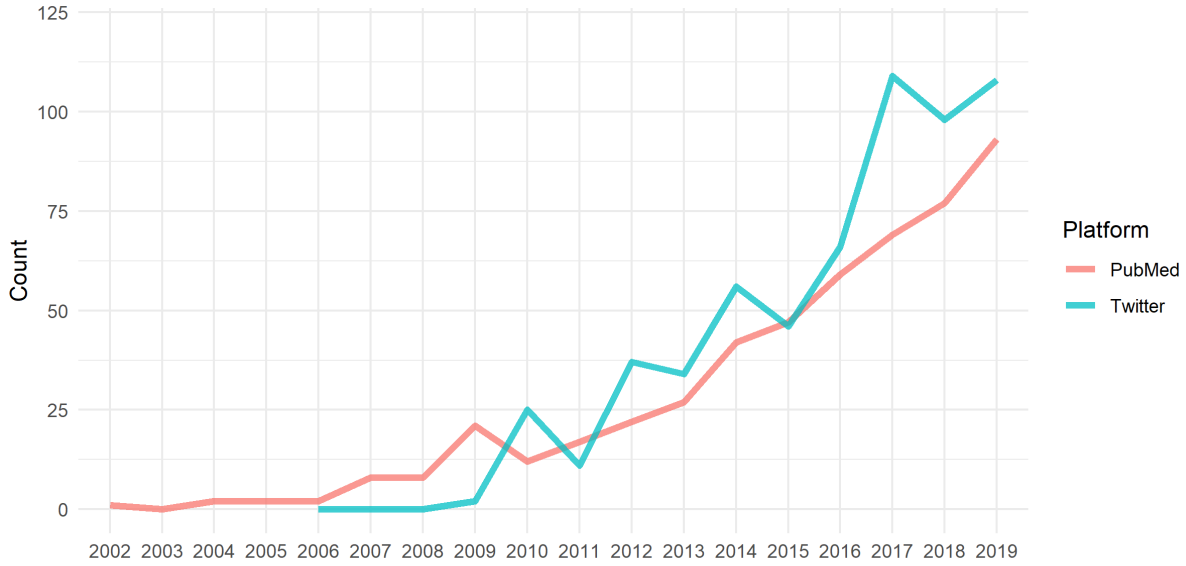


Figure 5.2 - The increasing use of the terms "DCE" or "discrete choice experiment" in PubMed and Twitter to 2019. Note Twitter was created in 2006.

Regulatory agencies have provided guidance on using DCEs for assessing patient preferences (7, 41). The Food and Drug Administration (FDA) has stated several benefits of quantitative research, including selecting patients who will benefit from a treatment, defining 'minimum clinically meaningful benefit' and in improving the generalisability of research (7) (Figure 5.3). There has been interest from licencing authorities in using DCEs to evaluate patient willingness to accept therapeutic risks (29). It has also been suggested that the UK National Institute for Health & Care Excellence (NICE) should use DCEs when evaluating new technologies for National Health Service (NHS) use (13). In pharmacogenetics, these could be designed to ascertain whether a test should be provided, and/or provide insight on the ideal configuration of a testing service (13, 42).

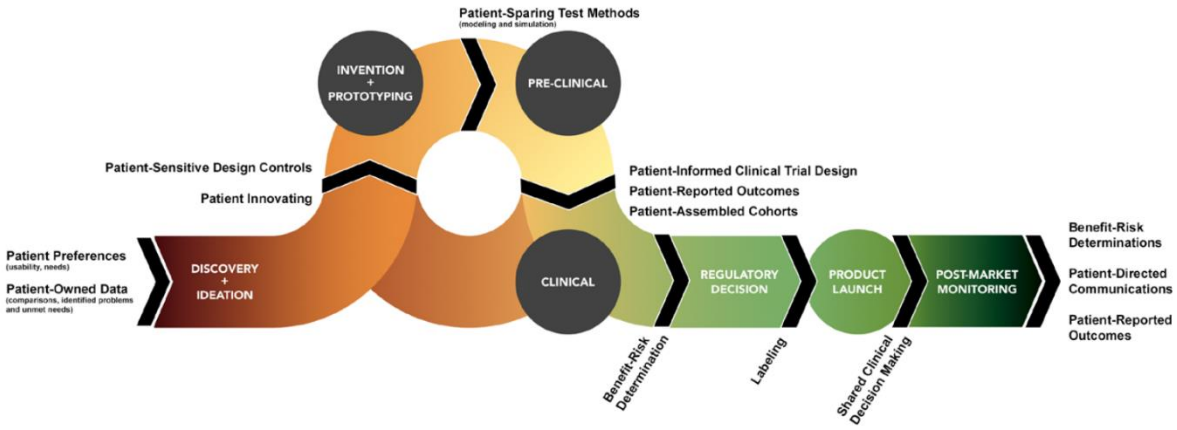


Figure 5.3 - FDA perspective on including patient perspective at every stage of a drug development program. Patient preferences should be incorporated during all stages of drug or product development, from discovery to post-market monitoring. From Patient Preference Information Guidance, FDA 2018 (7).

DCEs can contribute to health policy and service delivery by allowing quantification of preferences and trade-offs, and predicting uptake rates (14, 16, 39). Quantifying preferences can guide the implementation of an intervention, and estimates of uptake rates allow calculation of the potential overall costs of an intervention. They can be used to measure outcomes for inclusion in economic evaluations (16) and patient preferences for funding of health programmes (23, 43).

They are also important for learning about the potential acceptance of interventions by patients and the general public. As the use of pharmacogenetics increases (44), it is essential that these preferences are measured and incorporated into the implementation of new interventions. This may include involvement in the regulatory assessment of such interventions (45). This will increase the likelihood of new interventions and tests being accepted.

While qualitative methods are often used to assess patient preferences, quantitative methods such as DCEs provide a different perspective, particularly suited to complex decisions and scenarios (46). The DCE method is therefore well-placed for the evaluation of multifaceted pharmacogenetic interventions. It is the most widely used method for the evaluation of stated preferences in healthcare (31).

Using pharmacogenetics to reduce the risk of adverse drug reactions (ADRs) is a well-documented and growing field (47, 48), but public awareness of its potentials is low (49, 50) (see also Chapter 6). There has therefore been little work to measure

the preferences of the public for pharmacogenetic testing. I therefore planned a DCE to measure the preferences of the general public for genetic tests to prevent different ADRs.

The design and implementation of a DCE can be split into stages, from defining the problem to analysing policy impacts (Figure 5.4)(13, 26). This chapter focusses on the theory of DCEs, and investigates existing DCEs in pharmacogenetics and ADRs through a systematic review. Chapter 6 contains details of the extensive qualitative work undertaken to inform the selection of my own DCE attributes and levels, and details of the design used for my own DCE. Finally, Chapter 7 presents the results of my DCE investigating patient preferences for genetic testing and discusses how preferences differ between high and low evidence scenarios.

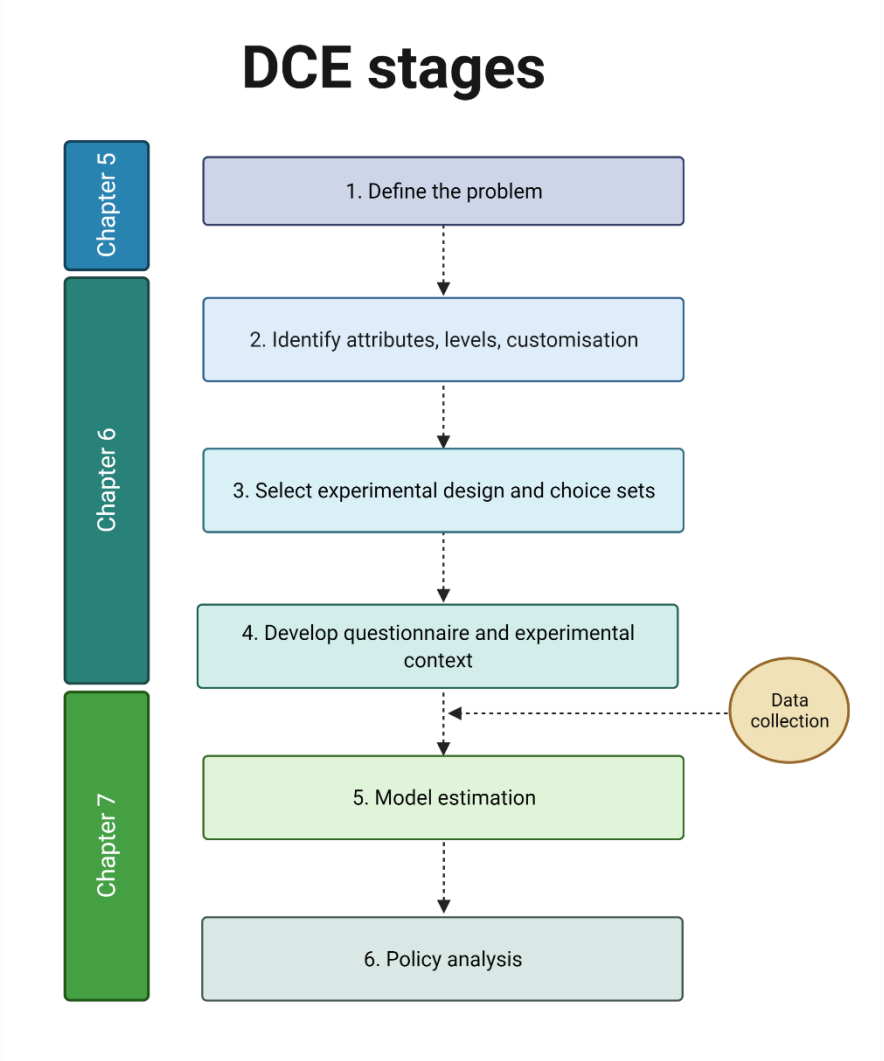


Figure 5.4 – DCE stages as defined by Street, et al. (2008) (26), with reference to the chapters of this thesis that correspond to each stage. Created using BioRender.com.

Prior to the systematic review, I will first discuss some essential DCE design terms, and examine how previous systematic reviews of DCEs have conducted their searches and analyses.

5.1.1 DCE design terms

There are many terms specific to DCEs that first require definition (Table 5.1). Additionally, Louviere, *et al.* (2010) provide a thorough review of the distinctions (21).

Term	Definition	Reference
Conjoint analysis	A generic term to describe several ways of eliciting preferences. Is based on conjoint measurement The implicit values for an attribute of an intervention are derived from some overall score for a profile consisting (conjointly) or two or more attributes	Louviere, *et al.* (2010) (21) Bridges, *et al.* (2011) [29]
Conjoint measurement	A mathematical theory concerned with the behaviour of number systems	Louviere, *et al.* (2010) (21)
D-efficiency	A measure of design efficiency that minimises design error (D-error)	Vanniyasingam, *et al.* (2018) (51) Walker, *et al.* (2018) (52)
Dominance tests	A method of testing for rationality in a DCE by providing choice sets where one alternative is clearly superior	Ryan & Gerard, (2003)(40)
External validity	Comparison of hypothetical and actual behaviour in a DCE	Ryan & Gerard, (2003)(40)
Orthogonality	When the occurrence of any two levels of different attributes is uncorrelated Each pair of levels appears equally often across all pairs of attributes within the design	Marshall, *et al.* (2007)(53) Reed Johnson, *et al.* (2013) (31)
Random utility theory	A theory proposing a latent construct of 'utility' in a person's head, consisting of systematic (explainable) and random (unexplainable) components. Underlie DCE theory	Thurstone (1927) (19) Louviere, *et al.* (2010) (21)
Stated preference	Use of survey methods where individuals are asked hypothetical questions about how much they would be willing to pay or willing to accept in compensation A method that is used to elicit an individual's	Walley, *et al.* (2004) (43) Louviere, *et al.* (2010) (21) Louviere, *et al.* (2000) (54) Abdullah, *et al.* (2011) (55)

	preferences for alternatives (goods, services, courses of action) in a survey context Contrasted with 'revealed preference' which focuses on existing markets and systems	
Utility	A general index of individual satisfaction The value placed on a good or service by any individual as a measure of its usefulness	Marshall, *et al.* (2007)(53) Walley, *et al.* (2004)(43)
Validity	The extent to which quantitative measures of relative importance, or trade-offs, reflect the true preferences of patients	Janssen, *et al.* (2017) (41)
Willingness to pay	Method for deriving preferences for treatment options based on determining what society is willing to pay in monetary terms by asking hypothetical questions	Walley, *et al.* (2004) (43)

Table 5.1 - Definitions of terms used in discrete choice experiments

Design is important to consider when planning a DCE. Arguably, the simplest type of design to visualise is the **full factorial design,** where all combinations of attributes and levels are presented to each participant (16, 52). However, these are rarely used as they can quickly become impractical and costly (25). As the number of possible combinations increases so does the number of questions to each participant. This increase in participant burden increases the risk of participants quitting the DCE (25, 54). One solution is to split questionnaires, but this increases costs. Instead using a subset of all possible combinations is known as a **fractional factorial design** (52). These subsets can be chosen randomly, but statistical methods are more often used to make the selection. This is the most common method for constructing DCEs (4). It is important to measure how efficiently the chosen fraction represents all possible combinations (51, 56). An efficient design is one that is orthogonal (the levels of each attribute vary independently of each other), level balanced, with minimal overlap, and with utility balance (the utilities of alternatives within choice sets are the same) (31, 52, 57, 58). **D-efficiency** is a measure of design efficiency that minimises design error (known as the D-error) (51, 52). Other terms used include **main effects** (a design that is only able to estimate the effect of each attribute independently) and **main and interaction effects** (a design able to estimate main effects and the interactions between attributes) (25).

The analysis of a DCE may take several forms, including simple linear regression, conditional logit, and hierarchical Bayes models (32). The analysis of DCEs is considered in more detail in Chapter 7.

There are also several methods for measuring the validity of a DCE (the extent to which it reflects the true preferences of patients) (41). A detailed overview of validity concepts and measurements is provided by Janssen, *et al.* (2017), and some of the important concepts are summarised here.

Theoretic (or face) validity is the extent to which DCE results are consistent with expectations (41). This is a common test for validity, often done as part of DCE analysis plans. It is examined by looking at expected directions of effect, comparing results to similar DCEs, and ensuring a robust DCE development process with sufficient qualitative work. Within-DCE randomisation (e.g. comparing labelled and unlabelled choices) may also be part of this assessment (8).

External validity is the extent to which preference results can be used to predict real life choices, outside of the experiment (40, 41). This is not always possible since the scenarios examined by DCEs are often hypothetical and/or simplifications of real choices. It has been used successfully in retrospective ('what *would* you have chosen?') DCEs (59), and in DCEs of policy makers (18).

5.1.2 Previous systematic reviews of DCEs

The detailed development of my DCE is covered in Chapter 6. Before beginning my own, I reviewed the literature for previous DCEs that evaluated preferences in pharmacogenetic testing related to ADRs.

When systematically searching the literature for DCEs, it is important to include all the terms that are used to refer to the method. Therefore, I first looked for existing systematic reviews of DCEs to inform the search, and combined search terms from these papers to use in the search. These were located by searching the literature for 'discrete choice experiment', 'dce' and 'systematic review', and related words.

Ten previous systematic reviews of DCEs were located, the earliest from 2003 and the newest from 2019 (Appendix Table 9). Four systematic reviews were linked, each using the same methods to update on the field of health related DCEs (30, 39, 40, 60). A total of 1142 DCEs were included across the reviews.

The first of these four was Ryan and Gerard (2003), who systematically reviewed DCEs in a health economics context (40). This was the earliest systematic review to use the search terms that reflect the different terms used to refer to DCEs. These terms were used by updated systematic reviews in 2010 (39) and 2014 (30). The review included 34 experimental DCEs in health economics published from 1990-2000.

The next paper, an updated systematic review of DCEs in health economics was published by de Bekker-Grob, *et al.* in 2012 (ePub 2010) (39). This was further updated in 2014 by Clark, *et al.* (30). The updated study located 179 DCEs for analysis. The authors categorised DCE attributes into six domains: money, time, risk, health status, health care, and other.

The latest update in this series was in 2019, covering DCEs published 2013-2017 (60). Soekhai, *et al.* located 301 papers, showing how the usage of DCEs in the literature continues to grow. The authors utilised data from the previous 3 reviews to evaluate how DCE methods have changed from 1990 to 2017. Changes include the increasing use of software (particularly Ngene) and online survey administration.

Other systematic reviews covered further specific areas of DCE design or of health (Appendix Table 9) (28, 61-65).

These systematic reviews used many different terms to locate DCEs (Appendix Table 10). It was important to include all these terms in the systematic review search, in order to capture all relevant literature. These search terms were combined to form the final search strategy.

The aim of this review was to identify and evaluate all DCEs conducted within the context of pharmacogenetics relating to adverse drug reactions (ADRs). Studies conducted in any population were included, with a focus on examining design features of the studies as well as outcome measures, population metrics, and disease domains. The aim was to collect this information to inform my own subsequent DCE in pharmacogenetic tests for ADRs and highlight potential areas where more research is required. The attributes and levels of different DCEs and their methodological features were examined, such as the inclusion of a 'no test' third option and the comparison of multiple DCEs. I also believe the review will provide a useful overview of the field and guide future practice in a patient- and public-centred manner.

5.2 Methods

Inclusion criteria were any previous DCE in pharmacogenetics that considered the prevention or management of ADRs. The ADR could be the focus of the included pharmacogenetic marker, or an additional attribute. An ADR was defined as any adverse consequence resulting from a drug or intervention, either acute or chronic. Studies in any population and any language were included.

Papers were excluded if they used non-choice conjoint methods (e.g. rating based, best-worst scaling, ranking-based), as were technical/theoretical/methods papers without experimental data, and review articles.

The Medline database was searched on 28th November 2018 using a structured search strategy, informed by searches used by previous systematic reviews of DCEs. The search included terms to reflect the varied terms used in the literature to refer to DCEs as identified in the review of previous systematic reviews, and these were combined with words that encompass the range of terms used to refer to both pharmacogenetics and personalised medicine (66) (Table 5.2). These pharmacogenetics terms were chosen based on my previous review (Chapter 3). The search was repeated (also in the Medline database) on 4th December 2019 during which one new paper was located (67). No limits or filters were used when searching.

1	“conjoint”
2	“conjoint analysis”
3	“conjoint measurement”
4	“conjoint stud*”
5	“conjoint choice experiment*”
6	“part-worth utilities”
7	“functional measurement”
8	“paired comparison*”
9	“pairwise choice*”
10	“discrete choice experiment*”
11	“DCE”
12	“discrete choice mode(l)ling”
13	“discrete choice conjoint experiment”
14	“stated preference*”
15	1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14
16	“biomarker*”
17	“pharmacogenetic*”
18	“pharmacogenomic*”
19	“personalised medicine”
20	“personalized medicine”
21	“precision medicine”
22	“stratified medicine”

23	“genetic testing”
24	16 OR 17 18 OR 19 OR 20 OR 21 OR 22 OR 23
25	15 AND 24

Table 5.2 - terms used when searching the Medline database for DCEs relating to ADRs in pharmacogenetics

Records were screened against inclusion and exclusion criteria by two reviewers by title and abstract (DJ, with AJ checking a random 10%) and then by full text (DJ and AJ). Disagreements were resolved by discussion with the two authors. As the aim of this systematic review was to learn more about DCEs in pharmacogenetics for the purpose of designing a DCE, a third reviewer was not required at this stage to resolve disagreements.

Data was extracted using a standard data extraction sheet prepared for this review (Appendix Table 11), and this was also informed by the aforementioned previous systematic reviews (28, 30, 39, 40), with added items relating specifically to pharmacogenetics. Details of populations, countries of origins, methods of survey administration, and of attributes and levels were collected. Details of DCE design and methods used to select attributes and choice sets were also extracted.

Studies were not examined for risk of bias. Summary effect measures were not appropriate for this review due to heterogeneity in included studies, but the findings have been summarised. The information from all included papers was combined into a summary and the details of each paper were then presented individually. This allowed me to learn from the methods used in each paper.

Analysis and figures were completed in R and R Studio (68).

5.3 Results

5.3.1 Results of systematic literature search

The initial search of the Medline database yielded 565 papers, which was reduced to 23 after removing duplicates and screening by titles and abstracts. After screening by full text, papers were excluded for being technical/theoretical only, and for being review papers. A total of 13 papers remained for analysis (Figure 5.5). During analysis, two of these papers (MacDonald et al (2016) (69) and Marshall et al (2016) (70)) were found to refer to the same survey, consisting of a decisional conflict scale (DCS) and a DCE. MacDonald, *et al.* focuses on the DCS part of the experiment, and Marshall, *et al.* focuses on the DCE. The papers do not refer to each other, presumably due to being published close together in time (in different

journals). The Marshall *et al*. (70) paper was retained as this one focussed on the DCE rather than the DCS. One extra paper was located in a later search (67). The final analysis therefore included 13 papers (59, 67, 70-80). A full detailed list of the papers included can be found in Appendix Tables 12 and 13. As some papers contained more than 1 DCE (72, 74, 78-80), a total of 19 experiments were included.

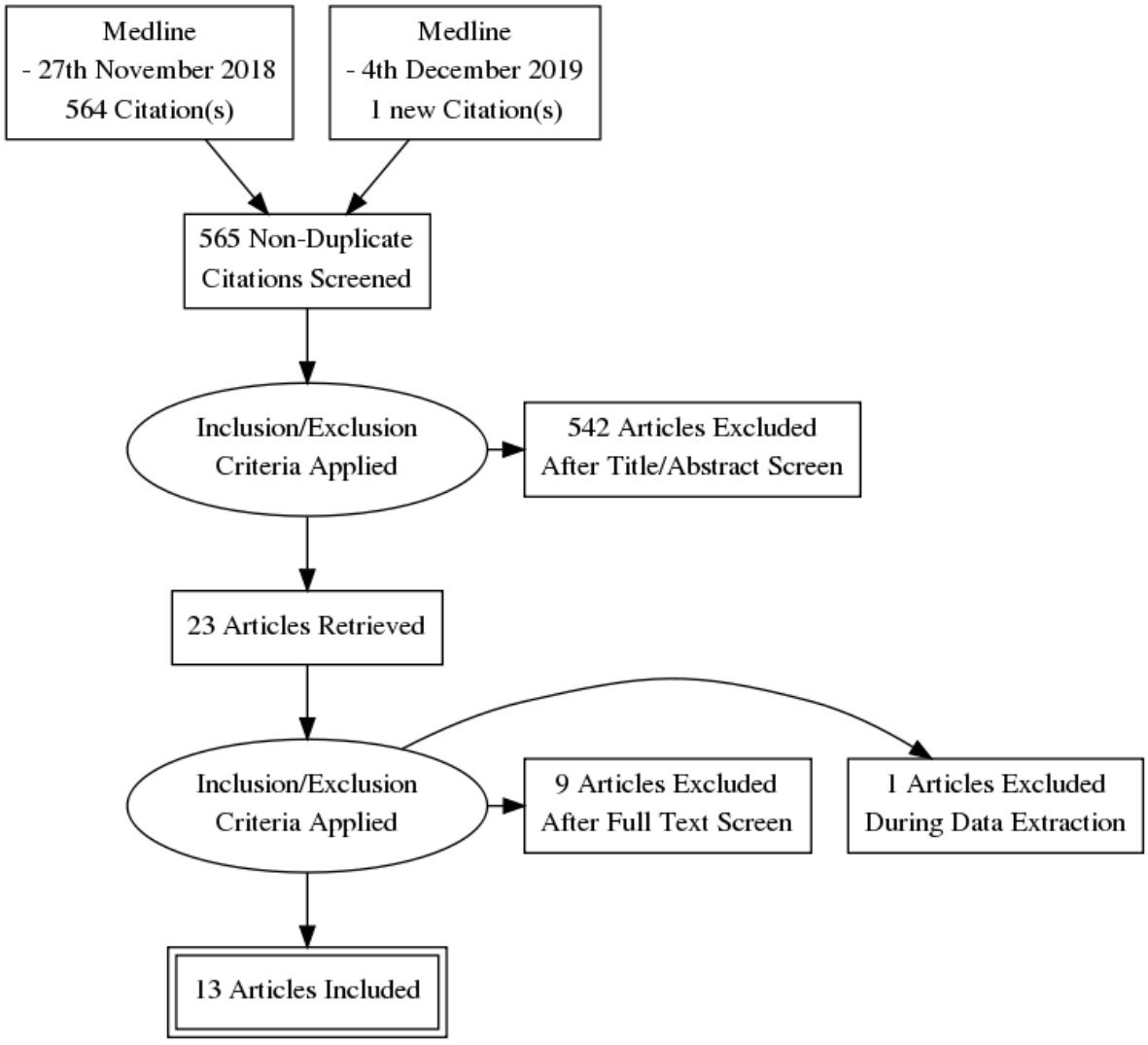


Figure 5.5- PRISMA flow chart (81, 82). Produced with an online tool at http://prisma.thetacollaborative.ca/generator (83)

DCEs represented cancer (59, 70-73, 78), psychiatric disease (67, 76), cardiovascular disease (74), autoimmune disease (79), epilepsy (80), and gout (75). One paper did not specify a disease area (77). Papers were published between 2009 and 2018 and were conducted in patients, healthcare professionals (HCPs), and the general public. ADRs were mostly considered in terms of risk, although

some included additional parameters such as ADR severity and duration. Full details of the included papers can be found in Appendix Tables 12 and 13.

5.3.2 Demographic details

The USA was the most represented country, with 4 papers. Other countries represented were the UK, Singapore, Canada, and Denmark. One paper was an international collaboration (59). Five papers were in the field of breast cancer (59, 70-73), with the remaining papers in autoimmune disease (79), cancer (78), cardiovascular disease (74), depression (76), epilepsy (80), gout (75), and schizophrenia (67). One paper did not examine any particular disease area (77). The most commonly recruited population was patients (those with the condition being studied), followed by the general public, then healthcare professionals. Most surveys were administered online. Several used market research companies to collect participants (70-72, 76, 78). The mean sample size was 440 (median 323), with a range of 67 – 1096 (Figure 5.6).

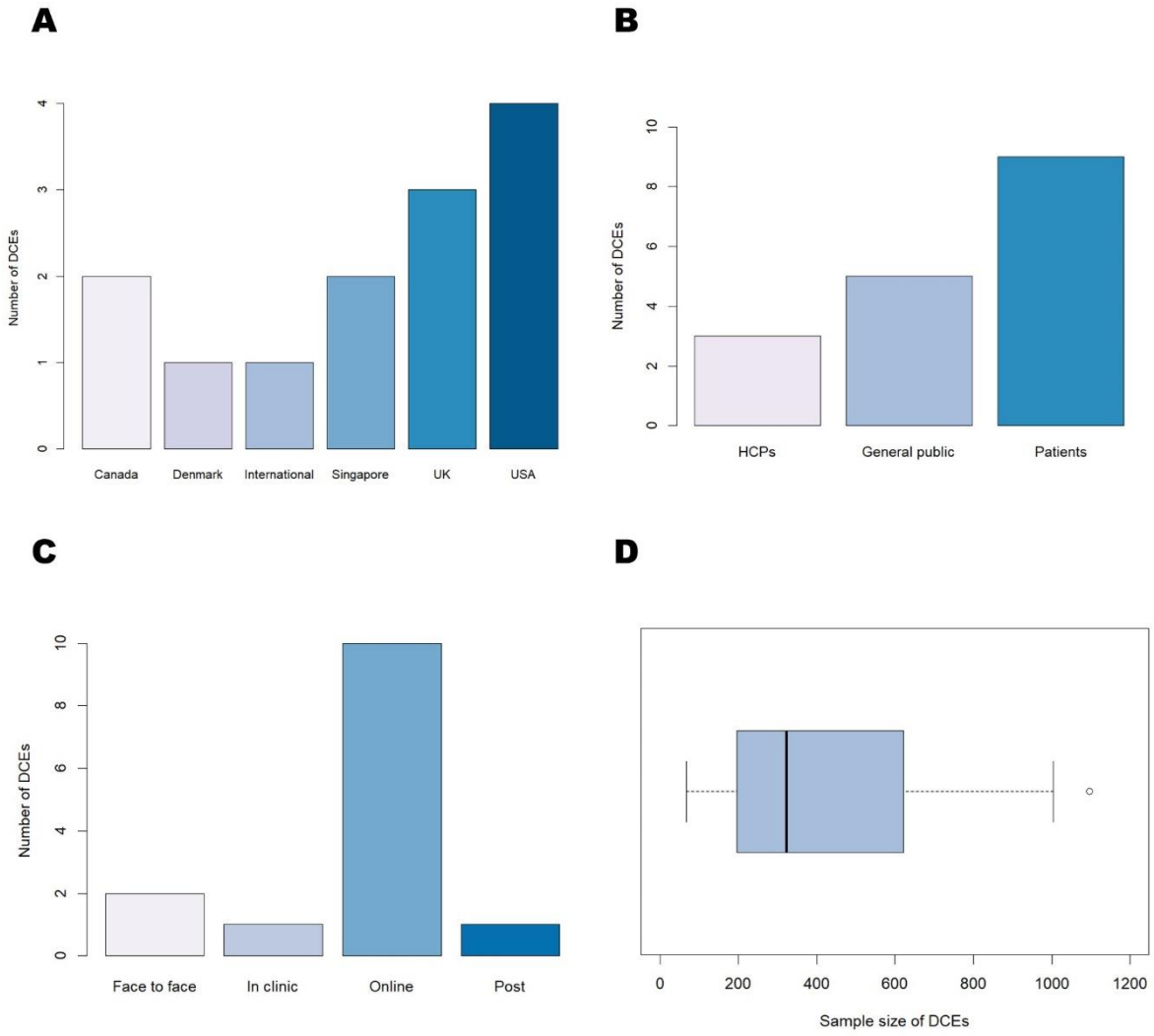


Figure 5.6 - Basic information on the DCEs included in this review. A) Location of DCEs, B) Population included in the DCE as participants, C) Method of DCE administration. 'Face to face' refers to where the DCE was conducted with support of a researcher. 'In clinic' refers to a DCE that gave out questionnaires for patients to complete in the clinic, without support, D) Boxplot showing the range in DCE sample sizes. Totals may sum to >10 as some papers included more than 1 DCE. HCPs = healthcare professionals

Participant ages and genders were averaged from papers that provided the requisite information. The mean age of the participants was 49.3, and 63.5% were female. Two studies recruited 100% female participants (59, 70). When these were excluded, the average percentage of female participants fell to 54.4%.

5.3.3 Attributes and levels

Attributes were sorted into domains as used by Clark, *et al.* (2014)(30), de Bekker-Grob, *et al.* (2012) (39), and Soekhai, *et al.* (2019) (60). These domains originated with Ryan & Gerard (2003) (40). These are: monetary measures (e.g. cost of test), time (e.g. time in hospital), risk or probabilities (e.g. of toxicity, likelihood of benefit), health status (e.g. side-effect severity), health care (e.g. route of drug administration), and other. Of 68 extracted attributes, risk was the most common type (41.2%, 28 attributes) (Figure 5.7). The most common number of attributes to include in a DCE was 4 (mean 5.1, range 3 – 7). A no-test option was included in 46% of papers (6/13). Of these, 3 reported the rates at which participants consistently chose no-test (9% (75), 12% (80) and 13.3% (76)).

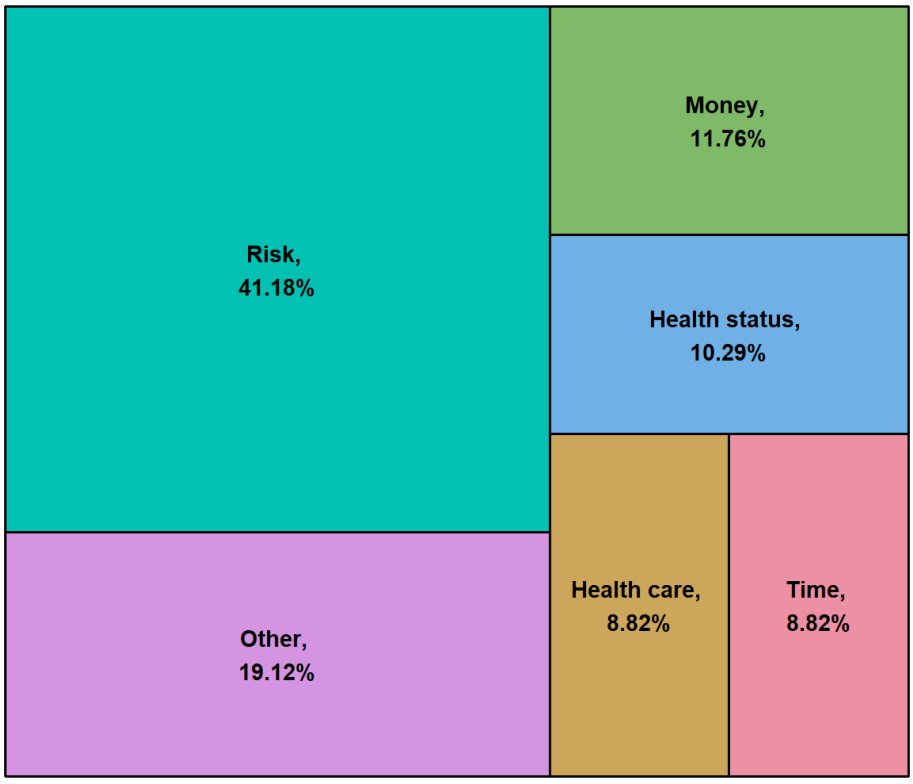


Figure 5.7 - Domains of the attributes of located DCEs.

The mean response rate, in studies that reported it was 58.6%; however, 5 papers did not report response rates. Only one DCE did not report details of qualitative work (71). Of the others, interviews, pilot testing, and seeking expert opinion were other reported types of qualitative work. These were most commonly performed for attribute selection and pre-testing of the DCE (Figure 5.8). Further details of the

methods used to choose attributes and levels are included in the discussion of each paper in detail.

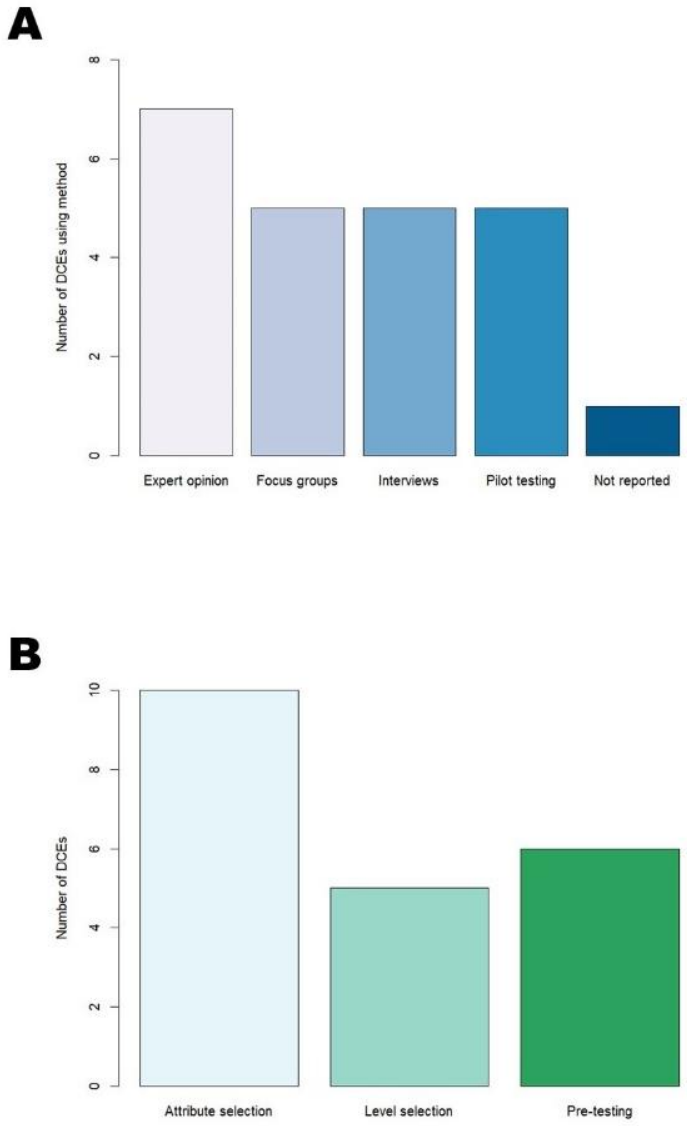


Figure 5.8 - Qualitative work of DCEs included in the review. A) Types of qualitative work done, B) When in the DCE development process that qualitative work was performed.

In terms of DCE design, the most common choice was a fractional factorial design. Four papers did not report their design (71-74). The most common method for creating choice sets was the use of D-efficiency. Two DCEs did not report their method for this (73, 77). Sawtooth and SAS were the most commonly used software for DCE design, and four DCEs did not report which they used (71, 73, 79, 80). Around half the DCEs included a ‘no test’ or ‘neither’ option, where participants

could choose none of the choices in a question (71, 75-77, 80, 84). The median number of choice tasks for each participant was 12 (range 4-26), and 8 was the most common number of choice tasks.

Included DCEs have been split by their target population for reporting. I report full details of DCE design, attributes and levels (and their selection), and the learning from each paper applicable to my own DCE. These data are summarised in Appendix Tables 12 and 13.

5.3.4 DCEs in patients

Five DCEs recruited patients as their target population (59, 71-73, 75) – defined as those suffering from an illness, including the illness being investigated in the DCE.

5.3.4.1 Ballinger et al 2017

This DCE recruited 417 $HER2^-$ breast cancer patients in the USA currently undergoing treatment and asked them to choose between four different chemotherapy regiments with different levels of relative reduction in risk of ADRs (Table 5.3) (71). The DCE also incorporated a biomarker analysis by varying the risks of each toxicity and the likelihood of benefit based on hypothetical biomarkers for peripheral neuropathy and congestive heart failure. Participants could also choose to not receive either treatment. The aim was to examine patients' preferences for treatment and determine if they were willing to trade toxicity and benefit. The authors did not specify any qualitative work performed to develop the DCE.

Attribute	**Levels**					
Peripheral neuropathy likelihood (%)	0	10	15	20	40	60
Relative recurrence risk reduction (%)	20	30	35	40	50	
Peripheral neuropathy severity/ duration	Severe/ during treatment	Moderate/ a year	Moderate/ rest of your life	Severe/ rest of your life		
CHF likelihood (%)	0	1	5	10		

Table 5.3 - Attributes and levels of Ballinger et al 2017 (71). CHF = congestive heart failure

To prevent bias, this survey did not name the specific drugs used to define the levels. Benefit and risk profiles of the drugs were based on previously published clinical trial evidence. Prior to the DCE portion of the survey, participants were asked to choose their perceived risk of breast cancer recurrence without chemotherapy. This perceived risk was then used to customise the subsequent DCE so the reduction in risk of recurrence was relative to each participant's perception of their recurrence risk. Participants were also asked about their previous experience of chemotherapy and toxicities.

Details of DCE design were not provided. Choice sets were chosen 'pragmatically' (details were not provided) and analysis done using a hierarchical Bayesian routine. The survey was conducted online. The response rate was not provided.

Of the participants, 88% were Caucasian and 65% were aged 50 or over. Specific details of participants' ages were not provided, so this DCE was excluded from the above calculation of the mean overall age in DCEs. All patients had completed chemotherapy and 90% had been diagnosed with breast cancer more than 1 year before completing the DCE.

The largest shifts in preference were caused by recurrence risk reduction and the likelihood of peripheral neuropathy. Participants that initially had a higher perceived risk of breast cancer recurrence were more favourable towards regimes with greater toxicity (and accompanied higher likelihood of benefit). However, participants with previous experience of peripheral neuropathy were more likely to choose a chemotherapy regime with moderate risk of peripheral neuropathy (and higher likelihood of benefit), than one with no risk of peripheral neuropathy.
Modelling using the hypothetical biomarker data showed that patients homozygous for a variant that confers a higher risk of an ADR would be more likely to choose a chemotherapy regime with a lower risk of the ADR. This shows that participants are able to perceive changes in risks due to pharmacogenetics, and that this modifies preferences. This work also shows that previous experience with ADRs affects preferences in these scenarios. A question asking if participants have previously experienced a disease is a useful addition to a DCE.

5.3.4.2 Dong et al 2016

A 2016 DCE in Singapore measured patient preferences for avoidance of a severe ADR (Stevens-Johnson syndrome, SJS), in the context of allopurinol treatment for

gout (75). Allopurinol is the main treatment for gout but is associated with a risk of SJS, particularly in patients with the *HLA-B*58:01* allele (85). The authors aimed to examine patient preferences for genetic testing in this context in order to inform policymakers and clinicians. Attributes were selected after in-depth interviews with 10 patients and the survey was pre-tested in interviews with a further 50 patients. Attributes and levels are summarised in Table 5.4.

Attributes	**Levels**			
Chance of getting severe side-effect	1 out of 1 million	1 out of 5000	1 out of 1000	1 out of 600
Cost of genetic test (SGD)	20	200	400	1000
Cost of gout medicines over 2 years if test positive (SGD)*	250	400	1500	4000
Your physician's recommendation	No information	An alternative is the physician recommendation	An alternative is not the physician recommendation	
Most common choice	No information	An alternative is the most common choice	An alternative is not the most common choice	
Cost of gout medicines over 2 years if test negative † (SGD)	200			

*Table 5.4 - Attributes and levels of Dong et al 2016 (75). SGD = Singapore dollars. *Chance of a test positive is 2 in 10. †Chance of a test negative in 8 in 10.*

Patients were asked to choose between two treatment options that included genetic testing, and a third treatment choice without genetic testing. One of these options was marked with a 'doctor recommended' banner. The 'cost of gout medicines over 2 years' attribute was split by probabilities. In the event that the test was positive for the risk allele (2 in 10 chance), patients would not be able to use allopurinol to manage gout, incurring higher medicine costs for alternative drugs. If the test was negative (8 in 10 chance), medication costs would be lower. Patients in Singapore pay for healthcare, although it is subsidised by the government (86).

The survey was designed using SAS software with a fractional factorial main and interaction effects design. Choice sets were created to maximise D-efficiency, and a latent class logit model was used for analysis. The survey was conducted in-person with the help of an interviewer. The response rate was not reported.

The authors recruited a convenience sample of 189 diabetic patients from diabetes clinics. This population was chosen as they are at a higher risk of developing gout than the general population (87), but would not currently be prescribed allopurinol. Similarly, males were oversampled (65.6% of the survey population) since gout is more common in men. Respondents were 61.4% Chinese, with a mean age of 57.1 A small number of participants (5.8%) had gout but did not receive medication for it.

Modelling the results of the DCE revealed two groups of participants – the 'risk averse' and the 'cost conscious'. The 'risk averse' group, comprising 63% of respondents, always preferred to test, with cost of test having a minimal impact on their decision. This group were willing to pay up to S$1215 (£675 as of January 2021) to reduce the risk of developing SJS. Meanwhile, the 'cost conscious' group were more sensitive to the cost of the test (WTP was not calculated as the risk coefficients did not significantly differ from 0). Across both groups, the recommendation of a clinician was a significant predictor of test uptake. A combination of this recommendation and an option being the most common choice amongst other patients increased uptake of the genetic test more than a 75% reduction in test cost did.

A simulation of uptake rates calculated an uptake of 65.1% for the 'most realistic' testing scenario, but this was significantly different between risk averse (>95%) and cost conscious (8.8%) groups. This is very useful for policy making in this context, and analysing both groups separately provides a more accurate picture of uptake rates. This highly policy-relevant measurement of uptake is a value I will aim to output from my own DCE.

5.3.4.3 Issa et al 2013

This DCE evaluated the preferences of US breast and colorectal cancer patients for diagnostic genetic testing (72). The example used for breast cancer testing was the commercial Oncotype DX scoring system. The score is highly correlated with the likelihood of breast cancer recurrence, and this is used to guide breast cancer treatment decisions (such as the decision to use chemotherapy or more conservative treatment options) (88). Individual mutation testing for *KRAS* and *UGT1A1* was examined in colorectal cancer. The authors aimed to quantify patient

willingness to pay for these tests and examine their preferences for the characteristics of these tests. Attributes and levels (Table 5.5) were developed by conducting six focus groups with breast or colorectal cancer patients (n=44). These results were published in a 2011 paper (89). Participants were broadly in favour of pharmacogenetics and personalised medicine, but expressed concerns about the evidence behind tests ("I would need more data, specifics, how long has it been tested, how many people, … everything. I would like to look at it entirely before I agree to [being tested]"). Privacy and data security were also common concerns ("Shouldn't it – who knows the results – be limited to the medical field, healthcare providers or something?").

Attribute	**Levels**					
Cost of testing to you personally (US$)	25	100	500	1000	2000	4000
Chance the test will correctly predict response to treatment	55%	70%	80%	90%	96%	99%
What information will the test provide?	Recurrence risk and how likely you are to benefit from chemotherapy and how likely you are to develop severe side effects from chemotherapy	Recurrence risk and how likely you are to benefit from chemotherapy	Recurrence risk and how likely you are to develop severe side effects from chemotherapy	How likely you are to benefit from chemotherapy and how likely you are to develop severe side-effects from chemotherapy	How likely you are to benefit from chemotherapy	
Who has access to your test results	Patient and doctor	Patient, doctor, and	Patient, doctor, insurance			

		insurance company	company, and employer
How will your test results be used?	Doctor will decide how best to treat you based on your results	You will decide how to use the results, regardless of your risk of recurrence	Your insurance company will use the test results to determine your coverage

Table 5.5 - Attributes and levels of Issa et al 2013 (72)

Both breast cancer and colorectal cancer patients completed the same questionnaire. The survey was designed using Sawtooth Software, but the specific design was not reported. Choice sets were created using random pairing. The analysis model was also not reported. The survey was conducted online, with a response rate of 42.2%.

Of the 300 participants that were recruited, 150 were breast cancer patients and 150 were colorectal cancer patients. Mean ages were 54.5 years and 42 years, respectively. Both populations were majority Caucasian or white. Breast cancer patients were more knowledgeable about tumour-specific genes than colorectal cancers in a pre-DCE survey.

Test accuracy was the most important attribute across all patients. Accuracy >90% was associated with positive preferences. When it comes to the cost of the test, 22.5% of patients were willing to pay for testing, with willingness falling when costs exceeded $500. Participants had a strong preference for only the patient and doctor having access to results, and a strong negative preference for an insurance company using their test results to determine coverage.

This DCE focussed on patient-relevant outcomes and revealed interesting similarities in preferences for genetic testing in two different groups of patients. The US-centric nature of the DCE, with insurance companies as part of two different attributes, limits its use outside of these contexts. However, the finding that patients cared highly about test accuracy is important, and underlines the importance of communicating this level of detail to patients. The detailed reporting of focus groups

used to determine attributes and levels is also useful for my DCE design. The themes of evidence and privacy will be interesting to explore from a UK perspective.

5.3.4.4 Liede et al 2017

This breast cancer DCE was an international study that recruited carriers of the *BRCA1* and *BRCA2* genes to examine preferences for preventative treatments (59). Those who are positive for these genes have an increased risk of breast and ovarian cancers (90). The aim was to examine preferences for treatments (such as hormone therapy and preventative surgeries) that reduce their risk but that may have associated ADRs (such as effects on fertility or uterine cancer risk). The attributes for this DCE (Table 5.6) were developed by consulting clinical experts and the survey was pre-tested in interviews with potential participants. No further details of qualitative work to develop the DCE were provided.

Attribute	**Levels**			
Reduction in risk (%)	90	75	50	40
How long you take the medicine (years)	1	3	5	
Effect on fertility	No effect	Cannot get pregnant during treatment	Can never get pregnant	
Effect on female hormones	No effect	Temporary menopause-like symptoms	Early menopause	
Risk of teeth and jaw problems	No risk	1%	5%	
Route of administration	Daily pill	Injection every 3 months	Injection every 6 months	
Risk of uterine cancer	No risk	1%		

Table 5.6 - Attributes and levels of Liede et al 2017 (59).

The attributes and levels represented four treatment options – mastectomy, oophorectomy, and two different medications. Participants could also choose to receive screening only. Participants were also asked what treatments they had

already undergone to reduce their risk of breast and ovarian cancer. The authors were able to use this information to test the external validity (if results can accurately predict choices outside of the survey (41)) of their DCE.

The DCE was designed using SAS software in a fractional factorial design. Choice sets were created in a D-efficient manner and analysed using a random-parameters logit model. The survey was administered online and had a response rate of 53.5%.

A total of 622 participants were recruited in the USA (56%), Australia (20%), UK (19%) and Canada (5%) through research registries and patient organisations. The mean age of participants was 41. Details of race or ethnicities were not provided. Most participants had taken measures to reduce their cancer risk, including mastectomy (49.2%), oophorectomy (52.3%) and the use of prescription medication like tamoxifen (5.5%).

The most important attribute to participants was the reduction in risk of breast cancer. However, this differed among the 32% of women who wanted to have more children. In these women, the effect of the intervention on fertility was the most important attribute. Other attributes had lesser effects on preferences. The effect on female hormones, and the risk of teeth and jaw problems, only affected preferences at their highest values. The type of ADR can be said to impact on preferences, and testing multiple ADRs may better capture clinical realities.

The DCE showed that women preferred mastectomy to prescription medicines, but preferred the medicines to oophorectomy. However, when examining real-world choices, women were more likely to have undergone an oophorectomy than to have taken any prescription medication. Reasons for this included: their doctor had not recommended medication; concerns about side-effects; and the association of the medications with cancer treatment. This shows the importance of including external validity checks in a DCE in scenarios where this is possible.

5.3.4.5 Smith et al 2014

This DCE tested patient preferences for two different chemotherapy drugs (paclitaxel and capecitabine) in the USA (73). Metastatic breast cancer patients were recruited through patient advocacy organisations. The aim of the DCE was to evaluate preferences for drugs with different toxicity profiles and routes of administration, while also varying the likelihood of benefit to simulate the use of a biomarker to predict benefit and toxicity likelihoods. For example, the likelihood of benefit from paclitaxel was 50% when using the biomarker, compared to 20%

without the use of the biomarker (Table 5.7). A pilot study in 2010 was the only reported qualitative work prior to the DCE. Participants were shown two different profiles and asked which treatment they would choose, then asked if they would take that treatment over no treatment.

Attribute	Levels (paclitaxel)		Levels (capecitabine)	
	Toxicity BM	Benefit BM	Toxicity BM	Benefit BM
Route of administration	IV	IV	Oral	Oral
Likelihood of benefit [BM range]	33%	20% [50%]	27%	13% [40%]
Likelihood of toxicity [BM range]	27% [60%]	27%	10% [40%]	10%
Toxicity type/ severity/duration	Moderate PN, 1 year	Moderate PN, 1 year	Severe diarrhoea, during treatment	Severe diarrhoea, during treatment

Table 5.7 - Attributes and levels of Smith et al 2014 (73). BM = biomarker, IV = intravenous, PN = peripheral neuropathy

This DCE did not report any details of design, software, methods used to create choice sets, or analysis. The only analysis reported was the percentage of respondents choosing treatment under each scenario. The survey was administered online, and the response rate was not reported.

A total of 641 respondents were recruited, and were 99.7% female and 90.6% Caucasian. Most had been diagnosed over 1 year prior to the DCE and were undergoing treatment. 47.6% and 43.5% of women had previously received paclitaxel and capecitabine, respectively.

Respondents were more likely to choose a treatment with greater likelihood of benefit and lower likelihood of toxicity. Preferences were more sensitive to changes in likelihood of benefit than risk of toxicity. Subgroup analyses revealed that women who were younger, and those that had children, were more likely to choose treatment. Biomarker modelling showed that benefit biomarkers had more influence on decision making than toxicity biomarkers. A biomarker predicting benefit from capecitabine had the largest individual effect of the biomarkers tested. Overall, most respondents indicated they would choose the treatment on offer, regardless of

biomarker information. However, the profile of the biomarker (how well it predicts benefit and toxicity) has a large impact on this decision. The four biomarkers tested all caused a decision change in respondents in the range of 5.4% to 34.1%. This shows that participants are able to comprehend and manage decisions in light of genetic biomarker information and that both benefit and toxicity are important (although prediction of benefit is more important).

5.3.5 DCEs in HCPs

One DCE was located that recruited only healthcare professionals (67).

5.3.5.1 Boeri et al 2018

This DCE recruited UK psychiatrists to evaluate preferences for prescribing drugs to patients with schizophrenia (67). The aim was to determine the overall maximum acceptable risk (MAR) acceptable in exchange for an increase in benefit of a treatment. This survey was unusual in two ways – firstly, it only included health professionals, and secondly, most of the survey attributes were continuous variables. The authors did not include full details of the qualitative work but specified that they consulted 2 practising psychiatrists (included in the list of authors) for attribute and level selection, and pre-tested their survey among psychiatrists. Attributes and levels are shown in Table 5.8.

Attribute	**Levels**	
Patient has hyper-responsiveness genotype	Yes	No
PANSS score change	**Continuous** 3 – 26	
Number of acute treatment days in hospital	**Continuous** 17 – 45	
Risk of 10 kg weight gain	**Continuous** 30 – 70%	

Table 5.8- Attributes and levels of Boeri et al 2018 (67). PANSS = positive and negative syndrome scale. Higher scores indicate more severe symptoms (91).

Biomarker status was indicated with a binary variable. Participants were asked to choose between two different treatments for a patient, and asked to rate their confidence in their judgement. The hyper-responsiveness genotype was a hypothetical attribute that influenced a patient's drug response.

This DCE was alone in using NGene software for design. The authors used a fractional factorial design and maximised D-efficiency. Analysis was performed using a random parameters logit model. The survey was administered face to face, in professional development meetings. Of 70 psychiatrists invited to participate, 67 were included in the final sample (response rate of 95.7%).

The participants were 59% male, with an average of 10 years of clinical experience in their speciality. No details of age, race or ethnicity were provided.

The psychiatrists were significantly more likely to recommend treatment when lower (better) Positive and Negative Syndrome Scale (PANSS) scores were achieved. As the risk of ADRs (weight gain) or number of days in hospital increased, they were less likely to recommend treatment. The MAR of each psychiatrist (the percentage increase in risk of weight gain that they were willing to accept in exchange for a one-unit decrease in symptoms on the PANSS) ranged from 0.5 – 9.5%. Genotype information was not found to significantly influence decision making in this context. Subgroup analysis revealed that more experienced psychiatrists were less likely to consider genotype when making decision than those with less than 1 year of clinical experience. This was explored in detail in a separate paper (92). The results of this DCE show that HCPs are willing to accept genetic testing for ADRs, but that practitioners with less experience may be more likely to use it in practice.

HCPs were not included in my DCE. However, I included them in the qualitative work to choose attributes and levels, in order to ensure that results will be clinically meaningful and useful.

5.3.6 DCEs in the general public

Three DCEs recruited only members of the general public (70, 76, 77).

5.3.6.1 Herbild et al 2009

This second DCE in psychiatry was conducted in Denmark with the general public (76). The aim of the survey was to estimate preferences for pharmacogenetic testing in treating depression. Depression is treated with antidepressants, all of which have similar common associated ADRs (93, 94). Variation in *CYP2D6* is associated with varying responses to treatment and rates of ADRs (94-96). The survey was designed using three patient focus groups, and the results of these were published in 2007 (42). Themes of frustration with ADRs (“I don’t really know whether I dare change from the product (pharmaceutical) I’m taking now in order to hope for less side-effects,[…] could be that I’ll get even worse or return to my

depressive state - and that thought is unbearable"), and data privacy ("It might lead to some sort of a rollercoaster effect and all of a sudden you won't be able to get yourself a life-insurance") also appeared in these discussions.

Expert opinion was also utilised in the design stage. The attributes and levels of this DCE are summarised in Table 5.9.

Attributes	**Levels**							
Price of the test (in Danish Krone)	200	600	1000	1500	3000	6000	9000	18000
Likelihood of improvements from the test)	10%	50%						
Number of changes in antidepressant medication	2	3						
Time with dosage adjustments due to lack of effect and/or unacceptable ADRs	1 month	3 months						

Table 5.9 - Attributes and levels of Herbild et al 2009. ADR = adverse drug reaction.

Participants were asked to imagine they had been diagnosed with depression and to choose between scenarios representing treatment with antidepressants, with and without pharmacogenetic testing included.

The survey was administered online, with a response rate of 46%. SAS software was used to design a DCE with a fractional factorial design, maximising D-efficiency. Conditional logistic regression was used to analyse the results.

A total of 323 members of the Danish general public completed the survey, with a gender split of 53% female and 47% male. Age was not reported, although the authors note that those aged 18-24 were underrepresented. There were 19 respondents (5.9%) who did not like the idea of pharmacogenetic testing and would never want to be tested. The remaining respondents were willing to pay 1571 Danish Krona (DKK) (90% CI 809 – 2331) for a 10% likelihood of reducing the number of antidepressant changes. The WTP for a 10% likelihood of reducing the time with dosage adjustments was DKK604 (90% CI 230 – 986). Subgroup analysis showed no significant difference in results in participants previously diagnosed with depression.

The authors concluded that the mean WTP exceeded the usual cost of the pharmacogenetic test. However, other costs associated with testing (such as labour, materials) were not taken into consideration, and the authors recommended future analyses include these assessments. Participants valued reducing the number of antidepressant changes higher than reducing the time with dosage adjustments. This information is potentially useful for informing a clinical strategy for treating depression. The results of this DCE also show that participants living in countries with healthcare free at the point of service can consider the impact of cost when making decisions. Using WTP as an output in these countries has its limitations (10, 31), but is beneficial for calculating monetary costs of services and estimating trade-offs with other attributes.

5.3.6.2 Marshall et al 2016

This DCE investigated the impact of gene expression profiling (GEP) on decision making for chemotherapy in early-stage breast cancer (70). In this setting, there are concerns that many patients are over-treated, since only an estimated 15% of those treated with chemotherapy will experience a recurrence of their cancer (97). The aim of this DCE was to examine the preferences of Canadian women in the general public for chemotherapy, with and without GEP scores as guidance. Extensive qualitative work was reported for developing the DCE and results were published in journals (98-100). Focus groups and interviews were carried out, recruiting oncologists (99), and women with a history of breast cancer (98, 100). Oncologists mostly viewed GEP as a tool that enhanced their confidence in risk assessment, with one describing the results as a “tie-breaker” in difficult decisions. However, patients viewed the test as ‘more scientific’ and ‘magical’. Evidence behind the test was also considered, with the fact that the test was covered by insurance being

taken as presumed evidence of its validity ("I had no idea there was even another world out there that wouldn't be supporting the test… the fact that's covered [sic] tells me that it's absolutely supported, right?")

After developing attributes from this process, the clinical face validity of these was checked by medical oncologists. The survey was then pre-tested in women from the focus group, and in the general population. Additionally, the survey was translated to French for use in French-speaking areas of Canada. Final attributes and levels are shown in Table 5.10.

Attributes	**Levels**			
GEP test score (likely benefit from chemotherapy)	9 (low)	22 (uncertain)	44 (high)	GEP test not available
Doctor's estimate of cancer returning (without GEP test)	Low	Intermediate	High	
Trust in doctor	Do not trust	Somewhat trust	Totally trust	
Likelihood of temporary side-effects	Low	Moderate	High	
Likelihood of permanent side-effects	Low	Moderate	High	

Table 5.10 - Attributes and levels of Marshall et al 2016 (70). GEP = gene expression profiling

Since the survey population was the general public, background information on breast cancer and GEP testing was presented to respondents prior to the DCE. Respondents were asked to imagine they had early-stage breast cancer, and indicate scenarios under which they would be most likely to choose chemotherapy. These included scenarios with varying GEP test scores, and with no GEP test. Other attributes included the likelihood of temporary side-effects (such as nausea, vomiting, numbness in fingers, hair loss, fever, and infection) and the likelihood of permanent side-effects (leukaemia, heart muscle damage, early menopause). This was the only DCE to use categorical descriptions of ADR risk (low, moderate, high).

The DCE was designed using Sawtooth software with a fractional factorial main and interaction effects design. Choice sets were created to maximise D-efficiency and

analysed using a hierarchical Bayesian routine. The survey was administered online, and authors did not report the response rate.

A total of 1004 Canadian women with a mean age of 49 were recruited to the survey. Respondents were 84% white, and 49% reported having a relative with or who had suffered from breast cancer. The DCE indicated that the GEP test score had the greatest importance to participants when making decisions about chemotherapy. This was considered even more important than an estimate of risk given by a clinician. The likelihood of temporary side-effects was the least important attribute.

The authors used the data from the DCE to estimate chemotherapy uptake rates in high-, moderate- and low-risk profiles (according to GEP test scores). These uptake rates (78%, 55%, and 33% respectively) are important data with a potential direct impact on health policy. Use of the GEP score alongside clinician assessments may reduce the risk of overtreatment in breast cancer patients. The learning from this DCE is the discussion of evidence in the published qualitative work. Patients in these groups assumed that if the test was covered by insurance, its validity was assured. While this is not necessarily the case, it provides a potentially useful shorthand for 'high' levels of evidence. In my DCE, I have used a similar shorthand – a high evidence test is one that is widely used and recommended by several countries' health authorities.

5.3.6.3 Marshall et al 2017

This DCE was contained within a technical paper that focussed on the design challenges of estimating preferences for whole genome sequencing (77). The authors used a DCE to illustrate some of the issues and offered potential solutions. For example, one issue encountered is that genetic testing often offers multiple cascading uncertainties e.g., if the test is positive, this cost is encountered, and if the test is negative, a different cost is encountered. The authors proposed simplifying these problems as much as possible, and randomising patients at different decision points. Participants in this DCE were randomised between two scenarios before making choices on attributes and levels (Table 5.11). A total of 410 members of the general public completed the survey. The survey was designed using expert opinion and 13 interviews. No details of qualitative methodology or results were given.

Attribute	Levels		
Chance that a side-effect makes you unable to do everyday activities or take care of yourself	None	5 out of 100 (5%)	20 out of 100 (20%)
Ongoing out of pocket cost	None	$50 a month ($600 a year)	$200 a month ($2400 a year)
Follow-up requirement	Check-up every 6 months	Check-up every year	Invasive test every year

Table 5.11 - Attributes and levels of Marshall et al 2017 (77).

Participants were randomised to receive either a favourable scenario (with a low chance of death and better quality of life) or a less favourable scenario (higher chance of death, poor quality of life) as context for their decision making. They were also randomised between choosing between different medications, or choosing between different surgeries. They were then asked to choose between two medications or surgeries with differing levels. They could also choose 'watchful waiting', an option with no interventions but a check-up every 6 months.

The survey was designed using an SAS algorithm. The method of choosing choice sets was not specified. Results were analysed using a random parameters logit regression model. The survey was administered online and had a response rate of 47.0%. The chance of an ADR was presented with pictograms of 100 people, with the number of affected coloured in (e.g. 5 coloured in to represent 5%).

Demographic information of the participants was not reported. Preferences were consistent between medicine and surgery options, although participants in the surgery arm were more likely to choose watchful waiting over an intervention. This DCE was unique in describing the ADR in a functional way (the impact on quality of life). This could theoretically allow comparison of different ADRs, that have the same impact. Although the scenarios presented here are likely too general to form evidence for any specific intervention, this DCE case study is a good example of how to present complex information to participants relating to ADR risks and quality of life. The pictograms are a good way to present information. I investigated their use further in later qualitative work.

5.3.7 DCEs in multiple populations

Four DCEs included multiple populations – patients plus the general public (74, 78), and patients plus HCPs (79, 80).

5.3.7.1 Chan et al 2013

Warfarin genetics is a well-researched area within pharmacogenetics with several studies showing its effectiveness (101-104). This DCE in Singapore investigated the WTP of both warfarin patients and the general public for a genetic test for warfarin dosing (74). The aim was to quantify the value of genetic testing to discuss its suitability for widespread use. Attributes (Table 5.12) were defined using pilot testing with patients, and the questionnaire was further piloted in 10 patients before wider use. Methods for choosing attributes and levels were not provided in further detail.

Attribute	**Levels**			
Cost of test (S$)	100	225	375	600
Number of INR tests needed before dose stabilisation	5	13	21	
Risk of serious side-effects (% per year)	1	5	9	
Nature of test	Genetic	Non-genetic		

Table 5.12 - Attributes and levels of Chan et al 2013. INR = International Normalised Ratio. S$ = Singapore dollars

Participants were presented with two hypothetical tests, either genetic or non-genetic, and asked to choose just one. There was no 'opt-out' option. Participants were then asked if they would actually take the test they had chosen.

Sawtooth software was used to create the choice sets, but a specific design type or method was not reported. The authors analysed data using a hierarchical Bayes method. The DCE was completed online by the general public and in anticoagulation clinics by warfarin patients. Response rates were 83.5% (general public) and 53.8% (patients).

Of the 197 warfarin patients who completed the DCE, the mean WTP for a genetic test over a non-genetic test was S$91 (around £50 in January 2021). In contrast, the mean WTP in the 187 members of the public who completed a survey was S$20 (£11). The risk of serious side-effects (major bleeding or clotting) was the most important attribute to both groups. Patients and the general public were willing to pay S$63 and S$109 respectively, for every decrease in percentage risk of side-effects. This shows that genetic testing for warfarin dosing may be economically acceptable in Singapore, although differences between groups show this is context-dependent.

This DCE provides an interesting examination of genetic exceptionalism. Participants were willing to pay more for a genetic test over a non-genetic test. However, the authors noted that many of their participants had difficulties understanding the genetics of testing. Further qualitative work to understand the difference in WTP between genetic and non-genetic testing would be a valuable addition.
I have included a warfarin example in my DCE. This paper provides a useful blueprint for the presentation of warfarin and associated ADRs for presentation to the general public.

5.3.7.2 Najafzadeh et al 2013

This DCE, conducted in Canada, examined the preferences of the general public for a hypothetical genetic test to guide cancer treatment (78). The DCE also included a small number of current or former lymphoma patients. The DCE aimed to see how preferences differed between these two groups, and how the type of cancer and its prognosis affected preferences (Table 5.13). The authors designed the survey using expert opinions of physicians who worked with cancer patients, and three pilot surveys. These were conducted in patients (n=7) and the general public (n=50) and were also used to inform expectations of directions of effect.

Attribute	**Levels**				
Genetic test procedure	Mouth swab	Blood sample	Tumour biopsy	Bone marrow biopsy	Liver biopsy

Untreated responders (%) [1 – sensitivity]	5	20	35	50
Unnecessary treatment of non-responders (%) [1 – specificity]	5	20	35	50
Genetic test cost ($)	50	500	1000	1500
Severity of side effects	Severe	Moderate	Mild	
Likelihood of side effects (%)	5	50	95	
Genetic test turnaround time	2 days	7 days	12 days	

Table 5.13 - Attributes and levels of Najafzadeh et al 2013 (78). Untreated responders = % of patients that would be cured but will not receive it because of an inaccurate genetic test result. Unnecessary treatment of non-responders = % of patients who would not benefit from new medication, but will receive it as a result of an inaccurate genetic test result.

Participants were presented with one of two scenarios to base their decisions within. The first was an aggressive, fast-acting, but curable cancer. The second was a slow-acting, incurable cancer. The descriptions were similar to different lymphoma types, although the scenarios were not linked to any specific cancers, to increase generalisability of the results. Participants were then asked to select the characteristics of a genetic test they would choose in a given scenario. They could also choose not to have a test. The patient group were only ever shown the first scenario (curable cancer).

The authors used Sawtooth software to produce their DCE using a fractional factorial design. Choice sets were created to maximise D-efficiency and analysed with a conditional logit model. The survey was administered online. Response rates were 65% and 69% in the two general public groups, and 64% among the patient group.

A total of 1096 participants were recruited. Mean ages in the general public groups were similar (48.2 and 47.6) and the patient group was slightly older on average (mean of 58.2). Groups were similar in gender balance. Participant ancestries were not reported.

Willingness-to-pay for genetic testing was similar between both scenarios presented. Among the general public, preferences for levels were also consistent, although participants presented with scenario B had stronger preferences for higher test sensitivity and specificity. Differences were more pronounced when comparing a sample of the general public to patients. These groups received the same scenario, but patients had strong preferences for high test sensitivity compared to the general public ($p < 0.001$). In patients, a test with 50% sensitivity was associated with a strong negative preference. There were also differences in preferences for the genetic test procedure, with patients preferring bone marrow biopsies to liver biopsies. The general public found both tests equally unfavourable.

This study demonstrates an elegant method for comparing the preferences of the general public and patients. The differences in preferences present several interesting hypotheses for future qualitative work with each of these groups. A limitation of this study was the limited sample size, particularly in patients. A larger sample size would have enabled comparison of both scenarios across both patients and the general public. This was also one of the more complex DCEs located, in terms of the information presented to participants. With coefficients all in the expected direction, it appears that participants adequately understood the concepts of sensitivity and specificity. While this needs to be verified with further qualitative work, the knowledge that participants can tolerate this level of information is useful knowledge for my own DCE.

5.3.7.3 Payne et al 2011

This DCE compared preferences for pharmacogenetic testing in both patients and healthcare professionals (79). This DCE took place within the previously discussed TARGET trial (105). The aim of this DCE was to compare preferences for the characteristics of a pharmacogenetic test to predict the risk of an ADR from azathioprine treatment. Qualitative work consisted of expert opinion, focus groups and interviews, and pilot testing with patients (n=25) and HCPs (n=17). The results of this qualitative work were published in 2007 (106). Patients broadly understood the concepts involved in pharmacogenetics and were generally supportive. Concerns about tests revealing susceptibility to other diseases, and anxieties about test results, were also discussed. Both groups also agreed that genetic test results should be delivered by someone able to offer a high level of explanation (“I would expect someone to give the results back to me who has totally understood the

results. I don't care if it's the doctor or the nurse, but someone who doesn't, and somebody who can't answer my questions, I don't want the results from them because it's a waste of my time and a waste of their time.")

A common list of attributes was developed based on this qualitative work, so patients and healthcare professionals received the same questionnaire (Table 5.14). In total, 159 patients and 138 healthcare professionals were recruited.

Attribute	**Levels**			
Information given to patient about the test	None	Low	Moderate	High
Predictive accuracy	50%	60%	85%	90%
How the sample is collected	Blood test	Mouthwash	Finger prick	Mouth swab
Turnaround time of test	2 days	7 days	14 days	28 days
Who explains the results to the patient	GP	Pharmacist	Hospital doctor	Nurse

Table 5.14 - Attributes and levels of Payne et al 2011 (79)

The survey was constructed with a fractional factorial design. The software used for this was not specified. The method used for creating choice sets was selected from a book (Street and Burgess, 2007) (107), but the exact method was not specified. Data was analysed using a random effects probit model. The survey was administered through the post, with response rates of 50% in patients and 34% in healthcare professionals.

The mean age of patients was 45.8, and 56% were female. Demographic information was not reported for the healthcare professionals. Both groups preferred tests with high predictive accuracy and short turnaround times, and preferred a hospital doctor to explain the results of the test. Both groups were also willing to trade turnaround times for improved predictive accuracy. However, only patients were willing to trade longer turnaround times for higher levels of information (19.3 days compared to 8.9 days for healthcare professionals). An interesting result is that

patients had higher preferences for receiving no information than low levels of information.

This DCE presents an interesting perspective on how patient and healthcare professionals' views can differ. It is clearly important to consider multiple stakeholders in pharmacogenetic decision making. The importance of test turnaround time has implications for service delivery. This DCE did not include a cost attribute, instead using turnaround time as a continuous variable for trading. This was due to objections by the TARGET trial ethics committee, and does better reflect the UK NHS free health care model. However, omitting a cost attribute may limit the use of these results in contexts outside of the UK. Conversely, including a cost attribute could reduce the generalisability of the results, since UK participants would be unfamiliar with healthcare costings. This shows that the inclusion of a cost attribute is not a straightforward decision, and there are methodological and ethical issues to consider. This is something explored in the ethics application for this study and was not met with any objections by the University of Liverpool committees.

5.3.7.4 Powell et al 2015

The final paper in this review was a UK study including two DCEs, one in clinicians and one in epilepsy patients. The aim of the DCE was to compare the preferences of each group for *HLA-A*31:01* testing prior to the prescription of carbamazepine. This gene is a known risk factor for carbamazepine-induced ADRs, including Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) (108-111). Unlike in the previous work, qualitative work prior to this DCE found no common list of attributes that could be used for both patient and clinician DCEs. Distinct attributes and levels were therefore developed separately (Table 5.15). This qualitative work consisted of semi-structured interviews with clinicians (n=8) and patients living with epilepsy (n=56). Further work to develop the clinicians' DCE was done with attributes chosen from a previous DCE (79), and interviews with 12 neurologists.

Attributes	**Levels**		
Clinician's DCE			
Cost of test (£)	35	100	200
Time to result (days)	2	4	7
PPV	2	25	70

NPV	70	85	99
Coverage of test	Severe ADRs only	Severe and mild ADRs	
BNF	Test not included in BNF	Test is included in BNF	
Patient's DCE			
Probability seizures stop	5 in 10	3 in 10	1 in 10
Probability of fewer seizures	3 in 10	1 in 10	
Probability of mild skin rash	1 in 100	26 in 100	
Probability of memory problems	1 in 100	7 in 100	
Probability of potentially life-threatening reaction	Rare: more than 1 in 10,000	Uncommon: more than 1 in 1000	

Table 5.15 - Attributes and levels of Powell et al 2015 (80). BNF = British National Formulary. NPV = negative predictive value. PPV = positive predictive value

Clinicians were asked to imagine a scenario where they had decided to prescribe carbamazepine to an epileptic patient. They were given the attributes and levels describing different genetic tests and asked if they would order the test before prescribing carbamazepine, or proceed with prescribing carbamazepine blindly. Patients were given the choice between two epilepsy medications, representing carbamazepine (with a higher risk of SJS) and an alternative.

A fractional factorial design was generated. The authors did not specify the software or method used to do this. For the clinician questionnaire, the choice sets were generated by pairing with a constant comparator – in this case, not ordering the test. The method of generating choice sets for the patient questionnaire was not specified. Data was analysed using a random effects logit model. Surveys were administered online and response rates were not reported.

The 83 clinicians in the questionnaire were mainly adult neurologists, and most (83%) had prescribed carbamazepine within the past month. Most (80%) stated that they had 'no/superficial awareness' of pharmacogenetic testing. The attribute that had the largest impact on whether clinicians chose a pharmacogenetic test or not was if the test was included in the British National Formulary (BNF). A tests'

inclusion increased the odds a participant would choose testing by 58%. Cost also had an impact on decision making. The odds of choosing to test decreased by 1% for every £1 increase in the cost of the test. The uptake rate for testing in the base case scenario was estimated at 49.9%.

The 82 epilepsy patients recruited were 66% female and 90.2% white, with a median age of 38. A third had previously taken carbamazepine, and one had experienced a severe skin reaction. In general, patients were willing to accept a reduction in medication efficacy in exchange for a reduced risk of side effects. This was consistent across all the side-effects in the survey (mild skin rash, memory problems, SJS/TEN), with the greatest reduction in efficacy tolerated for a reduced risk of memory problems. The uptake rate in a real-world scenario was estimated at 61%. Coefficients from this DCE were used in a utility model to compare utility across different scenarios. For example, where the cost of the test was £100, the probability of uptake was 49.9%. If this cost was reduced to £35, the probability of uptake increased to 68.1%. This sort of analysis provides an excellent DCE output that is easy for those without specific DCE knowledge to understand. This is therefore a potentially very useful policy output. I have therefore chosen to use this same strategy for my DCE analysis (Chapter 7).

This DCE is unique in comparing patients to healthcare professionals. That common attributes for both populations could not be decided upon shows the importance of qualitative work in all populations to be included in the final DCE.

5.4 Discussion

This review shows that DCEs are actively being used in pharmacogenetics and has provided valuable information about previous DCEs that I am able to use in my own DCE. Something that has become clear is the large gaps in the literature. With a large percentage of identified papers focussing on breast cancer, there is a clear need for experiments in other disease areas. Another takeaway is that where participants are given the option to choose ‘no test’, they choose this at a rate of around 10%, showing that a majority of people are open to genetic testing to prevent ADRs.

The majority of papers located came from the UK and USA. Only one paper was an international collaboration (59). This corresponds to the results of a recent published systematic review of DCEs in genetic testing (6). Papers focussing on one country

and its health system are useful in their context, but may struggle for wider generalisability in other populations. There is also a clear lack of research in Africa, South America, and Oceania. Some of this is explained by a relative lack of pharmacogenetic research in these populations, which is an ongoing and well-documented issue (112, 113).

Most DCEs were in patients (those suffering from the illness being investigated). A minority were in HCPs, but this included some papers that provided interesting comparisons between patients and HCPs (79, 80). These approaches offer new perspectives that are relevant to policy makers in their respective areas. This quantitative measurement of patient preferences is a particular strength of the DCE method, and should be used to influence characteristics of pharmacogenetic services. Results can also be used to estimate the uptake of these tests. Surveys in the general public also have their place in this, and this population can be viewed as 'potential patients' for future usage of pharmacogenetic testing.

There was a broad range of sample sizes in the included DCEs, with a mean of 440 (range 67 – 1096). Few papers included details of sample size calculations. This suggests that many are utilising the 'rule of thumb' for DCE sample size estimation (114), or indeed no calculation at all (see Chapter 6).

A recent systematic review examined the use of DCEs in genetic testing (6), a broader application than my focus on genetic testing and ADRs. These authors located 38 papers, of which 36 were DCEs (the remainder were conjoint analysis). Nine of my included papers were in this analysis (70, 72, 74-80). The reasons for the exclusion of the remaining 4 papers (59, 67, 71, 73) are not clear, since the authors of this review did not provide details of their inclusion criteria.

Several of my findings correlate with theirs. They found a dominance of European and USA/Canada analyses, and similar proportions of their papers included patients, HCPs, and the general public. The most common number of attributes of papers included in their review was 5 (range 3 – 12), compared to my 4 (3 – 7). Their work found that cost was the most common type of included attribute. It was not as common in my located papers, although this may not be significant due to the smaller number of included papers.

I have also learned several things that I can apply to my own DCE. The importance of thorough qualitative work is unmistakeable. There are also lessons about DCE design and analysis that can be utilised. Finally, there are several practical issues with delivering the experiment that can be addressed.

5.4.1 The importance of qualitative work

Qualitative research in DCE development is under-used (16) and is an essential part of planning a DCE (23, 28, 40). Researchers and clinicians will rarely understand attributes in the same way that patients and the public do (23). A 2014 systematic review of DCEs found that reports of qualitative work were decreasing, a trend the authors called 'worrying' (30).

Most studies did report their qualitative research, to varying degrees of detail. Only one did not provide any detail (71). The most frequent method used was expert opinion, and this was almost always combined with other methods (such as focus groups and individual interviews). Five DCEs reported conducting pilot studies before launching a full DCE. Three papers published separate papers with full details and results of qualitative work (72, 76, 79). This should be prioritised when restrictions are applied to the word count or level of detail in the published DCE.

Many insights can be gained from reading these published papers. In focus groups, themes of data security and privacy concerns around genetic testing were common. Level of evidence was also considered. In one discussion, it was assumed that if a test was available with insurance coverage, it would be valid and have high levels of evidence behind it (70). This assumption (translated to a regulatory approval for UK participants) provides a useful potential shorthand for explaining high and low levels of evidence to participants, without additionally having to explain concepts of RCTs, observational studies, and statistical significance.

As shown by Powell et al 2015 (80), it is essential to perform qualitative work in all populations that the final DCE will be tested in. In this case, the authors planned to conduct a DCE in both healthcare professionals and patients. However, qualitative work with these groups made it clear that there was little overlap between relevant attributes between the groups. The authors then made the decision to produce separate DCEs for each group. While this makes it harder to directly compare across groups, it does make the results from each DCE more applicable and relevant within groups. It also limits the burden on patients since concepts familiar to healthcare professionals (such as PPV, NPV) did not need to be explained.

In this review, many useful potential attributes and levels were located. For my initial qualitative work when designing my own DCE, I decided to incorporate many of these and get feedback from healthcare professionals. The previous studies also informed my other qualitative work, including the choice of focus groups, expert

opinions, and consultation with patients. The methodology and results from my qualitative work are detailed in Chapter 6.

5.4.2 DCE design and analysis

There are many methods for designing DCEs. The DCEs included in this review mostly used D-efficient fractional factorial designs. These have been previously found to be the most common method used for constructing DCEs (4).

None of the located DCEs used full factorial designs. Though these designs are often seen as impractical, Lancsar & Louviere (2008) pointed out that grouping questions into blocks and randomising participants to a block can make these designs feasible in some situations (16).

Around half of my DCEs included an opt-out or 'no test' option. Including an option to reject both choices is important for the accuracy of conclusions drawn from a DCE [18, 115], however this can be difficult in a healthcare setting (116). This has also been used to estimate uptake rates of screening [53]. When no 'opt-out' or 'no-choice' option is provided, but this would be a realistic choice in real-life, this is a 'forced-choice' design (14). Harrison, *et al.* (2014) recommended that opt-out options should be used if they reflect real-life situations (64), and others have suggested their use if the aim of the DCE is to derive welfare measures (16). The inclusion of an 'opt-out' or 'no-choice' option can reduce design efficiency, but this is offset by improved generalisability and real-world applicability of the results (16). Using a 'no test' option is important if this reflects the reality of the scenario being modelled (16). However, it needs to be carefully phrased and fully reflect that choosing none of the choices means a zero value for all attributes (117). Differences in wording can significantly alter respondent preferences for the 'no test' option, and more concerningly, the other options in a DCE (117). A recent systematic review found that 28.2% of DCEs in genetic testing included an 'opt-out' option (6), a lower rate than my located papers. The authors of this paper recommend the use of opt-out alternatives since genetic testing is normally optional. Not including an opt-out can therefore overestimate the demand for genetic testing (6).

There is a delicate balance to be struck between cognitive burden, efficient design, and market realism when designing a DCE (42). There was a median of 12 choice tasks per participant. The number of choice sets impacts on respondents choices

(118) . It is important to balance between optimising the statistical efficiency and not providing respondents with too much cognitive burden (25, 118). As cognitive burden increases, participants are more likely to make choices at random (25, 54). However, increasing the number of choice tasks increases D-efficiency (51).

Only one DCE looked at external validity (59). This is a powerful tool for evaluating DCE performance but is only possible in some scenarios. The example here recruited women with high-risk genotypes and a history of interventions to reduce breast cancer risk, then asked them to choose an intervention for reducing risk. They were then able to compare this to their previous choices and found that preferences differed to actual choices.

I will use a fractional factorial design for my DCE, and maximise D-efficiency. These were common choices in the included papers and are well-validated methods recommended by several authors and DCE guides (10, 25, 26, 56, 119).

I will also include an 'opt-out' option since this reflects real practice, where a patient would be able to opt-out of genetic testing. The assessment of external validity would be a good addition to this project, with funding for longer term follow-up of participants. A group of participants could be followed and their uptake of genetic testing in the future could be linked with their DCE responses. This is outside the scope of this PhD but would be an interesting challenge.

The analysis of a DCE allows for several potential outputs, such as WTP, MAR, and uptake. Examples of each of these were located in this review. I chose to base the analysis on that conducted by Powell, *et al.* (2015) (80). Coefficients obtained in this DCE were used to calculate the probability of test uptake in different scenarios. This output is policy-relevant and easy to understand. It also allows comparison across different potential genetic tests.

5.4.3 Practical issues

The majority of studies recruited participants that were older, white, and female. This is likely due to the high proportion of papers that focussed on breast cancer. The average age of participants, where reported, was 48.0. Considering that older people may be more likely to develop disease in general, this may be representative of the wider patient population. However, this does indicate a gap in the research for the preferences of younger populations. Genetic testing is an intervention relevant for more age groups, so I will aim to recruit a sample representative of the UK population.

Not reporting response rates is “unacceptable” (40) although this may not be reportable in some studies (e.g. when approaching people at random as in Severin, *et al.* (2013) (120) did at a conference). It is also harder to define ‘response rate’ if a survey is online with an open link. Around half of the studies did not report response rates. I will aim to report them if the online platform allows the recording of the number of participants that begin the study but do not complete it.

Online DCEs were the norm, continuing the trend observed by Soekhai, *et al.* (2019) from a series of 4 analyses (60). Only one survey was delivered by post (79). Post is historically a poor choice as it suffers from low response rates (121, 122), and the risk that participants may not fully understand the exercise (25). It may therefore be more suitable for surveys of healthcare professionals. Web-based designs are also hypothesised to be better for recruitment for longer DCE designs, as respondents cannot see the length of the questionnaire (118). A more recent analysis found that the device used by participants to access the questionnaire online does not affect outcomes, providing further evidence for the implementation of surveys online (123). Online administration was the most common method in a systematic review of DCEs in genetic testing (6). I am planning to conduct my DCE online for ease of recruitment and design.

Communication of risk to the general public can be difficult (124-126). It is important to consider how different people may interpret the same risks. Numeracy of participants is also a factor to consider (124). Visual aids can help in overcoming these issues (126). One paper in the systematic review successfully used pictograms to represent risks (77). As there are similar levels of risk with the ADRs in my study, I will use pictograms to communicate that risk where practical.

A final consideration is that participant preferences can change depending on their previous experience of disease or ADRs. This was seen in the DCE published by Ballinger, *et al.* (2017) (71). Participants without previous experience of the ADRs were more concerned about avoiding future ADRs, while those who had previously suffered them were more willing to risk suffering them again. With this in mind, I have incorporated an optional question into my DCE to ask if participants have previously suffered from the disease discussed in the scenario. I will also ask participants if they have previously had a genetic test.

5.4.4 Limitations

This systematic review has some limitations. The scope of this review is quite narrow, focussing on DCEs in pharmacogenetics that included ADRs in their

analysis. While this makes the results very useful for my own DCE, it limits the generalisability of the findings to wider pharmacogenetics. A systematic review focussing on broader genetic testing applications published in 2021 has several similarities with my results (6).

The 2013 International Society for Pharmacoeconomics and Outcomes Research (ISPOR) checklist provided a well-validated set of guidelines for the design and analysis of DCEs (31). Ideally, all DCEs should follow these guidelines. With further time and resources, I may have considered evaluating these DCEs for compliance with this checklist. Studies were also not assessed for quality or risk of bias. This is normally an important step when conducting a systematic review. It was not included since my systematic review is not of an intervention. This assessment is also not within the scope of the ISPOR checklist. There is the possibility that bias in this review will affect my DCE, but any bias is likely to be corrected by further qualitative work (Chapter 6).

DCEs also have more general limitations. The method can be difficult to apply. Use of the DCE method without a full understanding of the theory, methods and interpretation (16) can lead to false assumptions of preferences and has the potential to negatively affect health policy. DCEs can also be time-consuming to design and interpret, and their generalisability to wider policy has been questioned (16). External validity, or the correlation between what participants say they will do with what they actually do, is difficult to measure. A study by Lambooij, *et al.* (2015) investigated the link between people's plans to be vaccinated, and their actual vaccination behaviour (127). Although DCE responses could predict behaviour correctly 80% of the time, this means that DCE results may not accurately predict behaviour in 1 in 5 people.

WTP in a DCE also differs from real-life behaviour. A DCE asking people's WTP for an asthma intervention found that 38% of people would purchase the intervention at any price. However, given an opportunity to purchase the intervention, only 12% of people in a comparable group actually did (128).

5.5 Conclusion

The use of DCEs within healthcare and pharmacogenetics is likely to increase in the future (16). It is therefore essential that the theory and methods of this technique are fully understood. This powerful tool for assessing patient preferences is particularly useful for examining WTP (in money, or other measures), estimating uptake rates and determining the ideal characteristics of a service (14).

I completed this systematic review prior to my DCE, as I recommended when planning future RCTs in Chapter 3. I have learned several important points that will inform my own DCE. The importance of extensive qualitative work is made clear. Full details of qualitative work will be presented in Chapter 6. Examining DCE design led me to consider the use of an 'opt-out' option in my own DCE, to more accurately reflect real world scenarios. I have also learned about the practicalities of conducting DCEs, including the advantages of recruiting online, and the importance of a diverse pool of participants (where appropriate).

I have directly used some of the attributes identified here as a starting point for my qualitative work. This is discussed in full in Chapter 6.

5.6 References

1. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med. 2009;11(1):3-14.
2. Need AC, Motulsky AG, Goldstein DB. Priorities and standards in pharmacogenetic research. Nat Genet. 2005;37(7):671-81.
3. Weinshilboum R, Wang L. Pharmacogenomics: bench to bedside. Nature Reviews Drug Discovery. 2004;3(9):739-48.
4. Bridges JFP, Kinter ET, Kidane L, Heinzen RR, McCormick C. Things are Looking up Since We Started Listening to Patients. The Patient: Patient-Centered Outcomes Research. 2008;1(4):273-82.
5. Bridges JF. Lean systems approaches to health technology assessment: a patient-focused alternative to cost-effectiveness analysis. Pharmacoeconomics. 2006;24 Suppl 2:101-9.
6. Ozdemir S, Lee JJ, Chaudhry I, Ocampo RRQ. A Systematic Review of Discrete Choice Experiments and Conjoint Analysis on Genetic Testing. Patient. 2022;15(1):39-54.
7. Food and Drug Administration (US). Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling: Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders 2016 [Available from: https://www.fda.gov/media/92593/download.
8. de Bekker-Grob EW, Hol L, Donkers B, van Dam L, Habbema JD, van Leerdam ME, et al. Labeled versus unlabeled discrete choice experiments in health economics: an application to colorectal cancer screening. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2010;13(2):315-23.
9. Mandeville KL, Lagarde M, Hanson K. The use of discrete choice experiments to inform health workforce policy: a systematic review. BMC health services research. 2014;14(1):367.

10. Ryan M, Kolstad JR, Rockers PC, Dolea C. How to conduct a discrete choice experiment for health workforce recruitment and retention in remote and rural areas: a user guide with case studies. The World Bank; 2012.
11. Watson V, Luchini S, Regier D, Schulz R. Chapter 4 - Monetary analysis of health outcomes. In: Guerriero C, editor. Cost-Benefit Analysis of Environmental Health Interventions: Academic Press; 2020. p. 73-93.
12. de Corte K, Cairns J, Grieve R. Stated versus revealed preferences: An approach to reduce bias. Health Economics. 2021;30(5):1095-123.
13. Ryan M. Discrete choice experiments in health care. BMJ. 2004;328(7436):360-1.
14. Lancsar E, Fiebig DG, Hole AR. Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software. Pharmacoeconomics. 2017;35(7):697-716.
15. Bryan S, Dolan P. Discrete choice experiments in health economics. For better or for worse? Eur J Health Econ. 2004;5(3):199-202.
16. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics. 2008;26(8):661-77.
17. Holmes EAF, Plumpton C, Baker GA, Jacoby A, Ring A, Williamson P, et al. Patient-Focused Drug Development Methods for Benefit-Risk Assessments: A Case Study Using a Discrete Choice Experiment for Antiepileptic Drugs. Clinical pharmacology and therapeutics. 2019;105(3):672-83.
18. Linley WG, Hughes DA. Decision-makers' preferences for approving new medicines in Wales: a discrete-choice experiment with assessment of external validity. Pharmacoeconomics. 2013;31(4):345-55.
19. Thurstone LL. A law of comparative judgment. Psychological Review. 1927;34(4):273-86.
20. Lancaster KJ. A New Approach to Consumer Theory. Journal of Political Economy. 1966;74(2):132-57.
21. Louviere JJ, Flynn TN, Carson RT. Discrete Choice Experiments Are Not Conjoint Analysis. Journal of Choice Modelling. 2010;3(3):57-72.
22. McFadden D. The Choice Theory Approach to Market Research. Marketing Science. 1986;5(4):275-97.
23. Hall J, Viney R, Haas M, Louviere J. Using stated preference discrete choice modeling to evaluate health care programs. Journal of Business Research. 2004;57(9):1026-32.
24. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ. 2006;15(8):797-811.
25. Ryan M, Gerard K, Amaya-Amaya M. Discrete Choice Experiments in a Nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care. The Economics of Non-Market Goods and Resources. Dordrecht: Springer Netherlands; 2008. p. 13-46.
26. Street DJ, Burgess L, Viney R, Louviere J. Designing Discrete Choice Experiments for Health Care. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care. The Economics of Non-Market Goods and Resources. Dordrecht: Springer Netherlands; 2008. p. 47-72.
27. Coast J, Al-Janabi H, Sutton EJ, Horrocks SA, Vosper AJ, Swancutt DR, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. Health Econ. 2012;21(6):730-41.
28. Vass C, Rigby D, Payne K. The Role of Qualitative Research Methods in Discrete Choice Experiments. Medical decision making : an international journal of the Society for Medical Decision Making. 2017;37(3):298-313.

29. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2011;14(4):403-13.
30. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883-902.
31. Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Muhlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013;16(1):3-13.
32. Hauber AB, Gonzalez JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2016;19(4):300-15.
33. Harris CA, Shauver MJ, Yuan F, Nasser J, Chung KC. Understanding Patient Preferences in Proximal Interphalangeal Joint Surgery for Osteoarthritis: A Conjoint Analysis. J Hand Surg Am. 2018;43(7):615-24 e4.
34. Louviere JJ, Hensher D. DESIGN AND ANALYSIS OF SIMULATED CHOICE OR ALLOCATION EXPERIMENTS IN TRAVEL CHOICE MODELING. Transportation Research Record. 1982:11-7.
35. McIntosh E. Applied methods of cost-benefit analysis in health care: Oxford University Press; 2010.
36. Griffith GL, Edwards RT, Williams JM, Gray J, Morrison V, Wilkinson C, et al. Patient preferences and National Health Service costs: a cost-consequences analysis of cancer genetic services. Fam Cancer. 2009;8(4):265-75.
37. Ryan M. Using Consumer Preferences in Health Care Decision Making: The Application of Conjoint Analysis1996.
38. Marsh K, van Til JA, Molsen-David E, Juhnke C, Hawken N, Oehrlein EM, et al. Health Preference Research in Europe: A Review of Its Use in Marketing Authorization, Reimbursement, and Pricing Decisions-Report of the ISPOR Stated Preference Research Special Interest Group. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2020;23(7):831-41.
39. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21(2):145-72.
40. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Applied health economics and health policy. 2003;2(1):55-64.
41. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? Expert Rev Pharmacoecon Outcomes Res. 2017;17(6):531-42.
42. Herbild L. Designing a DCE to outlay patients' and the publics' preferences for genetic screening in the treatment of depression. 2007.
43. Walley T, Haycox A, Boland A. Pharmacoeconomics: Elsevier Health Sciences; 2004.
44. Pirmohamed M, Hughes DA. Pharmacogenetic tests: the need for a level playing field. Nature reviews Drug discovery. 2013;12(1):3-4.
45. Altar CA, Amakye D, Bounos D, Bloom J, Clack G, Dean R, et al. A prototypical process for creating evidentiary standards for biomarkers and diagnostics. Clinical pharmacology and therapeutics. 2008;83(2):368-71.

46. Food and Drug Administration (US). Patient Preference Information—Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labelling. Guidance for industry, food and Drug Administration Staff, and Other Stakeholders. 2016 [Available from: https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM446680.pdf.
47. Osanlou O, Pirmohamed M, Daly AK. Chapter Seven - Pharmacogenetics of Adverse Drug Reactions. In: Brøsen K, Damkier P, editors. Advances in Pharmacology. 83: Academic Press; 2018. p. 155-90.
48. Udagawa C, Zembutsu H. Pharmacogenetics for severe adverse drug reactions induced by molecular-targeted therapy. Cancer Sci. 2020;111(10):3445-57.
49. Daud ANA, Bergsma EL, Bergman JEH, De Walle HEK, Kerstjens-Frederikse WS, Bijker BJ, et al. Knowledge and attitude regarding pharmacogenetics among formerly pregnant women in the Netherlands and their interest in pharmacogenetic research. BMC Pregnancy Childbirth. 2017;17(1):120.
50. Rigter T, Jansen ME, de Groot JM, Janssen SWJ, Rodenburg W, Cornel MC. Implementation of Pharmacogenetics in Primary Care: A Multi-Stakeholder Perspective. Front Genet. 2020;11:10.
51. Vanniyasingam T, Daly C, Jin X, Zhang Y, Foster G, Cunningham C, et al. Investigating the impact of design characteristics on statistical efficiency within discrete choice experiments: A systematic survey. Contemp Clin Trials Commun. 2018;10:17-28.
52. Walker JL, Wang Y, Thorhauge M, Ben-Akiva M. D-efficient or deficient? A robustness analysis of stated choice experimental designs. Theory and Decision. 2017;84(2):215-38.
53. Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA. Measuring patient preferences for colorectal cancer screening using a choice-format survey. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2007;10(5):415-30.
54. Louviere JJ, Hensher DA, Swait JD, Adamowicz W. Stated Choice Methods : Analysis and Applications. Cambridge, UNITED KINGDOM: Cambridge University Press; 2000.
55. Abdullah S, Markandya A, Nunes PALD. Introduction to Economic Valuation Methods. Research Tools in Natural Resource and Environmental Economics2011. p. 143-87.
56. Louviere JJ. What You Don't Know Might Hurt You: Some Unresolved Issues in the Design and Analysis of Discrete Choice Experiments. Environmental and Resource Economics. 2006;34(1):173-88.
57. Zwerina K, Huber J, F. Kuhfeld W. A General Method for Constructing Efficient Choice Designs2000.
58. Kuhfeld WF, Tobias RD, Garratt M. Efficient Experimental Design with Marketing Research Applications. Journal of Marketing Research. 1994;31(4):545-57.
59. Liede A, Mansfield CA, Metcalfe KA, Price MA, Kathleen Cuningham Foundation Consortium for Research into Familial Breast C, Snyder C, et al. Preferences for breast cancer risk reduction among BRCA1/BRCA2 mutation carriers: a discrete-choice experiment. Breast Cancer Res Treat. 2017;165(2):433-44.
60. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. Pharmacoeconomics. 2019;37(2):201-26.
61. Guerra RL, Castaneda L, de Albuquerque RCR, Ferreira CBT, Correa FM, Fernandes RRA, et al. Patient Preferences for Breast Cancer Treatment

Interventions: A Systematic Review of Discrete Choice Experiments. Patient. 2019;12(6):559-69.
62. Trapero-Bertran M, Rodriguez-Martin B, Lopez-Bastida J. What attributes should be included in a discrete choice experiment related to health technologies? A systematic literature review. PLoS One. 2019;14(7):e0219905.
63. Bien DR, Danner M, Vennedey V, Civello D, Evers SM, Hiligsmann M. Patients' Preferences for Outcome, Process and Cost Attributes in Cancer Treatment: A Systematic Review of Discrete Choice Experiments. Patient. 2017;10(5):553-65.
64. Harrison M, Rigby D, Vass C, Flynn T, Louviere J, Payne K. Risk as an attribute in discrete choice experiments: a systematic review of the literature. Patient. 2014;7(2):151-70.
65. Harrison M, Milbers K, Hudson M, Bansback N. Do patients and health care providers have discordant preferences about which aspects of treatments matter most? Evidence from a systematic review of discrete choice experiments. BMJ open. 2017;7(5):e014719.
66. Pokorska-Bocci A, Stewart A, Sagoo GS, Hall A, Kroese M, Burton H. 'Personalized medicine': what's in a name? Per Med. 2014;11(2):197-210.
67. Boeri M, McMichael AJ, Kane JPM, O'Neill FA, Kee F. Physician-Specific Maximum Acceptable Risk in Personalized Medicine: Implications for Medical Decision Making. Medical decision making : an international journal of the Society for Medical Decision Making. 2018;38(5):593-600.
68. R Studio Team. RStudio: Integrated Development for R. RStudio, Inc. Boston, MA2016.
69. MacDonald KV, Bombard Y, Deal K, Trudeau M, Leighl N, Marshall DA. The influence of gene expression profiling on decisional conflict in decision making for early-stage breast cancer chemotherapy. European journal of cancer (Oxford, England : 1990). 2016;61:85-93.
70. Marshall DA, Deal K, Bombard Y, Leighl N, MacDonald KV, Trudeau M. How do women trade-off benefits and risks in chemotherapy treatment decisions based on gene expression profiling for early-stage breast cancer? A discrete choice experiment. BMJ open. 2016;6(6):e010981.
71. Ballinger TJ, Kassem N, Shen F, Jiang G, Smith ML, Railey E, et al. Discerning the clinical relevance of biomarkers in early stage breast cancer. Breast Cancer Res Treat. 2017;164(1):89-97.
72. Issa AM, Tufail W, Atehortua N, McKeever J. A national study of breast and colorectal cancer patients' decision-making for novel personalized medicine genomic diagnostics. Per Med. 2013;10(3):245-56.
73. Smith ML, White CB, Railey E, Sledge GW, Jr. Examining and predicting drug preferences of patients with metastatic breast cancer: using conjoint analysis to examine attributes of paclitaxel and capecitabine. Breast Cancer Res Treat. 2014;145(1):83-9.
74. Chan SL, Wen Low JJ, Lim YW, Finkelstein E, Farooqui MA, Chia KS, et al. Willingness-to-pay and preferences for warfarin pharmacogenetic testing in Chinese warfarin patients and the Chinese general public. Per Med. 2013;10(2):127-37.
75. Dong D, Ozdemir S, Mong Bee Y, Toh SA, Bilger M, Finkelstein E. Measuring High-Risk Patients' Preferences for Pharmacogenetic Testing to Reduce Severe Adverse Drug Reaction: A Discrete Choice Experiment. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2016;19(6):767-75.
76. Herbild L, Bech M, Gyrd-Hansen D. Estimating the Danish populations' preferences for pharmacogenetic testing using a discrete choice experiment. The case of treating depression. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2009;12(4):560-7.

77. Marshall DA, Gonzalez JM, MacDonald KV, Johnson FR. Estimating Preferences for Complex Health Technologies: Lessons Learned and Implications for Personalized Medicine. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2017;20(1):32-9.
78. Najafzadeh M, Johnston KM, Peacock SJ, Connors JM, Marra MA, Lynd LD, et al. Genomic testing to determine drug response: measuring preferences of the public and patients using Discrete Choice Experiment (DCE). BMC health services research. 2013;13(1):454.
79. Payne K, Fargher EA, Roberts SA, Tricker K, Elliott RA, Ratcliffe J, et al. Valuing pharmacogenetic testing services: a comparison of patients' and health care professionals' preferences. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2011;14(1):121-34.
80. Powell G, Holmes EA, Plumpton CO, Ring A, Baker GA, Jacoby A, et al. Pharmacogenetic testing prior to carbamazepine treatment of epilepsy: patients' and physicians' preferences for testing and service delivery. British journal of clinical pharmacology. 2015;80(5):1149-59.
81. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6(7):e1000100.
82. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.
83. The TA Collaborative. PRISMA Flow Diagram Generator [Available from: http://prisma.thetacollaborative.ca/.
84. Najafzadeh M, Davis JC, Joshi P, Marra C. Barriers for integrating personalized medicine into clinical practice: a qualitative analysis. American journal of medical genetics Part A. 2013;161A(4):758-63.
85. Tassaneeyakul W, Jantararoungtong T, Chen P, Lin PY, Tiamkao S, Khunarkornsiri U, et al. Strong association between HLA-B*5801 and allopurinol-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in a Thai population. Pharmacogenetics and genomics. 2009;19(9):704-9.
86. Pacific Prime Singapore. How does Singapore's healthcare system work? 2020 [Available from: https://www.pacificprime.sg/blog/singapores-healthcare-system/.
87. Suppiah R, Dissanayake A, Dalbeth N. High prevalence of gout in patients with Type 2 diabetes: male sex, renal impairment, and diuretic use are major risk factors. The New Zealand medical journal. 2008;121(1283):43-50.
88. Carlson JJ, Roth JA. The impact of the Oncotype Dx breast cancer assay in clinical practice: a systematic review and meta-analysis. Breast Cancer Res Treat. 2013;141(1):13-22.
89. Issa AM, Hutchinson JF, Tufail W, Fletcher E, Ajike R, Tenorio J. Provision of personalized genomic diagnostic technologies for breast and colorectal cancer: an analysis of patient needs, expectations and priorities. Per Med. 2011;8(4):401-11.
90. Antoniou AC, Cunningham AP, Peto J, Evans DG, Lalloo F, Narod SA, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. British journal of cancer. 2008;98(8):1457-66.
91. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull. 1987;13(2):261-76.
92. McMichael AJ, Boeri M, Rolison JJ, Kane J, O'Neill FA, Scarpa R, et al. The Influence of Genotype Information on Psychiatrists' Treatment Recommendations: More Experienced Clinicians Know Better What to Ignore. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2017;20(1):126-31.

93. Price JS, Waller PC, Wood SM, MacKay AV. A comparison of the post-marketing safety of four selective serotonin re-uptake inhibitors including the investigation of symptoms occurring on withdrawal. British journal of clinical pharmacology. 1996;42(6):757-63.
94. Evans WE, McLeod HL. Pharmacogenomics--drug disposition, drug targets, and side effects. The New England journal of medicine. 2003;348(6):538-49.
95. Thomas KL, Ellingrod VL. Pharmacogenetics of selective serotonin reuptake inhibitors and associated adverse drug reactions. Pharmacotherapy. 2009;29(7):822-31.
96. Bijl MJ, Visser LE, Hofman A, Vulto AG, van Gelder T, Stricker BH, et al. Influence of the CYP2D6*4 polymorphism on dose, switching and discontinuation of antidepressants. British journal of clinical pharmacology. 2008;65(4):558-64.
97. Persing M, Grosse R. Current St. Gallen Recommendations on Primary Therapy of Early Breast Cancer*. Breast Care. 2007;2(3):137-40.
98. Bombard Y, Rozmovits L, Trudeau M, Leighl NB, Deal K, Marshall DA. Access to personalized medicine: factors influencing the use and value of gene expression profiling in breast cancer treatment. Current oncology (Toronto, Ont). 2014;21(3):e426-33.
99. Bombard Y, Rozmovits L, Trudeau M, Leighl NB, Deal K, Marshall DA. The value of personalizing medicine: medical oncologists' views on gene expression profiling in breast cancer treatment. The oncologist. 2015;20(4):351-6.
100. Bombard Y, Rozmovits L, Trudeau ME, Leighl NB, Deal K, Marshall DA. Patients' perceptions of gene expression profiling in breast cancer treatment decisions. Current oncology (Toronto, Ont). 2014;21(2):e203-11.
101. Pirmohamed M, Burnside G, Eriksson N, Jorgensen AL, Toh CH, Nicholson T, et al. A randomized trial of genotype-guided dosing of warfarin. The New England journal of medicine. 2013;369(24):2294-303.
102. Anderson JL, Horne BD, Stevens SM, Woller SC, Samuelson KM, Mansfield JW, et al. A randomized and clinical effectiveness trial comparing two pharmacogenetic algorithms and standard care for individualizing warfarin dosing (CoumaGen-II). Circulation. 2012;125(16):1997-2005.
103. Gage BF BA, Lin H, Woller SC, Stevens SM, Al-Hammadi N, Li J, Rodríguez T, Miller JP, McMillin GA, Pendleton RC. Effect of Genotype-Guided Warfarin Dosing on Clinical Events and Anticoagulation Control Among Patients Undergoing Hip or Knee Arthroplasty: The GIFT Randomized Clinical Trial. JAMA. 2017;318(12):1115-24.
104. Jorgensen AL, FitzGerald RJ, Oyee J, Pirmohamed M, Williamson PR. Influence of CYP2C9 and VKORC1 on patient response to warfarin: a systematic review and meta-analysis. PLoS One. 2012;7(8):e44064.
105. Newman WG, Payne K, Tricker K, Roberts SA, Fargher E, Pushpakom S, et al. A pragmatic randomized controlled trial of thiopurine methyltransferase genotyping prior to azathioprine treatment: the TARGET study. Pharmacogenomics. 2011;12(6):815-26.
106. Fargher EA, Eddy C, Newman W, Qasim F, Tricker K, Elliott RA, et al. Patients' and healthcare professionals' views on pharmacogenetic testing and its future delivery in the NHS. Pharmacogenomics. 2007;8(11):1511-9.
107. Street DJ, Burgess L. The construction of optimal stated choice experiments: Theory and methods: John Wiley & Sons; 2007.
108. Grover S, Kukreti R. HLA alleles and hypersensitivity to carbamazepine: an updated systematic review with meta-analysis. Pharmacogenetics and genomics. 2014;24(2):94-112.
109. McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperaviciute D, Carrington M, et al. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. The New England journal of medicine. 2011;364(12):1134-43.

110. Niihara H, Kakamu T, Fujita Y, Kaneko S, Morita E. HLA-A31 strongly associates with carbamazepine-induced adverse drug reactions but not with carbamazepine-induced lymphocyte proliferation in a Japanese population. J Dermatol. 2012;39(7):594-601.
111. Ozeki T, Mushiroda T, Yowang A, Takahashi A, Kubo M, Shirakata Y, et al. Genome-wide association study identifies HLA-A*3101 allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. Hum Mol Genet. 2011;20(5):1034-41.
112. Limaye N. Pharmacogenomics, Theranostics and Personalized Medicine - the complexities of clinical trials: challenges in the developing world. Appl Transl Genom. 2013;2:17-21.
113. Klein ME, Parvez MM, Shin JG. Clinical Implementation of Pharmacogenomics for Personalized Precision Medicine: Barriers and Solutions. J Pharm Sci. 2017;106(9):2368-79.
114. Orme B. Sample size issues for conjoint analysis studies. Sawthooth Software Research paper Series Squim, WA, USA: Sawthooth Software Inc. 1998.
115. Johnson FR, Backhouse M. Eliciting stated preferences for health-technology adoption criteria using paired comparisons and recommendation judgments. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2006;9(5):303-11.
116. McIntosh E. Using discrete choice experiments within a cost-benefit analysis framework: some considerations. Pharmacoeconomics. 2006;24(9):855-68.
117. Determann D, Gyrd-Hansen D, de Wit GA, de Bekker-Grob EW, Steyerberg EW, Lambooij MS, et al. Designing Unforced Choice Experiments to Inform Health Care Decision Making: Implications of Using Opt-Out, Neither, or Status Quo Alternatives in Discrete Choice Experiments. Medical decision making : an international journal of the Society for Medical Decision Making. 2019;39(6):681-92.
118. Bech M, Kjaer T, Lauridsen J. Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. Health Econ. 2011;20(3):273-86.
119. Ryan M, Gerard K, Amaya-Amaya M. Using discrete choice experiments to value health and health care: Springer Science & Business Media; 2007.
120. Severin F, Schmidtke J, Muhlbacher A, Rogowski WH. Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments. Eur J Hum Genet. 2013;21(11):1202-8.
121. Adams J, Bateman B, Becker F, Cresswell T, Flynn D, McNaughton R, et al. Effectiveness and acceptability of parental financial incentives and quasi-mandatory schemes for increasing uptake of vaccinations in preschool children: systematic review, qualitative study and discrete choice experiment. Health technology assessment (Winchester, England). 2015;19(94):1-176.
122. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf). 2005;27(3):281-91.
123. Vass CM, Boeri M. Mobilising the Next Generation of Stated-Preference Studies: the Association of Access Device with Choice Behaviour and Data Quality. Patient. 2021;14(1):55-63.
124. Waters EA, Fagerlin A, Zikmund-Fisher BJ. Overcoming the Many Pitfalls of Communicating Risk. In: Diefenbach MA, Miller-Halegoua S, Bowen DJ, editors. Handbook of Health Decision Science. New York, NY: Springer New York; 2016. p. 265-77.
125. Frewer L. The public and effective risk communication. Toxicol Lett. 2004;149(1-3):391-7.
126. Garcia-Retamero R, Cokely ET. Communicating Health Risks With Visual Aids. Current Directions in Psychological Science. 2013;22(5):392-9.
127. Lambooij MS, Harmsen IA, Veldwijk J, de Melker H, Mollema L, van Weert YW, et al. Consistency between stated and revealed preferences: a discrete choice

experiment and a behavioural experiment on vaccination behaviour compared. BMC medical research methodology. 2015;15(1):19.
128. Blumenschein K, Johannesson M, Yokoyama KK, Freeman PR. Hypothetical versus real willingness to pay in the health care sector: results from a field experiment. J Health Econ. 2001;20(3):441-57.

Chapter 6: DCE development

6.1 Introduction

Patient involvement in research is increasing (1-4), particularly within the National Health Service (NHS) (1). This involvement is important when moving forward with new technologies and services, such as genetic testing. The NHS Genomic Medicine Service, launched in January 2019, provides access to gene panel testing to eligible patients through their clinicians (5, 6). The use of this is expanding and is predicted to soon be a routine part of clinical practice (7). The potential cost of this service is a concern. However, the cost of adverse drug reactions (ADRs) to the NHS is also significant (8), with 92,114 emergency hospital admissions due to ADRs in 2014/2015 (9). The use of genetic testing to prevent ADRs has the potential to reduce these costs and improve patient care.

Another challenge in the implementation of this service is patient acceptance. For example, data security is one major concern (6, 10, 11). A successful Genomic Medicine Service must incorporate the preferences of patients and the general public for genetic testing.

Discrete choice experiments (DCEs) have been extensively used to quantify the stated preferences of individuals and stakeholders for goods and services (12-19). The method allows for the calculation of utility, uptake rates, willingness-to-pay, and other highly policy-relevant outcomes (20-22). However, alongside this power comes the importance of rigorous development and pre-testing. For a DCE to be relevant and its findings to be generalisable, qualitative work with the target population during development is essential (17, 19, 23-25).

Street, *et al.* (2008) defined six stages of designing a DCE, from identifying the problem to analysing the impact of results on policy (Figure 6.1) (26). After defining the problem the DCE will address, the attributes and levels need to be defined. There are several methods for choosing attributes, and published DCE studies vary in the level of detail provided on this process (24, 27). Louviere remarked in 2006 that “the level of knowledge and understanding of statistical design theory exhibited in most published DCE papers is very low and/or exhibits significant errors.”(28)

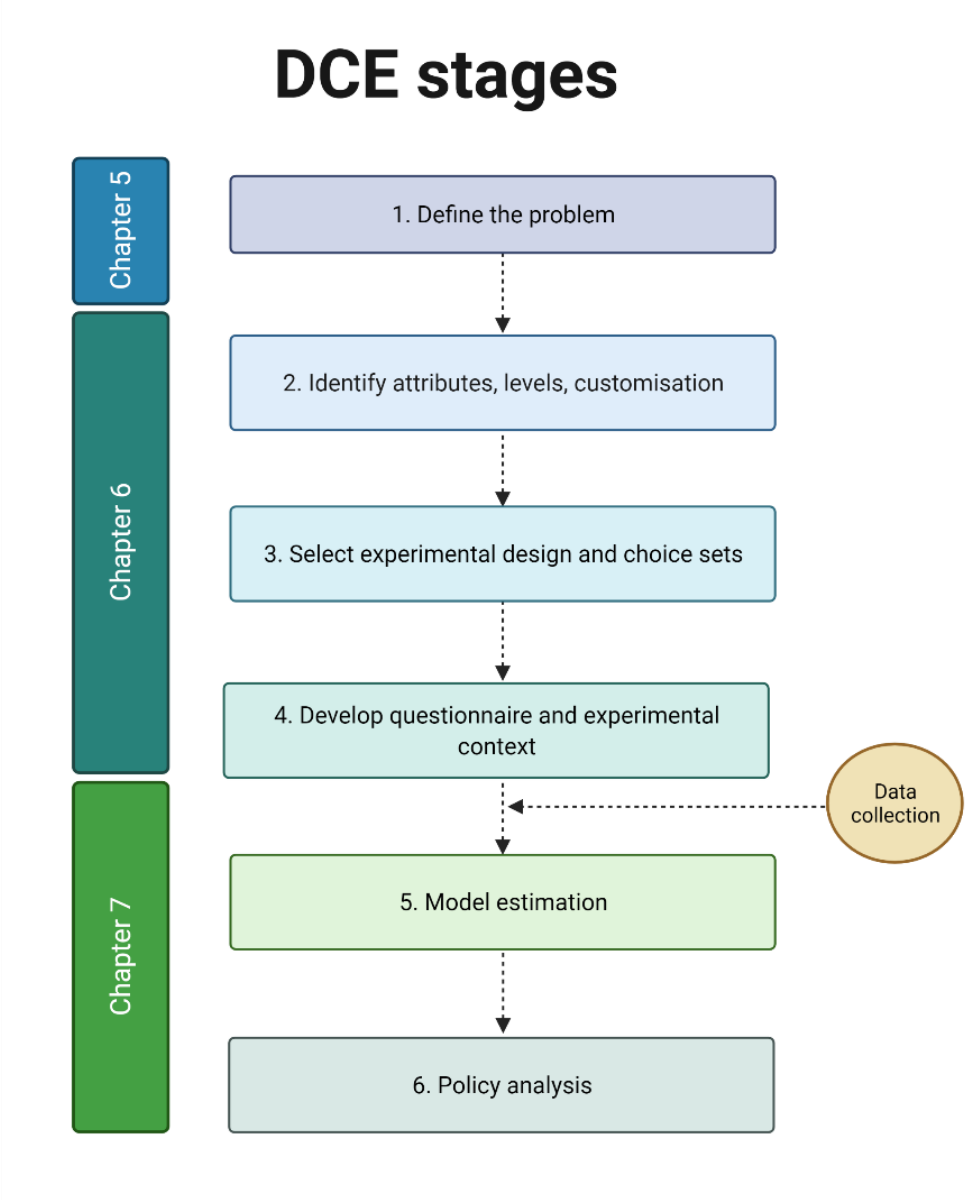


Figure 6.1 - DCE stages as defined by Street, et al. 2008 (26), with reference to the chapters of this thesis that correspond to each stage. Created using BioRender.com.

Attributes need to be:

- factors important to both patients and policy makers,
- have plausible levels that can be evaluated,
- be able to be traded (27).

There are many available methods for choosing attributes (24) but qualitative work to develop them is "highly recommended".(24, 27, 29) Street, *et al.* also provided a list of the minimum information that should be reported to show the development of attributes for a DCE. A further standard for DCE development was published by the

International Society for Pharmacoeconomics and Outcomes Research (ISPOR) in 2013 (19).

Qualitative research can be used in any stage of the DCE process to inform development. This may include attribute/level development, pre-testing, and debriefing (23, 30). Its use is increasingly advocated in order to improve the quality of choice experiments (30, 31). In my systematic review of previous pharmacogenetics DCEs (Chapter 5), I identified several methods used for qualitative research. These included surveys of healthcare professionals (HCPs) and patients, focus groups, and eliciting expert opinions.

I have chosen to design a DCE to quantify the preferences of the general public for genetic testing to prevent ADRs. In order to ensure the results of the DCE are generalisable and relevant, I have undertaken a programme of qualitative work that elicits the views of HCPs, patients, and the general public.

6.2 Methods

6.2.1 Overview

The importance of qualitative work has been explored in the previous chapter. I planned an extensive series of qualitative work in multiple groups to fully inform my DCE design (Figure 6.2).

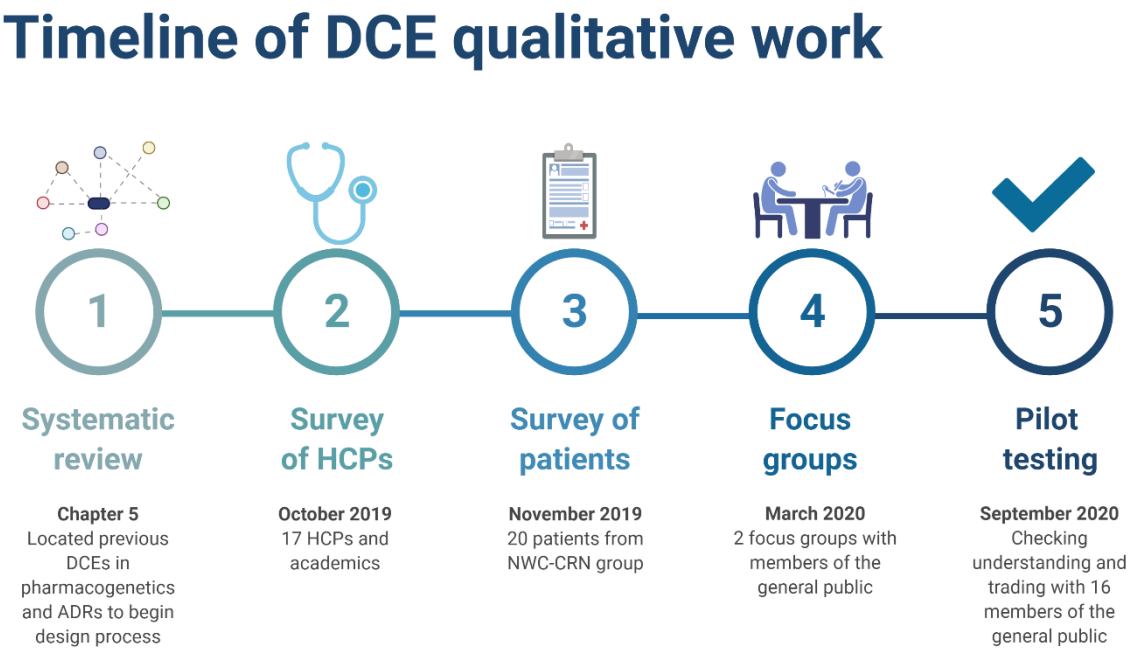


Figure 6.2 - Timeline of DCE qualitative work. HCPs= healthcare professionals. NWC-CRN = North West Coast Clinical Research Network. Created with BioRender.com

In the systematic review, I found that expert opinion was the most commonly used qualitative method in the development of DCEs. I began my work with this. It is also important to incorporate the views of patients, and the views of the population to be used for the final survey. I therefore planned surveys and focus groups in these populations.

The results of the qualitative work were used to inform the final DCE design. I conducted further reviews of the evidence to estimate the risk of ADR associated with each drug chosen for the final design.

This work was approved by the University of Liverpool Health and Life Sciences Research Ethics Committee (Human participants, tissues and databases), reference 4736.

6.2.2 Qualitative work: Survey of healthcare professionals

6.2.2.1 Aim

The aim of this survey was to gain an overview of some potential attributes for a DCE in pharmacogenetics, from healthcare professionals and academics in the

field. The previous systematic review (Chapter 5) informed the attributes included in this initial survey.

6.2.2.2 Methods

The scenario used in this survey was an early idea for the final DCE scenario. Participants were asked to imagine a scenario where a patient underwent a pharmacogenetic test and then had a choice between two medications – one which their genetic test indicated they were at a high risk of a severe ADR, and one that was less effective but also had a lower risk of the ADR. Participants were then asked to choose the most important characteristic for this genetic test, from groups of related possible attributes identified from the systematic review (Table 6.1). The survey was distributed in October 2019 using the SurveyMonkey platform (32), and was sent to academics and clinicians, working in pharmacogenetics, located through the networks of supervisors Professor Andrea Jorgensen, Professor Dyfrig Hughes, and Professor Sir Munir Pirmohamed. A link to the survey was also distributed at the International Clinical Trials Methodology Conference 2019 (33). There was no compensation provided for this survey. The full survey is provided in Appendix 6.1 (survey of healthcare professionals).

Test characteristics	**Medication choices**	**Test information**	**Practicalities**
Time to result	Efficacy of first- and second-choice medications	Information on specific gene polymorphism(s)	How sample is collected (saliva, blood, etc)
Cost of the test	Risk of severe ADRs	A panel of several pharmacogenes	Who is involved in ordering, interpreting and explaining results to patients
Level of evidence for testing	Risk of mild ADRs	Whole genome sequencing	Privacy of test results
Coverage of the test	Cost/cost-effectiveness	Easily understandable interpretation of test result	

PPV			
NPV			
If test included in BNF			
+ Reasons for selection			

Table 6.1 - potential attributes for discrete choice experiment included in the survey of healthcare professionals. The list of attributes was split into 4 groups to ease participant burden. Each group also included an 'Other' option where participants were invited to write in their own answer. Participants were also given a free text section to explain the reasons for their selection (optional). BNF = British National Formulary. NPV = negative predictive value. PPV = positive predictive value.

6.2.2.3 Results

The final sample size was n = 17. Most participants were recruited from academia (n = 8, 47.1%). Out of all participants, the majority had never ordered a genetic test themselves (n = 10, 58.8%). Just over half had used the results of genetic testing to inform prescribing or treatment of a patient (n = 9, 52.9%) (Figure 6.3A). Note that the survey did not confirm if all participants were prescribers.

A

	Academic (N=8)	Hospital doctor (N=3)	Industry (N=1)	Multiple (N=2)	Other (N=3)	Overall (N=17)
Have ordered a genetic test						
No	5 (62.5%)	1 (33.3%)	1 (100%)	0 (0%)	3 (100%)	10 (58.8%)
Yes	3 (37.5%)	2 (66.7%)	0 (0%)	2 (100%)	0 (0%)	7 (41.2%)
Have used the results of a genetic test						
No	5 (62.5%)	0 (0%)	1 (100%)	0 (0%)	2 (66.7%)	8 (47.1%)
Yes	3 (37.5%)	3 (100%)	0 (0%)	2 (100%)	1 (33.3%)	9 (52.9%)

Figure 6.3 - A) details of participants in healthcare professionals survey (n=17), including whether they had ordered a genetic test for a patient or used the results of a genetic test (ordered by themselves or

others) to inform prescribing or treatment or a patient. Those that answered 'Other' were 'pharmacy', 'consultant clinical scientist', and 'older people's medicine'. B) Participant preferences for test characteristics. C) Participant preferences for medication choices. D) Participant preferences for test information. E) Participant preferences for the practicalities of genetic testing. ADRs = adverse drug reactions. BNF = British National Formulary. NPV = negative predictive value. PPV = positive predictive value.

In the first group of attributes (test characteristics), 'level of evidence for testing' was strongly favoured over other potential attributes (Figure 6.3B). In the second group (medication choices), 'efficacy and effectiveness of first- and second-choice medications' and the 'risk of severe ADRs' were the most important attributes (Figure 6.3C). In the third group (test information), 'easily understandable interpretation' was strongly favoured (Figure 6.3D). In the final group (practicalities), participants were most concerned about 'who is involved in ordering/interpreting/explaining results to patients' (Figure 6.3E).

Participants were also asked to provide brief rationales for their choices. Level of evidence was highly valued:

"If I was ordering a test, I would want to know that it was clinically relevant unless I was ordering a test as part of a research study or clinical trial"

"Data from RCTs critical as to the absolute involvement of mutation in a gene(s) leading to ADR."

In relation to ADRs, participants indicated that reducing the risk of severe ADRs was a high priority:

"Do no harm. Better to prescribe slightly less efficacious if reduction in severe adverse effects"

"Risk of severe ADRs and efficacy/effectiveness of treatments are most important but preventing severe ADRs would be a priority."

"do no harm"

Finally, participants were asked to add any further characteristics they thought were important but had not yet been mentioned. These are shown in Table 6.2.

Any further characteristics
Speed of test result availability. This might be less of a problem in primary care but much more problematic in secondary care if for example the test takes 3-4 days to come through as the patient could have been discharged by then.

In mental health there has been less research into pharmacogenetics. This is hindering clinical uptake. And the limited clinical use in turn makes it harder to get pgx research funded in this area. It's a catch 22 situation that reflects how far we are from parity of esteem with physical health.
Would be good to find out how the public prioritises potential health benefit gained from pgx testing versus cost of testing.
Level of evidence for testing - Sensitivity/specificity of the test - Reports with clear interpretation - Health professional must explain to patients to maximise compliance
Cost of test - in terms of saving to the NHS
With most drugs genetics only plays a part in explaining efficacy/safety. There are also other patient and clinical characteristics (e.g. age, body size, sex, concurrent illness and drugs) that need to be considered as potential contributors.

Table 6.2 - Responses given by healthcare professionals when asked to define any further characteristics they thought were important in a genetic test. NHS = National Health Service.

These results provided an important basis for future work. I now wished to compare these views of professionals with those of patients.

6.2.3 Qualitative work: Survey of patients

6.2.3.1 Aim

The aim of this piece of work was to gain a perspective on pharmacogenetic testing from people with long term conditions (‘patients’). This is a group that have experience of interacting with health systems, and are more likely to have suffered ADRs, since they are more likely to be taking medications. Patient-centred research ensures greater research quality and relevance (34), and is a prerequisite for many research grants (34, 35).

6.2.3.2 Methods

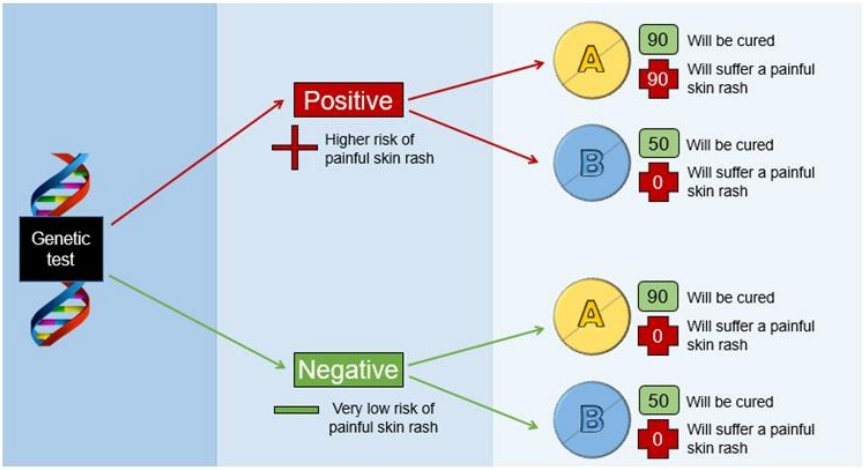
In November 2019, I contacted the North West Coast Clinical Research Network (NWC-CRN) to gain a patient's perspective on genetic testing for the prevention of ADRs. There is a group within the network made up of people with long-term conditions who are willing to be involved in research (36). There was no compensation provided for this survey. The full survey is provided in Appendix 6.2 (survey of patients).

Participants were presented with a simpler version of the survey given to HCPs, with a choice between medicines A and B, with varying risks of severe ADRs, and an accompanying genetic test (Figure 6.4A). They were asked to choose the top 5 things that were most important to them about the hypothetical genetic test, and the most important thing to consider when deciding whether or not to use the test (from a smaller list) (Figure 6.4B).

I also tested different ways of communicating risks (Figure 6.4C-F), to help inform this display in the final DCE. The survey was also conducted on the SurveyMonkey

platform (32).

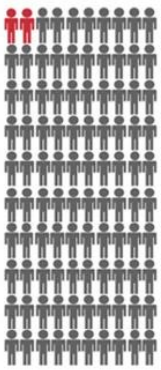
A



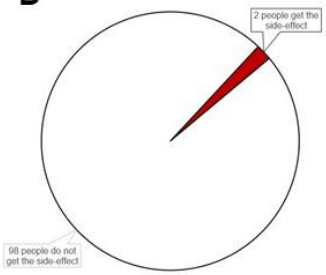
B

Most important thing to consider?	Top 5 things by importance
Severity of the disease being treated	Severity of the disease being treated
How much evidence there is to show that the genetic test works in predicting risk	How much evidence there is to show that the genetic test works in predicting risk
How accurate the genetic test is	How accurate the genetic test is
Risk of getting the serious side-effect if you take medicine A	Risk of getting the serious side-effect if you take medicine A
How well medicine A and B work	How well medicine A and B work
Who you see for the genetic testing service	Who you see for the genetic testing service
How the test is done	How the test is done
Privacy of your test results	Privacy of your test results
	Risk of developing a severe side-effect
	Risk of developing a mild side-effect
	Time to wait for the result
	Level of information you receive from the test
	If your doctor recommends you receive the test
	If most other people in your situation choose to take the test
	Cost of the test to you personally
	Cost of the test to the NHS
	Cost of medicine A and B to the NHS

C



D



E

2 out of 100 people experience the serious side-effect

F

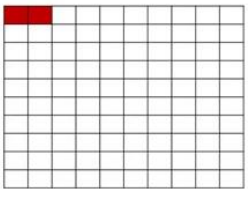


Figure 6.4 - A) Diagram presented to patients of the risk of an ADR (painful skin rash) with medications A and B. The test and medications presented here are hypothetical. B) A list of genetic test characteristics that patients were asked to choose between. They were asked to choose the most important characteristic, and the top 5 most important characteristics. C-F show different ways of representing the same level of risk. Patients were asked to choose which they thought was the clearest. C) pictogram, D) pie chart, E) written, F) boxes.

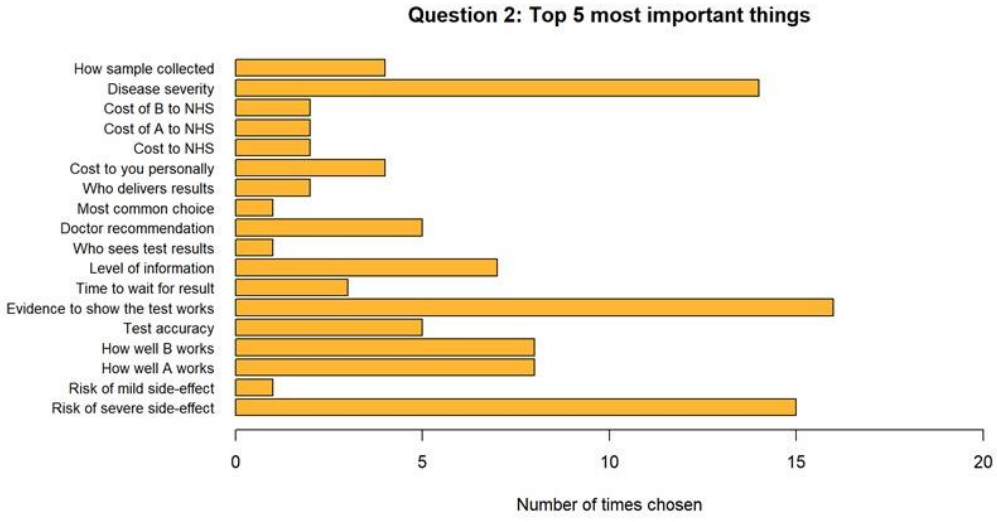
6.2.3.3 Results

A total of 20 patients were recruited. Only 1 had previously had a genetic test, 1 did not know, and 2 preferred not to answer. The group was 55% male, and 75.0% of the participants were over the age of 55 (Figure 6.5A). This was as expected due to the known composition of the group.

A

Have you had a genetic test before?	Age	Gender
Yes (N=1)	18-24: 0 (0%) 35-44: 0 (0%) 45-54: 0 (0%) 55 and over: 1 (100%) Prefer not to answer: 0 (0%)	Female: 1 (100%) Male: 0 (0%) Prefer not to answer: 0 (0%)
No (N=16)	18-24: 1 (6.3%) 35-44: 1 (6.3%) 45-54: 2 (12.5%) 55 and over: 11 (68.8%) Prefer not to answer: 1 (6.3%)	Female: 7 (43.8%) Male: 8 (50.0%) Prefer not to answer: 1 (6.3%)
Don't know (N=1)	18-24: 0 (0%) 35-44: 0 (0%) 45-54: 0 (0%) 55 and over: 1 (100%) Prefer not to answer: 0 (0%)	Female: 0 (0%) Male: 1 (100%) Prefer not to answer: 0 (0%)
Prefer not to answer (N=2)	18-24: 0 (0%) 35-44: 0 (0%) 45-54: 0 (0%) 55 and over: 2 (100%) Prefer not to answer: 0 (0%)	Female: 0 (0%) Male: 2 (100%) Prefer not to answer: 0 (0%)
Overall (N=20)	18-24: 1 (5.0%) 35-44: 1 (5.0%) 45-54: 2 (10.0%) 55 and over: 15 (75.0%) Prefer not to answer: 1 (5.0%)	Female: 8 (40.0%) Male: 11 (55.0%) Prefer not to answer: 1 (5.0%)

B



C

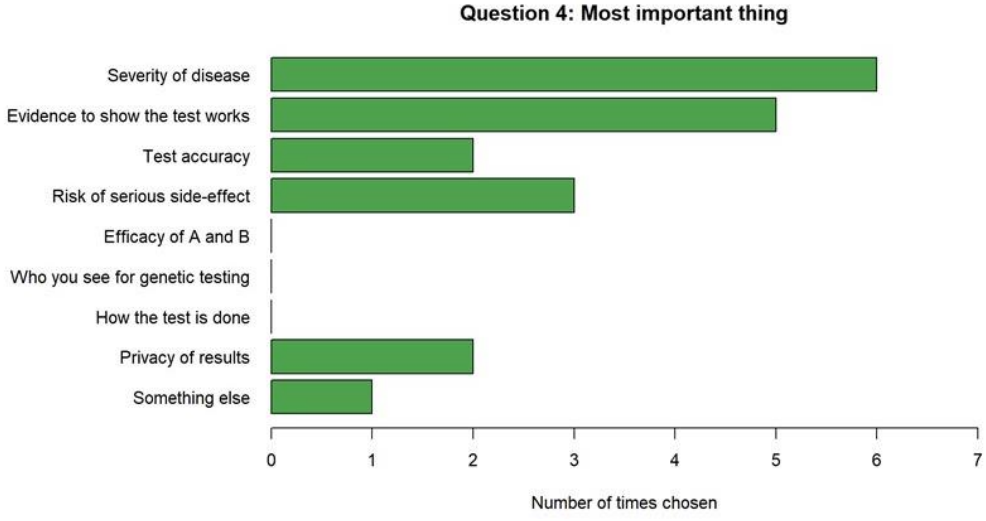


Figure 6.5 - A) characteristics of patients recruited for this survey. B) Choices made by patients when asked to choose the top 5 most important attributes. C) Choices made by patients when asked to choose the single most important attribute.

When asked to choose the top 5 most important things from a list of attributes, the highest number of participants were concerned about the evidence to show that the genetic test worked (Figure 6.5B). When asked to choose the single most important thing about the hypothetical genetic test, the severity of the underlying disease was the most important attribute.

Participants were also asked to provide brief rationales for their choices. On the level of evidence to show a test works to predict the risk of a side-effect:

"I feel that if I was to consider the test I would like to be appropriately informed before I decided to go ahead."

"I would hope that there would be an action resulting from having a genetic test....either to give or not to give a particular medication, otherwise there would be no point in undertaking the test."

The severity of the disease being treated was also a consideration:

"If genetic testing could have predicted this problem I would have been regularly tested and at the first sign of trouble it could have been treated."

"If one sees another family member suffering badly, then I believe I would brave enough to have a test."

Other concerns included privacy and cost:

"The sharing of the information could be more damaging than the disease being treated."

"Cost & time taken to develop the test and administer it need to have a significant impact."

Participants were also asked which ways of communicating risk were clearest. Pictograms and pie charts were the most popular choices (Figure 6.6). As there was no clear answer, this was tested further in focus groups.

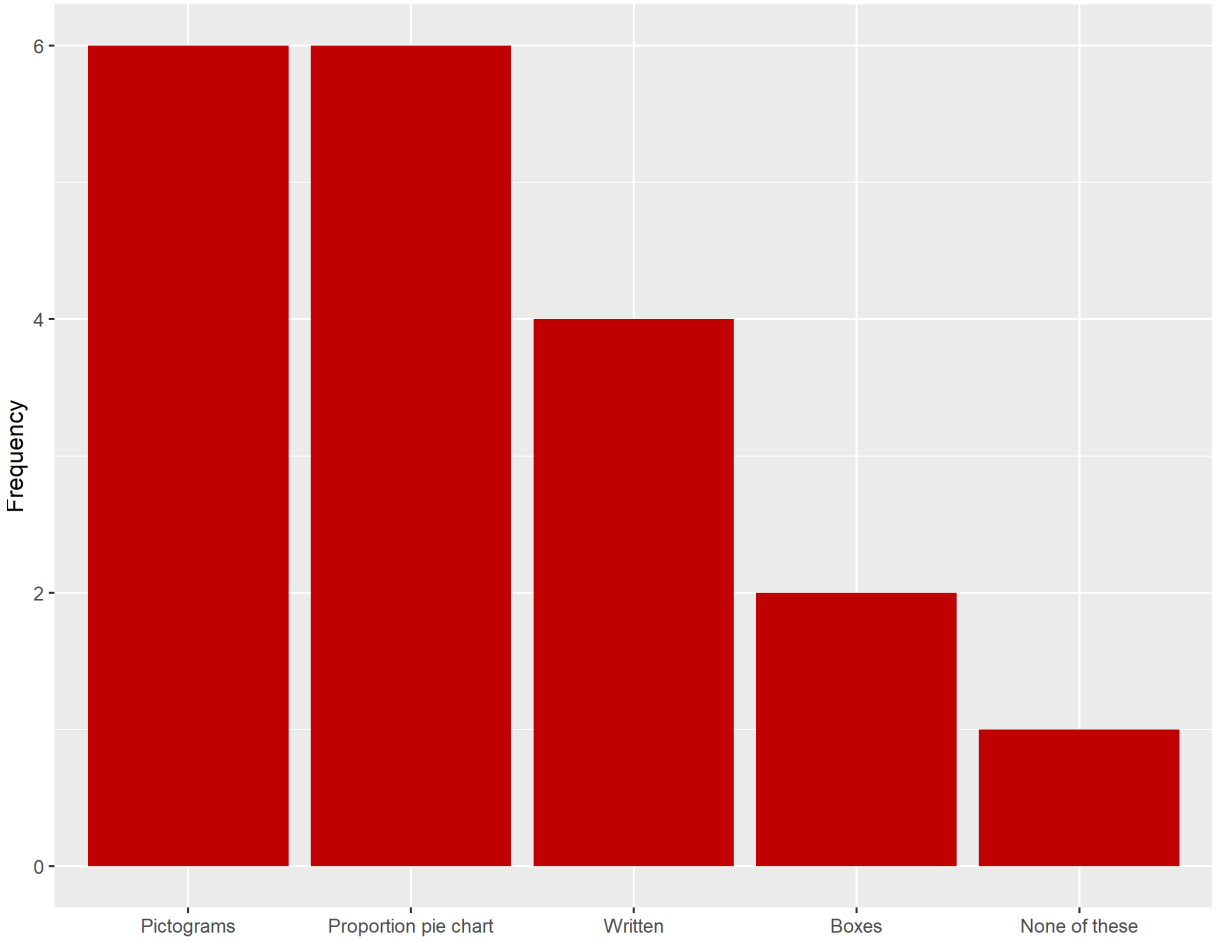


Figure 6.6 - Testing which method of presenting risk is most preferred by patients (n=20).

6.2.4 Qualitative work: Focus groups with the general public

6.2.4.1 Aim

Focus groups allow wider discussion of topics, and may generate more information than might be gathered from a single individual (37, 38). Two focus groups with members of the general public were conducted in early March 2020. Further groups were planned but unable to take place due to the COVID-19 pandemic.

6.2.4.2 Methods

Participants were given a brief presentation explaining genetic testing, and how it can be used to predict drug response (focusing on the prevention of ADRs). They were asked to imagine a hypothetical genetic panel that would test for 12 genes. The results of testing one of these genes would be immediately used, some would be useful in the future, and some would never be used (Figure 6.7A). Sessions were recorded and transcribed with full consent of participants. Discussions were categorised into themes using NVivo 12 software (QSR International) (39).

Results

Two groups were recruited (n = 6 and n = 3). This was done so that two dates could be offered for participants to attend. Details of participant ages or genders, or whether they had previously had genetic tests, were not collected. Participants received a £10 voucher for participation.

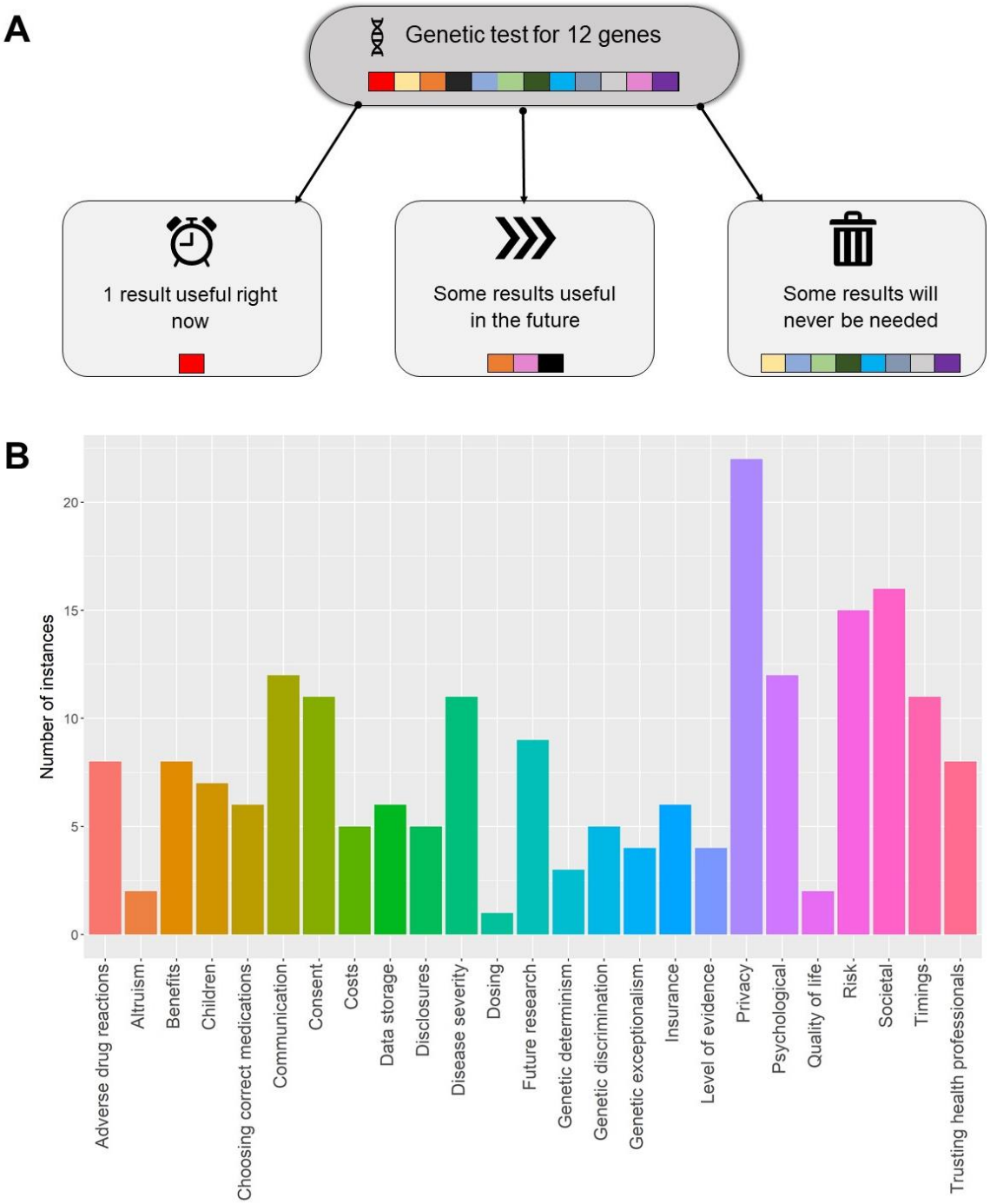


Figure 6.7 – A) Diagram shown to focus group participants to explain how a genetic panel test could produce results with different outcomes – in this hypothetical panel of 12 genes, 1 test result is useful now, some are useful in the future, and some results will never be needed. B) Number of instances of each category of discussion in focus groups conducted March 2020 (n=9). Groups were coded from transcripts using NVivo software (39). Number of instances refers to the number of sections where the topic was discussed, across both focus groups.

Privacy concerns dominated both discussions (Figure 6.7B). This was also seen in focus groups used in development of other DCEs (Chapter 5) (40, 41). Some concerns include:

"it depends how its stored, where its stored, who has access to it. If its anonymised then you're just a number."

"Cause if its just your doctor that has access to that information, that's very different than if they're sharing it with all insurance companies and everyone."

"I've worked in an NHS setting and seen what data security looks like. Or not looks like"

Other concerns were around societal impacts of genetic testing:

"In a hundred years time, people will probably be thinking, why on earth weren't we doing it 100 years ago?"

and risks associated with drugs and testing:

"anything that's above a 50 50 chance. Then I'd probably wanna know about it".

Level of evidence was less discussed in these groups. If it had been possible to run further focus groups, I would have tried to guide the conversation more in this direction. Other reports of focus groups also provided guidance in this area (40, 42, 43). Some participants also brought up the altruistic value of 'donating' their data to research and this was generally positively received.

Participants were also asked their opinions on representing risk, using the same diagrams as in the previous survey. Pictograms were the most preferred (7/9 participants) followed by pie charts. I therefore decided to use pictograms to represent risk going forward.

6.2.5 Initial choice of attributes and levels

Following this set of work, a set of 5 attributes and their accompanying levels were chosen, as shown in Table 6.3.These were based on results from each of the above sets of work (survey of healthcare professionals, survey of patients, focus groups with the general public), along with the previously discussed results of the systematic review.

Attribute	Rationale	Levels	Rationale
Cost of the test to the NHS	Allows quantification of preferences (21, 23) and is based on a recommendation from a supervisor (DH)	£20	These levels are based on Illumina sequencing (44) and a personal communication (DH)
		£40	
		£60	
Use of your data for further research by universities and researchers	Reflects the importance of privacy, particularly to focus group participants	Yes, and they can contact me (linked to medical record)	Allows the incorporation of altruistic donation of data for research, discussed in focus group
		Yes, but no contact (anonymous)	
		No	
Number of drugs the test can be used to inform	This will capture the use of panel vs single gene testing	1	These levels chosen based on a personal communication (DH)
		25	
		50	
Number of genetic tests, besides this one, that you might require over the next 10 years	Similarly to the above, this captures whether a single gene test is sufficient for future use	0	These levels chosen based on a personal communication (DH)
		1	
		2	
Chance of serious side-effect from any treatment over the next 10 years	The risk of a serious ADR is the main aim of this DCE.	1 in 10	These levels chosen arbitrarily for now.
		1 in 100	
		1 in 1000	

Table 6.3 - Attributes and levels used in a pilot test of the DCE. ADR = adverse drug reaction.

I also decided to include a choice of two different tests, and allow participants to choose 'neither', or ‘no test’. This option more closely reflects clinical practice.

6.2.6 Qualitative work: Pilot testing

6.2.6.1 Aim

The aim of pilot testing is to understand any issues in the DCE design that should be modified before its final deployment. An online pilot test was conducted in a

small group of participants from the focus group that consented to further contact. This also included some participants that had expressed interest in participating in the focus groups but were unable to attend on the times and dates offered.

The more specific aim of this work was to test a group of 5 potential attributes that were chosen based on the results of the previous sets of qualitative work (see above). I also wanted to check the wording of the explanations provided to participants.

6.2.6.2 Methods

In this survey I tested whether participants would trade in two choice tasks, and checked the level of understanding for explanations of different potential levels. Initially, participants were asked to imagine they had been diagnosed with a type of colorectal cancer and their doctor had recommended capecitabine. Details of a capecitabine ADR and a genetic test to predict their risk were provided. Information about each potential attribute was provided. Participants were then asked to choose between two tests (or no test) described by the 5 different attributes from Table 6.3. Only two choice tasks were used, so the remainder of the survey could be used to examine participant views about the survey introduction and explanations. Participants were asked if they understood the explanation of each attribute, and asked to rank (on a scale of 0 to 100, where 0 is not important and 100 is the most important) the importance of each attribute in their decision making during the choice tasks. This would allow measurement of whether each attribute was considered important, and to gain an initial idea of their relative importance. In exchange for participation, participants were entered into a prize draw to win a £10 voucher.

6.2.6.3 Results

A total of 16 participants were recruited, mostly aged 25 to 34 (43.8%) and 50% male (Figure 6.8A). Overall, participants found the questions easy to complete. No participants found the survey ‘very difficult’ to complete (Figure 6.8B). Participants were then asked about their understanding of each attribute, and the importance of each attribute on a numerical scale (from 0 to 100).

Each attribute was well understood. No participant indicated they did not understand any attribute. The risk of ADRs, and the number of drugs the test could be used to inform were the most important attributes on a 1 (not important) to 100 (the most important) scale (Figure 6.8C).

A

	Overall (N=16)
Age	
25 to 34	7 (43.8%)
35 to 44	2 (12.5%)
45 to 54	4 (25.0%)
Missing	3 (18.8%)
Gender	
Female	5 (31.3%)
Male	8 (50.0%)
Missing	3 (18.8%)

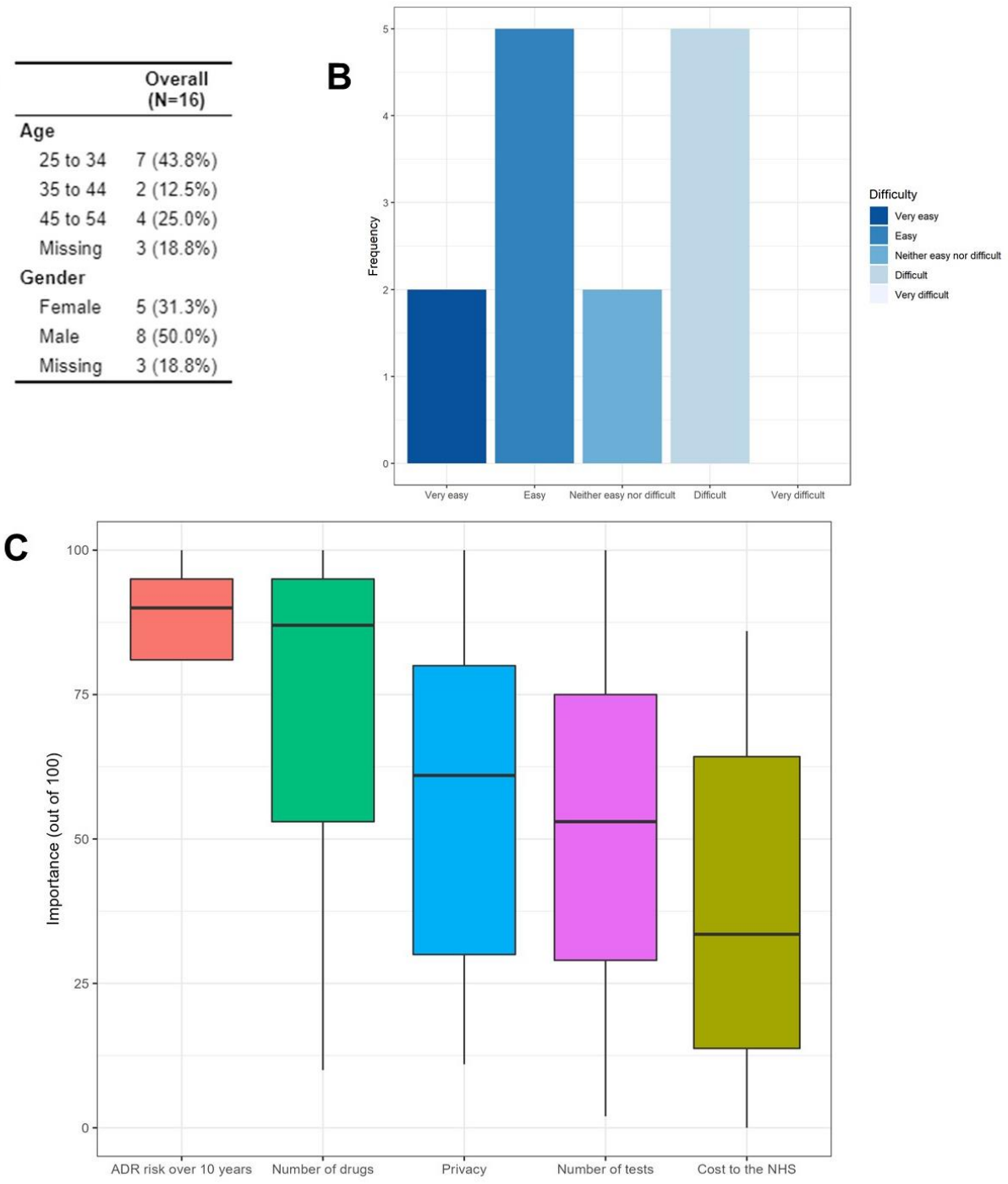


Figure 6.8 - Results of pilot testing (n= 16). A) Participant age groups and genders. B) Participant ratings of the ease of completing choice tasks. C) Importance of each potential DCE attribute to participants on a scale of 0 (not important) to 100 (the most important). Number of drugs is the number of drugs the test can be used to inform, number of tests is how many genetic tests that participants may require over the next 10 years.

The risk of an ADR over 10 years was rated the most important attribute (mean 82/100). Participants were less concerned about the cost of genetic testing to the NHS (39/100). Privacy was not rated as important as expected based on the focus

group opinions (58/100). The ‘number of drugs’ attribute was rated as more important than ‘number of tests’. As these represent similar characteristics of a genetic test, I decided to keep only the ‘number of drugs’.

Informal feedback from this survey received by email also indicated that some people who wished to participate were put off by the subject matter (colorectal cancer) because of its severity and risk of death. I chose to continue with potentially distressing scenarios, but to provide warnings before starting the DCE, further stress the hypothetical nature of the test, and provide links to support telephone numbers and websites.

6.3 Results: DCE design

6.3.1 Design and choice of attributes and levels

Upon reviewing the results of pilot testing and combining this with results of the other qualitative work, along with further email consultation with experts (since it was not possible to run further focus groups), the choice of attributes was amended to bring the risk of serious ADRs into greater prominence and still incorporate issues of genetic panel testing and level of evidence.

I wanted to examine how level of evidence would inform DCE decision making. However, there was uncertainty about how to incorporate this into an attribute without having to explain concepts such as RCTs, meta-analyses, and quality of evidence to participants, which would make the DCE too long which could put participants off completing it. It was therefore decided to produce 2 sets of DCEs, one with ‘high’ levels of evidence for their genetic test, and one with ‘low’ levels. High levels of evidence were defined as Level 1 (A or B) in the Pharmacogenomics Knowledgebase (PharmGKB) clinical annotations (45, 46). This is the top level of evidence in PharmGKB and is awarded to variant-drug combinations that have been recommended in a clinical guideline (46). Low levels of evidence were Levels 2 to 4 (the lowest level).

I decided to produce 8 different DCEs, each with different gene-drug-ADR combinations. Half of these would be gene-drug combinations with high levels of evidence, and half with low levels of evidence. Participants were then randomised to receive one of the eight DCEs. These would also each have their own ADRs and associated risks, which would varied within a specific attribute.

I reviewed PharmGKB and consulted with experts to choose the gene-drug-ADR combinations (Table 6.4). I chose a high evidence and low evidence example for each of 4 indications: HIV treatment, cancer, epilepsy, and cardiovascular disease.

Gene(s)	Drug	Indication	ADR	Evidence (clinical annotations)	Evidence (drug label annotations)
High evidence					
*HLA-B*57:01*	Abacavir	HIV	Abacavir hypersensitivity syndrome	Level 1A	Testing required (FDA)
					Testing required (EMA)
					Testing required (Canada)
					Testing required (Swiss)
					Informative PGx (Japan)
DPYD	Capecitabine	Cancer	Neutropenia	Level 1A	Testing recommended (EMA)
					Actionable PGx (FDA)
					Actionable PGx (Japan)
					Actionable PGx (Canada)
					Actionable PGx (Swiss)
*HLA-A*31:01*	Carbamazepine	Epilepsy, pain, others	SJS/TEN	Level 1A	Testing recommended (Swiss)
CYP2C9/ VKORC1	Warfarin	Cardio-vascular	Bleeding	Level 1A	Actionable PGx (FDA)

					Actionable PGx (Canada)
Low evidence					
CYP2B6	Efavirenz	HIV	DILI	Level 2A *†	Actionable PGx (FDA)
					Actionable PGx (EMA)
					Actionable PGx (Japan)
					Actionable PGx (Canada)
					Actionable PGx (Swiss)
UGT1A1	Irinotecan	Cancer	Neutropenia	Level 2A	Testing recommended (Japan)
					Actionable PGx (FDA)
					Actionable PGx (EMA)
					Actionable PGx (Canada)
					Actionable PGx (Swiss)
*HLA-A*24:02*	Phenytoin	Epilepsy	SJS/TEN	Level 3	None
SLCO1B1	Atorvastatin	Cardio-vascular	Myopathy	Level 3	Actionable PGx (Swiss)

*Table 6.4- Gene drug ADR combinations used. Clinical annotations are assigned by PharmGKB to describe the level of evidence for a variants phenotypic impact. Drug label annotations are referenced on PharmGKB, collated from USA, European, Swiss, Japanese, and Canadian health bodies. *Level 2A for prevention of ADRs, other levels for other uses. † Since this study was conducted, this has been upgraded to Level 1A. ADR = adverse drug reaction. DILI = drug-induced liver injury. EMA = European Medicines Agency. FDA = Food and Drug Administration (USA). PGx = pharmacogenetics. SJS/TEN = Stevens-Johnson syndrome, toxic epidermal necrolysis.*

For each combination, an estimate of the risk of each ADR with and without genetic testing was required. These estimates would then be incorporated into an attribute (and its levels) unique to each indication, i.e., both epilepsy DCEs would have the same levels. The levels should encompass the range of each set of estimates. These estimates would also be used at the analysis stage to calculate the utility of single gene vs gene panel testing. To obtain these estimates, I searched the literature and conducted mini meta-analyses where appropriate.

6.3.1.1 Developing risk estimates: risk of serious side-effect from this medication

The risk of the ADR associated with each drug, without genetic testing, was calculated using cohort studies (where unselected patients receiving the drug of interest were followed to identify those developing the ADR). The risks from these studies were meta-analysed (where appropriate) to provide an estimate of the risk of each ADR where testing is not done. ADRs were chosen based on ADRs that were known to have a pharmacogenetic component within each drug.

The risk of the ADR when genetic testing was in use was calculated using studies that evaluated patients receiving each drug by genotype. The risk for all patients was assumed to be the same as the risk as patients with low-risk genotypes, as these patients would still receive the drug if genetic testing was used. Patients with high-risk genotypes would not be given the drug (or in some cases, a reduced dose of the drug instead).

However, where a trial of genotyping vs not genotyping existed, this data was used directly as the risk estimate. Where existing systematic reviews or meta-analyses were available, these results were used to calculate risk estimates. PharmGKB resources were used as a further source of data. Where possible, data collected in UK or European populations was used, as variant gene frequency varies across different populations.

The 95% confidence intervals for estimates were obtained using the BinomCI R package from DescTools (47).

6.3.1.1.1 Abacavir – *HLA-B*57:01* – AHS

Abacavir hypersensitivity syndrome (AHS) is an abacavir ADR causing fever, chills, rash, vomiting, and fatigue (48, 49). The reaction occurs in 4.3% of patients receiving abacavir (50), and can be fatal (49). AHS is strongly associated with the *HLA-B*57:01* allele (51).

Calculating the risk for this gene-drug combination was straightforward since an RCT of genotyping vs no genotyping was available, the PREDICT-1 trial (52). The trial randomised HIV patients requiring abacavir to receive *HLA-B*57:01* results prior to prescription, or to receive prescription without knowing *HLA-B*57:01* screening results. Patients were mostly White (82.8 and 82.9% in screening and control groups), with a mean age of 42. There were also more males than females in the trial. Patients positive for *HLA-B*57:01* did not receive abacavir.

Hypersensitivity reactions were assessed clinically, and immunologically confirmed by patch testing. Rates of clinically confirmed hypersensitivity were used for these risk estimates since these are more generalisable to actual clinical practice (51) (Table 6.5).

Mallal et al 2008 PREDICT-1 (52)			
Outcome	n	Risk of outcome	95% CI
AHS/Total	1772	79/1772 = 0.045	0.036 - 0.055
AHS/ clinically diagnosed, screened group	803	27/803 =0.034	0.023 - 0.048
AHS/ clinically diagnosed, control group	847	66/847 = 0.078	0.062 - 0.098
AHS/ immunologically diagnosed, screened group	802	0/802 = 0	-
AHS/ immunologically diagnosed, control group	842	23/842 = 0.027	0.018 – 0.041

*Table 6.5 – A previous randomised controlled trial of HLA-B*57:01 and abacavir hypersensitivity syndrome where patients were randomised to genotyping or not prior to prescription. The abacavir hypersensitivity syndrome could be clinically diagnosed or immunologically diagnosed. Clinical diagnosis was used as this is more generalisable. AHS = abacavir hypersensitivity syndrome.*

Using these data, the risk of AHS without genetic testing is 0.078 (95% CI 0.062 – 0.098), and the risk with genetic testing is 0.034 (95% CI 0.023 – 0.048).

6.3.1.1.2 Capecitabine – *DPYD* – Neutropenia

Neutropenia (a low neutrophil count) increases mortality by leaving patients more vulnerable to infection (53, 54). It can also compromise and delay treatment of the underlying cancer (54). Neutropenia is associated with capecitabine treatment (55-57).

Capecitabine is first metabolised to its active form, fluorouracil. This active form is an antimetabolite that slows tumour growth (58). Fluorouracil is metabolised into inactive components by the enzyme dihydropyrimidine dehydrogenase (DPD). This enzyme is encoded by *DPYD* (58, 59). Patients with DPD deficiencies are more likely to suffer from ADRs in capecitabine treatment, including neutropenia (59).

A safety and cost analysis of *DPYD* genotyping in Dutch patients treated with fluoropyrimidine-based chemotherapy calculated the risk of grade >=3 toxicity with and without genotyping (60). The risks were calculated using data from a prospective cohort study combined with historical controls located in the literature. The majority of participants received capecitabine (90%), the remainder received fluorouracil.

Participants (n=2038) were prospectively genotyped for the *DPYD*2A* risk allele. Wild-type (WT) patients received standard dosing, and patients with the risk allele received reduced doses of capecitabine (Table 6.6). Of the patients that received therapy (n=1631), 134 experienced a grade >=3 neutropenia. This (0.082) is the risk of neutropenia with genotyping.

For the risk without genotyping, data from this paper was also used. Of the total population, 98.8% were WT and 1.2% had *DPYD*2A* polymorphisms. In total, 23.12% of the WT population experienced a toxicity of grade >=3. In the *DPYD*2A* population, 72.9% experienced this.

From another piece of supplementary data to this paper, neutropenia was found to represent 35.1% of all toxicities grade 3 and above in WT patients. In the *DPYD*2A* population, this proportion was 60%. Combining all this information allows the calculation of the probability of grade >=3 neutropenia:

$$(0.988 * 0.2312 * 0.351) + (0.012 * 0.729 * 0.600) = 0.0854$$

Deenen et al 2016 (60)			
Outcome	n	Risk of outcome	95% CI
Neutropenia/ Total in a genotype-guided approach	1631	134/1631 = 0.0822	0.070 – 0.096

Table 6.6 – Previous studies of DPYD and neutropenia in patients receiving capecitabine.

6.3.1.1.3 Carbamazepine – *HLA-A*31:01* – SJS/TEN

Data from Plumpton, et al 2015 (61) was used to obtain an estimate for the utility of testing for *HLA-A*31:01* in carbamazepine patients. This was an economic analysis using a decision analytic model to estimate the cost-effectiveness of screening for *HLA-A*31:01* in epilepsy patients. The estimates were produced by incorporating data from an RCT of treatments for epilepsy (the SANAD trial) (62) and a genome-wide association study in Northern Europeans (63). This data was used as it was a robust analysis that incorporated many of the data sources that would have been used in this analysis, in the manner of the other drug-gene-ADR sets listed here. The risk of SJS/TEN without *HLA-A*31:01* screening was 1.18 per 10,000 patients. With screening, this was 0.87 per 10,000.

Data from a cohort study in Japan that examined the utility of *HLA-A*31:01* testing to prevent cutaneous ADRs was also evaluated for inclusion (64). This study in a Japanese population is less relevant to my DCE population than the Plumpton, *et al.* paper, that used Northern European data. Therefore, only included the estimates from the Plumpton paper were used as the risk estimate (Table 6.7).

Plumpton et al 2015 (61)			
Outcome	n	Risk of outcome	95% CI
SJS/TEN in epilepsy patients, without genetic testing	n/a	0.000118	-
SJS/TEN in epilepsy patients, with genetic testing	n/a	0.000087	-

*Table 6.7 – Results of a previous study of HLA-A*31:01 and SJS/TEN in patients receiving carbamazepine.*

6.3.1.1.4 Warfarin – *CYP2C9*/*VKORC1* – bleeding ADRs

The anticoagulant warfarin is associated with bleeding events which may require blood transfusions (65) and is the cause of around 10% of hospital admissions related to ADRs (8). Bleeding events are a result of incorrect dosing, causing under-coagulation. Dosing is traditionally done using clinical algorithms, but algorithms based on pharmacogenetics are becoming more widely used (66-68). The genes *CYP2C9* and *VKORC1* are associated with warfarin response and are used in genetic dosing algorithms (68-70).

Two previous RCTs were located that compared genotype-guided and clinical dosing of warfarin (68, 71). However, neither of these trials were powered to detect bleeding events. An older RCT with a similar design had the same problems (72).

Estimates were therefore based on the 2017 GIFT RCT of genotype guided (*CYP2C9*, *CYP4F2*, *VKORC1*) vs clinically guided warfarin dosing (73). The primary endpoint of GIFT was a composite outcome that combined bleeding, clinical, and laboratory outcomes. Each outcome was also evaluated individually as secondary outcomes (Table 6.8).

GIFT randomised 1650 patients planned to undergo hip or knee arthroplasty (replacement), who required treatment with warfarin. Patients were 90% white with a mean age of 72. There were 87 patients in the genotype-guided and 116 in the clinically-guided group that experienced at least 1 composite end-point (p = 0.02).

Gage et al 2017 (73) **GIFT trial**			
Outcome	N	Risk of outcome	95% CI
Composite primary end point/ total	1597	203/1597 = 0.127	0.112 - 0.144
Composite primary end point/ genotyped group	808	87/808 = 0.108	0.088 – 0.131
Composite primary end point/ clinical-guided group	789	116/789 = 0.147	0.124 – 0.173
Major bleeding/ total	1597	10/1597 = 0.006	0.003 – 0.011
Major bleeding/ genotyped group	808	2/808 = 0.0020	0.0007 – 0.0090

Major bleeding/ clinically-guided group	789	8/789 = 0.010	0.005 – 0.020

Table 6.8 – Previous study of genetics-guided and clinically-guided warfarin treatment and the rates of a composite primary endpoint (including major bleeding), and major bleeding alone.

For the final risk estimates, I relied upon the major bleeding outcome of the GIFT trial. This is a simpler outcome than the primary composite outcome and was chosen for use here as it would be simpler to explain to DCE participants in the scenario. The risk of major bleeding in patients with risk alleles was 0.010 (95% CI 0.005 – 0.020), and the risk in patients without risk alleles was 0.0020 (95% CI 0.0007 – 0.0090).

The risk with genotyping was 0.0020 (0.0007 – 0.0090). The risk without genotyping (in the clinically guided group) was 0.010 (0.005 – 0.020).

6.3.1.1.5 Efavirenz – *CYP2B6* – DILI

Drug-induced liver injury (DILI) is a syndrome that causes hepatic necrosis, jaundice, and abdominal pain (74). It is the most common cause of acute liver failure in the United States (74, 75). DILI is a diagnosis based on exclusion of other causes of liver injury (76). DILI has an annual incidence of 14-19 per 100,000 persons, and is a leading cause of attrition in drug development (77). There are no specific diagnostic markers for DILI, but patterns of elevated liver enzymes that resolve when the drug is stopped or the dose is lowered. DILI is associated with the HIV drug efavirenz, particularly in patients with the *CYP2B6*6/*6* allele (78).

PubMed was searched for studies that recruited efavirenz patients and evaluated them for DILI. The search terms were: ("drug induced liver injury"[All Fields]) AND ("efavirenz"[All Fields]). No RCTs were located that randomised patients between genotyping and non-genotyping for efavirenz treatment. There were also no meta-analyses located. Three studies were used to calculate the risk estimate (Table 6.9).

A prospective cohort study in Ethiopia recruited anti-retroviral naïve HIV-positive participants to evaluate predictors of drug-induced liver injury (DILI) (79). DILI was defined according to criteria of the Council for International Organizations of Medicine Science (CIOMS) criteria (80). Participants received anti-retroviral therapies based on efavirenz (stavudine/lamivudine/efavirenz, or zidovudine/lamivudine/efavirenz, or tenofovir/lamivudine/efavirenz). Participants

with high liver enzymes before starting the study were excluded from analysis. Of the 261 remaining patients, 41 (15.7%) developed DILI after efavirenz-based treatment. DILI was associated with the different *CYP2B6*6/*6* alleles.

A further study in Ethiopia recruited 4 groups of patients into a cohort study, one of these being patients with HIV receiving efavirenz-based regimens alone (81). In this group, 24 out of 273 patients experienced DILI. This study did not genotype participants.

A cohort study in Tanzania by Mugusi, *et al*. included and genotyped 253 patients with HIV only and 220 HIV/TB patients (82). The first group received efavirenz-based regimen, the second also received rifampicin-based therapy. In total, 37 patients suffered DILI. In the HIV-only group, 15 patients developed DILI. Genotyping data was only provided with these groups combined.

Yimer et al 2012 (79)			
Outcome	n	Risk of outcome	95% CI
DILI/ Total	261	41/261 = 0.157	0.118 – 0.206
DILI/ *CYP2B6*1*1* (WT)	111	12/111 = 0.108	0.063 – 0.180
DILI/ *CYP2B6*1/*6*	114	20/114 = 0.175	0.117 – 0.256
DILI/ *CYP2B6*6/*6* (high risk)	20	5/20 = 0.250	0.112 – 0.469
Mugusi et al 2012 (82)			
DILI/ Total HIV-only patients	253	15/253 = 0.059	0.036 – 0.096
DILI/ Total HIV and TB patients	473	37/473 = 0.078	0.057 – 0.106
DILI/ *CYP2B6*1/*1* (WT), HIV and TB patients	147	6/147 = 0.041	0.019 – 0.086
DILI/ *CYP2B6*1/*6*, HIV and TB patients	148	16/148 = 0.108	0.068 – 0.168
DILI/ *CYP2B6*6/*6* (high risk), HIV and TB patients	54	6/54 = 0.111	0.052 – 0.222
Yimer et al 2014 (81)			
DILI/ Total	273	24/273 = 0.088	0.060 – 0.127

Table 6.9 - Previous studies of CYP2B6 and DILI in patients receiving efavirenz. DILI = drug induced liver injury.

Using these data in a proportion meta-analysis, the risk of DILI with genetic testing is 0.07 (0.03 – 0.13). The risk without testing is 0.10 (0.07 – 0.15) (Figure 6.9).

All three of these studies were conducted in African countries, potentially making the data less applicable to my DCE population. However, the frequency of the high risk *CYP2B6*6/*6* allele is comparable across European and African populations (3.4% and 5.8% respectively) (83). Similar risk profiles can therefore be assumed in this case.

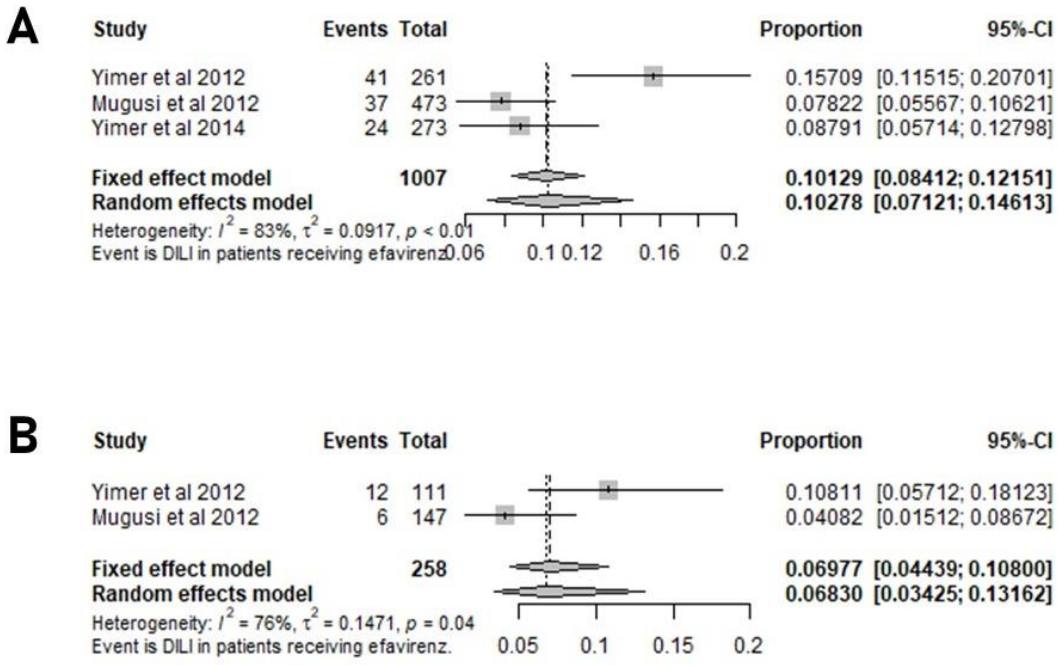


Figure 6.9 - Efavirenz meta-analysis A) risk of DILI without genetic testing, B) with genetic testing. DILI = drug-induced liver injury.

6.3.1.1.6 Irinotecan – *UGT1A1* – neutropenia

Irinotecan, used for the treatment of colorectal cancer (84), is associated with the development of neutropenia (85, 86). *UGT1A1* variants (particularly *UGT1A1*28*28*) are associated with increased risks of neutropenia (86). This is also dependent on dosage. A 2007 meta-analysis of irinotecan haematologic toxicity

(grade III and IV by Common Terminology Criteria [CTC]) found that the risk of toxicity increased with increasing irinotecan dose in patients with the *UGT1A1*28/*28* genotype, but not with the *UGT1A1*1/*1* genotype (87).

For neutropenia in irinotecan treatment, two meta-analyses were located (Table 6.10). Also located was a Cochrane review that compared the efficacy and safety of irinotecan monotherapy and irinotecan in combination with other drugs (85). From this, data on neutropenia in irinotecan monotherapy was extracted. Doses used in the irinotecan-only arms of the trials varied from 125 mg/m^2 to 350 mg/m^2. The effect of dose was not considered in the meta-analysis. Grade 3/4 neutropenia (by CTC guidelines) affected 129 out of 672 patients in the irinotecan-only arm. This meta-analysis did not consider the genotypes of the participants.

One further meta-analysis of irinotecan and neutropenia was located. This meta-analysis included trials investigating the link between *UGT1A1* variants and toxicities (88). The genetic analysis included 2334 patients across 30 studies. All trial types were included in this analysis. Neutropenia was defined as grade >=3 using varying validated criteria. Doses of irinotecan varied from 50-375 mg/m^2. The numbers of participants used in these meta-analyses were extracted.

Wulaningsih et al 2016 (85) **Cochrane review (systematic review and meta-analysis)**			
Outcome	n	Risk of outcome	95% CI
Neutropenia/ Total (5 RCTs)	672	129/672 = 0.192	0.164 – 0.223
Yang et al 2018 (88) (systematic review and meta-analysis)			
Neutropenia /Total	4075	737/4075 = 0.181	0.169 – 0.193
Neutropenia/ *UGT1A1*1/*1* (WT)	2334	382/2334 = 0.164	0.149 – 0.179
Neutropenia/ *UGT1A1/*28/*28* (risk variant)	275	82/275 = 0.298	0.247 – 0.355
Neutropenia/ *UGT1A1*6* AA genotype (variant)	144	48/144 = 0.333	0.262 – 0.414
Neutropenia/ *UGT1A1*6* GG genotype (variant)	1322	225/1322 = 0.170	0.151 – 0.191

Table 6.10 – Previous meta-analyses of UGT1A1 and neutropenia in patients receiving irinotecan. RCTs = randomised controlled trials.

To calculate the risk of neutropenia without genetic testing, these studies were combined in a random effects meta-analysis (Figure 6.10). The risk of grade >=3 neutropenia without testing is 0.182 (0.172 – 0.194). The risk with genetic testing was taken as the risk in WT participants in the Yang, *et al.* study (88). The risk with testing is 0.164 (0.149 – 0.179).

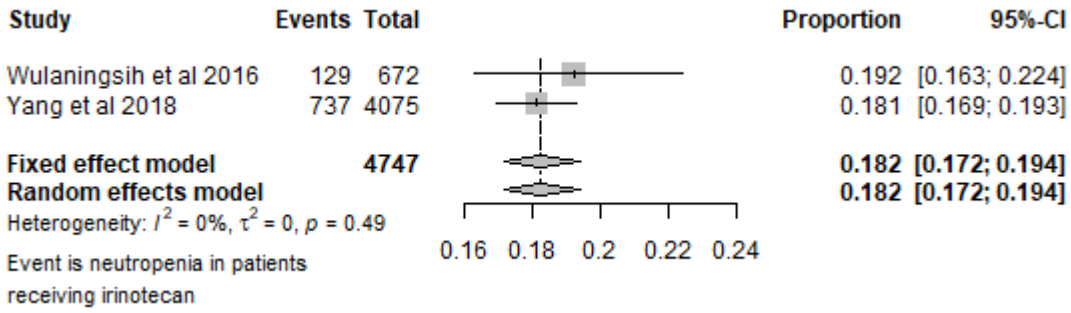


Figure 6.10 - risk of grade 3 or above neutropenia in patients receiving irinotecan, without genetic testing.

6.3.1.1.7 Phenytoin – *HLA-A*24:02* – SJS/TEN

Phenytoin is mainly used in the treatment of seizures (89-91). Treatment is associated with SJS/TEN (91), and there is some evidence that patients with *HLA-A*24:02* variants are at increased risk of this ADR (92).

Searching Medline and Google Scholar, I did not locate any RCT or meta-analysis data on the link between *HLA-A*24:02* and phenytoin-induced SJS/TEN. Results were instead combined from three case population studies in Spain (93), South Korea (94), and Germany (95), that examined SJS/TEN cases in patients newly prescribed phenytoin (Table 6.11). None of these papers genotyped participants.

Rodríguez-Martín et al 2019 (93)			
Outcome	n	Risk of outcome	95% CI
SJS/ New users of phenytoin	10162	7/10162 = 0.0007	0.0003 – 0.0010
Chung et al 2020 (94)			

SJS/ New users of phenytoin	50978	51/50978 = 0.0010	0.0008 – 0.0013
Mockenhaupt et al 2005 (95)			
SJS/ New users of phenytoin	36171	30/36171 = 0.0008	0.0006 – 0.0011

Table 6.11 – Details of SJS/TEN in previous trials of patients receiving phenytoin.

The combined risk of SJS/TEN associated with phenytoin use (the risk in patients without genetic testing) was 0.0009 (95% CI 0.0007 – 0.0011) (Figure 6.11).

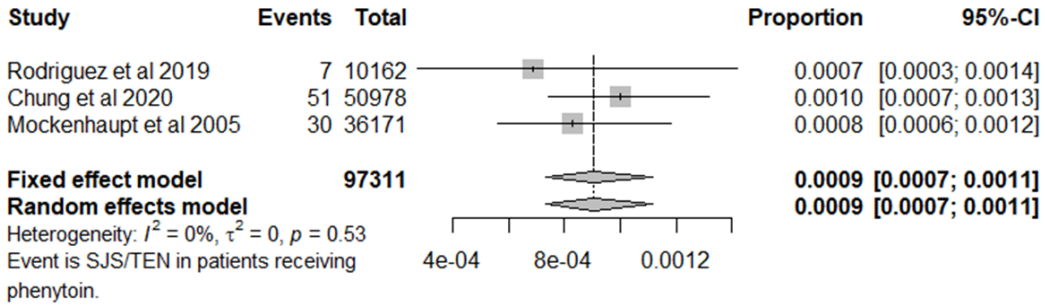


Figure 6.11 - Proportion meta-analysis of phenytoin and SJS/TEN

Only one paper was located that explored the risk of SJS/TEN in phenytoin users with respect to *HLA-A*24:02* (92). This was a case-control study that enrolled 91 SJS/TEN patients (13 induced by phenytoin) and 322 drug-tolerant controls (40 to phenytoin). Out of the phenytoin-induced SJS/TEN cases evaluated for *HLA-A*24:02*, 6 were positive (46.2%), compared to 5/40 (12.5%) in the drug-tolerant control group (OR 6.0, 95% CI 1.42 – 25.27, p = 0.027).

To produce an estimate of the risk of SJS/TEN when genetic testing is used, this information was combined with details of the risk of *HLA-A*24:02* in the population. This risk is assumed to be the same as the risk of ADR in non-carriers of *HLA-A*24:02*, since this population would still receive phenytoin after a genetic test.

Where;

$a = probability\ of\ ADR\ in\ HLAA2402\ carriers$

$b = probability\ of\ ADR\ in\ non\ carriers$

The OR from the case-control study was 6.0 (92), which can be used as a risk estimate since the outcome is rare (96). Then:

$$a/b = 6$$

The frequency of *HLA-A*24:02* was calculated from datasets held in the allelefrequencies.net database (97). Two English datasets gave the frequency as 13.8% and 12.3%. The median of these (13.05%) was used as the frequency of *HLA-A*24:02*. The frequency of non-carriers is therefore 86.95%.

Therefore;

$$(Frequency\ of\ carriers * probability\ of\ ADR\ in\ risk\ allele\ carriers) + (Frequency\ of\ noncarriers * probability\ of\ ADR\ in\ non\ carriers) = Probability\ of\ ADR\ in\ population$$

$$(0.1305 * a) + (0.8695 * b) = 0.0009$$

Solving these equations (using $a/b = 6$) gives the probability of ADR in non-carriers as 0.000545 (95% CI 0.000168 – 0.001043). This is the risk of SJS/TEN with genetic testing. Confidence intervals were obtained by performing the same analysis with the highest and lowest values given for the risk in users of phenytoin and the relative risk in *HLA-A*24:02* carriers.

6.3.1.1.8 Atorvastatin – *SLCO1B1* – Muscle ADRs

Statins are associated with several muscular ADRs, including pain, aching, stiffness, and more serious ones such as weakness, myopathy, and rhabdomyolysis (98, 99). Statin-induced rhabdomyolysis led to the withdrawal of cerivastatin in 2001 (100). Reporting of muscle symptoms is highly heterogeneous (98, 99, 101). *SLCO1B1* encodes the OATP1B1 protein, responsible for statin transport into hepatocytes (102). While the mechanism of statin-induced myopathy is not clear, the presence of *SLCO1B1* variants increases the risk of these ADRs (103). Genotyping before prescribing statins is not widely practiced (104, 105).

Clinically diagnosed myopathy (biochemical abnormality like rhabdomyolysis or necrotising myositis) was chosen as the atorvastatin ADR. While patient-reported myalgia is a more common ADR, a large proportion of myalgia in statin patients is not due to the statin (106-108). One 2017 meta-analysis found no difference in rates of muscle symptoms between patients on statins and those on placebo (OR 1.2, 95% CI 0.88 – 1.62, p = 0.25) (108).

After exploring the literature, very little data for atorvastatin alone was located. The studies that were located were heterogeneous in their results (109-112). For this reason, it was decided to base the estimate upon a large systematic review and meta-analysis that included 94,283 patients across a mostly Caucasian population (113). This study found 1938 cases of myopathy over 120,094 person-years in patients receiving placebos. The rate of myopathy (individual trials' own definitions were used) was therefore 0.016 (95% CI 0.015 – 0.017) over one year. However, since a pharmacogenetic test would be used to inform the first prescription of a statin, then an approximation of the probability of myopathy in one month (assuming constant risk) would be:

$$P\_myopathy = 1 - \exp\ \ [\![-0.16\ /12]\!]\ \ = 0.00134$$

The authors calculated a relative risk of myopathy in statin users of 1.08 (95% CI 1.01 – 1.15) compared to placebo. The risk in statin users, over one month, is therefore 0.00134*1.08 = 0.00145. This is the risk of ADR without genetic testing. The 95% confidence intervals were calculated using the confidence intervals for each estimate, producing a final estimate of the risk of ADR without genetic testing of 0.00145 (0.00126 – 0.00163).

For the risk of ADR with genetic testing, data from a 2019 UK meta-analysis was used, which found a non-significant increase in the risk of severe myopathy with atorvastatin in patients with a *SLCO1B1* risk allele (rs4149056, T521C) compared to those with *SCLO1B1* WT (OR 1.49, 95% CI 0.79 – 2.84, p = 0.2133) (103).

The same methods as for phenytoin were used to calculate the probability of ADR with genetic testing (the risk of ADR in non-carriers). This risk is assumed to be the same as the risk of ADR in non-carriers of *SLCO1B1*, since this population would still receive atorvastatin after a genetic test.

Where;

$a\ = probability\ of\ ADR\ in\ SLCO1B1\ risk\ allele\ carriers$

$b = probability\ of\ ADR\ in\ non\ carriers$

The risk ratio from the previous systematic review and meta-analysis was 1.08 (1.01 – 1.15) (113). Then:

$a/b = 1.08$

The frequency of *SLCO1B1* was calculated from data contained within the ALFA Allele Frequency project (114). The frequency of the *SLCO1B1* rs4149056 risk variant in a European population was estimated as 0.1586 (the frequency of non-carriers is therefore 0.8414). No confidence intervals were provided. This data is based on data from 251,826 genomes (115).

The increased risk of myopathy with the *SLCO1B1* risk allele was found to be 1.49 (0.79 – 2.84) in a UK meta-analysis (103).

Therefore;

$$(Frequency\ of\ carriers * probability\ of\ ADR\ in\ SLCO1B1\ risk\ allele\ carriers) + (Frequency\ of\ noncarriers * probability\ of\ ADR\ in\ non\ carriers) = Probability\ of\ ADR\ in\ population$$

$$(0.1586 * a) + (0.8414 * b) = 0.00145$$

Solving these equations gives probability of ADR in non-carriers (therefore the probability of myopathy in statin users, with genetic testing) as 0.001301 (95% CI 0.000956 – 0.001637). Confidence intervals for these numbers were obtained by performing the same analysis with the highest and lowest values given for the risk in users of statins and the relative risk in *SLCO1B1* risk allele carriers.

This is arguably the drug-gene-ADR combination with the least evidence for its use. The link between *SCLO1B1* and statin-induced myopathy is not consistent (98, 111, 113, 116-121). These estimates also differ from those produced by a European Atherosclerosis Society Consensus Panel, which noted a probability of statin-associated myopathy in statin users of 1 per 1000 to 1 per 10,000 people (98).

6.3.1.1.9 Summary of risk estimates

Through various methods, estimates of the risk of each ADR associated with each drug were obtained, with and without the implementation of genetic testing. The estimates for the ‘high evidence’ combinations are likely to be more accurate since they rely on more and higher quality evidence than the estimates in the ‘low evidence’ combinations.

Frequencies were also compared to those stated in the summary product characteristics (SPc) for each drug, where available. These use common terminology to describe the frequency of events; common (> 1/100 to < 1/10); uncommon (> 1/1,000 to < 1/100); rare (> 1/10,000 to < 1/1,000); very rare (< 1/10,000) (122). These were used to describe the frequencies calculated (Table 6.12).

Combination	Risk with testing (95% CI)	Risk without testing (95% CI)	SPc prevalence	Lay frequency of my estimate	Source(s)
High evidence					
Abacavir - *HLA-B*57:01* - AHS	0.034 (0.023 – 0.048)	0.078 (0.062 – 0.098)	Not stated	Common	RCT of genotyping vs not genotyping (52)
Capecitabine – *DPYD* - neutropenia	0.082 (0.070 – 0.096)	0.0854	Common (1-10%)	Common	Prospective study with genotyping (60)
Carbamazepine – *HLA-A*31:01* – SJS/TEN	0.000087	0.000118 1 in 11,800	Uncommon (0.1 – 1%)	Very rare – with test Rare - without test	Economic analysis (61)
Warfarin – *CYP2C9/VKORC1* – bleeding	0.0020 (0.0007 – 0.0090)	0.010 (0.005 – 0.020)	Not stated	Uncommon	GIFT RCT of genotyping vs not genotyping (73)
Low evidence					
Efavirenz – *CYP2B6* – DILI	0.07 (0.03 – 0.13)	0.10 (0.07 – 0.15)	Not DILI specific	Common – with test Very common – with test	Meta-analysis (3 x cohort studies) (79, 81, 82)
Irinotecan – *UGT1A1* - neutropenia	0.164 (0.149 – 0.179)	0.182 (0.172 – 0.194)	Very common (>10%)	Very common	Meta-analysis (2 x MA) and data from 1 of these on genotype (85, 88)
Phenytoin – *HLA-A*24:02* – SJS/TEN	0.0005 (0.0002 – 0.0010	0.0009 (0.0007 – 0.0011)	Not stated	Rare	Meta-analysis (3 x population studies) (94, 95, 123) plus calculations

Atorvastatin – *SLCO1B1* – muscle ADRs	0.001301 (0.000956 – 0.001637)	0.001439 (0.001262 – 0.001628)	Rare (0.001 – 0.1%)	Uncommon	2 meta-analyses (103, 113) plus calculations

Table 6.12 - Summary of risk estimates of each ADR for each drug, with and without genetic testing. This is compared to the stated prevalence of each ADR in the SPc for each drug. The lay frequency of my estimate is stated. ADR = adverse drug reaction. AHS = abacavir hypersensitivity syndrome. DILI = drug induced liver injury. MA = meta-analysis. RCT = randomised controlled trial. SJS/TEN = Stevens Johnson syndrome/toxic epidermal necrolysis. SPc = summary of product characteristics.

These estimates were used to produce a range of levels for the risk of each ADR with and without genetic testing. I wanted to use the same levels across the same indications, i.e., both HIV indications would have the same levels. The calculated risk estimates provide a suitable range for levels across indications. For example, abacavir and efavirenz risk estimates range from 0.023 – 0.15 (including confidence intervals), so levels of 0.03, 0.05, and 0.15 were chosen. Using the risk estimates above, three levels for each drug-gene-ADR combination were produced (Table 6.13).

Drug – gene – ADR	Levels– risk of ADR from this drug		
	Level 2	**Level 1**	**Level 0**
High evidence			
Abacavir - *HLA-B*57:01* – AHS	0.03 (3 in 100)	0.05 (5 in 100)	0.15 (15 in 100)
Capecitabine – *DPYD* - neutropenia	0.02 (2 in 100)	0.1 (10 in 100)	0.2 (20 in 100)
Carbamazepine – *HLA-A*31:01* – SJS/TEN	Very rare (less than 1 in 10,000)	Rare (less than 1 in 1000)	Uncommon (less than 1 in 100)
Warfarin – *CYP2C9/VKORC1* – bleeding	0.001 (1 in 1000) Rare (less than 1 in 1,000)	0.005 (5 in 1000) Uncommon (less than 1 in 100)	0.01 (10 in 1000) Common (less than 1 in 10)
Low evidence			
Efavirenz – *CYP2B6* – DILI	0.03 (3 in 100)	0.05 (5 in 100)	0.15 (15 in 100)
Irinotecan – *UGT1A1* - neutropenia	0.02 (2 in 100)	0.1 (10 in 100)	0.2 (20 in 100)

Phenytoin – *HLA-A*24:02* – SJS/TEN	Very rare (less than 1 in 10,000)	Rare (less than 1 in 1000)	Uncommon (less than 1 in 100)
Atorvastatin – *SLCO1B1* – muscle ADRs	0.001 (1 in 1000) Rare (less than 1 in 1,000)	0.005 (5 in 1000) Uncommon (less than 1 in 100)	0.01 (10 in 1000) Common (less than 1 in 10)

Table 6.13 – Levels produced from risk estimates of the risk of each ADR with and without genetic testing, for each drug-gene-ADR combination. ADR = adverse drug reaction. AHS = abacavir hypersensitivity reaction. DILI = drug induced liver injury. SJS/TEN = Stevens Johnson syndrome/toxic epidermal necrolysis.

6.3.1.2 Other attributes

Other attributes and their levels were drawn from the qualitative work and the literature.

6.3.1.2.1 Use of your data for further research by universities and researchers

The focus group and survey findings indicated the importance of privacy to participants. In the focus group especially, privacy was the primary concern regarding genetic testing. This concern has also been found by other researchers (10, 11).

For this reason, an attribute and levels were designed that would capture privacy and include it in the DCE. A three-tier situation was proposed, where participants could choose to link their genetic data to their medical records, where they could have their genetic data used anonymously, or decline to have their data used (apart from as the results of their genetic test for ADR prediction). Anonymisation was the most common requirement of respondents in a recent survey of health data sharing (124). In this scenario, this data would only be used by universities and researchers. This more fully captures the balance between altruism and privacy reflected in the focus group discussions. In the 'no test' option, there is no use of data for further research.

Attribute	Levels		
	Level 2	**Level 1**	**Level 0**
Use of your data for further research by universities and researchers	Yes, and they can contact me (linked to medical record)	Yes, but no contact (anonymous)	No

6.3.1.2.2 Number of medicines the test can be used to inform

This attribute was chosen to reflect the possibility of genetic panel testing in the future. Pilot testing revealed that this was very important to participants. Levels were chosen based on consultation with experts about potential future panels [personal communication, Dyfrig Hughes]. In the 'no test' option, the test cannot inform any medicines.

Attribute	Levels		
	Level 2	**Level 1**	**Level 0**
Number of medicines the test can be used to inform	50	25	1

6.3.1.2.3 Cost of the test to the NHS

Although cost was not rated as hugely important by participants in the qualitative work, it was decided to include it as an attribute to allow the future calculation of willingness to pay and the estimation of trade-offs for the other attributes (19, 22). This attribute was framed as 'the cost to the NHS', since this is the model UK participants would be familiar with. The levels for this attribute were based on the prices given by Illumina for a global screening array kit (as of December 2020) (44).

Attribute	Levels		
	Level 2	**Level 1**	**Level 0**
Cost of the test to the NHS	50	30	10

6.3.1.2.4 Risk of serious side-effect from any medicine over the next 10 years, excluding this medicine

This attribute also reflects the utility of genetic panel testing vs single gene testing. The levels were chosen according to a 2019 paper that calculated the number of drugs with pharmacogenetic guidelines a UK primary care patient might expect to

be prescribed (125). They predicted a median of 2 pharmacogenetic drugs per patient over a 10-year period. After consulting expert opinion based on this paper, the average risk of avoidable ADRs across all pharmacogenetic drugs was assumed to be 0.05 (1 in 20). The top of the interquartile range for the number of drugs a patient might be prescribed was 3. This would give a risk of an ADR of 1 in 6. A final value of 1 in 5 (0.2) as a level as this is easier to visualise. For the final level for this attribute, 1 in 50 (0.02) was chosen.

Attribute	Levels		
	Level 2	**Level 1**	**Level 0**
Risk of serious side-effect from any medicine over the next 10 years, excluding this medicine	1 in 5 (0.2)	1 in 20 (0.05)	1 in 50 (0.02)

6.3.1.3 Final attributes and levels

The final DCE attributes and levels are shown in Table 6.14. The levels for the 'chance of serious side-effect from this medicine' attribute for each drug-gene-ADR combination were previously shown in Table 6.13.

Attribute	Levels	Notes
Chance of serious side-effect from this medicine		Vary by drug-gene-ADR combination. See above for calculations
Cost of the test to the NHS	£10	The Illumina prices for a chip with 650,000 markers on it (44). They work out as £40 a sample. No test is £0.
	£30	
	£50	
Use of your data for further research by universities and researchers	Yes, and they can contact me (linked to medical record)	Privacy attribute important to focus group. No test would correspond to 'no'.
	Yes, but no contact (anonymous)	

	No	
Number of medicines the test can be used to inform	1 (corresponds to single gene test)	These won't be marked as single gene or panel test. These will be categorical. No test is 0.
	25 (corresponds to smaller panel test)	
	50 (corresponds to panel test)	
Risk of serious side-effect from any medicine over the next 10 years *excluding this drug*	1 in 5	Displayed as '20 in 100' '5 in 100' '2 in 100'
	1 in 20	
	1 in 50	

Table 6.14 - Final attributes and levels in the DCE.

From the qualitative work, the following order of preferences was hypothesised for each attribute:

- Chance of serious side-effect from this medicine: lowest chance > highest chance
- Cost of test to the NHS: £10 > £30 > £50
- Use of your data for further research by universities and researchers: yes, but no contact > no > yes, and they can contact me (linked to medical record)
- Number of medicines the test can be used to inform: 50 > 25 > 1
- Risk of serious side-effect from any medicine over the next 10 years excluding this drug: 1 in 50 > 1 in 20 > 1 in 5

6.3.2 Sample size

Sample size calculation in discrete choice experiments is made difficult when the strength and direction of preferences are not known (126, 127). The 'rule of thumb' by Johnson and Orme states that the sample size (N) should be:

$$N > \frac{500c}{t * a}$$

Where c is the largest number of levels for any of the attributes, t is the number of choice tasks, and a is the number of alternatives (excluding the no choice option) (126, 127). Using this equation, the proposed DCE design leads to a minimum of 50 participants per DCE:

$$N > \frac{500 * 3}{15 * 2}$$

$$N > 50$$

However, one source recommends a sample size of 200 participants per group, to be able to compare across groups. These decisions also have to be weighted against costs and participant time (126). These equations also assume a large or infinite potential population (i.e., the 'product' is not a niche that only a small population would make choices about) (126). I chose to include 2000 participants, randomly allocated between all 8 surveys. This will give a sufficient sample size to compare across groups and exceeds the 'rule of thumb' minimum.

There is a fine balance required between obtaining maximum possible data, and overworking respondents. Participants are known to be more likely to make choices at random as the cognitive burden (which may include length of survey) increases (20, 128). One source recommends using the formula:

$$3(K - k + 1)$$

where K is the total number of levels across all attributes and k is the number of attributes (126). Using this formula, the minimum number of questions should be 33 per participant:

$$3(15 - 5 + 1) = 33$$

This is where one needs to mindful of participant burden, and the increased cost of reimbursing participants for their time. I therefore chose to design the survey with 15 questions per participant. This strikes an appropriate balance between data collection and cognitive burden for participants. Pilot testing of one DCE located in the Chapter 5 systematic review found that participants could complete up to 15 choice tasks "easily" (129).

6.3.3 Final DCE design

Based on the qualitative work, pictograms were chosen to represent risk. This was the option most often chosen by focus group participants, and was also popular in the survey of patients.

I chose to include a 'no test' option since this most accurately reflects the clinical reality of genetic testing. This is important for the accuracy of conclusions drawn from DCE data [130, 131]. A genetic test is unlikely to be mandatory, even when screening is strongly recommended.

The final DCE design was generated using Ngene software (132). Three levels in five attributes would produce a total of 3^5 (243) total scenarios. This was reduced to a fractional factorial design using a D-efficient multinomial logit (MNL) design. An MNL model assumes that the unobserved component (preferences) is uncorrelated across choices and individuals (133). The design was developed with constraints on which levels could appear together. For example, the cost of the test to the NHS could not be £10 where the test could be used to inform 50 medicines.

Full surveys as shown to participants are located in Appendix 6.3.

6.4 Discussion

I have completed a comprehensive program of quantitative and qualitative work in the development of this DCE. By surveying several different groups (healthcare professionals, patients, and the general public), I have gained valuable perspectives into problems encountered in the area of genetic testing. This qualitative work informed attribute and level selection, and I can be confident that these, as far as possible, represent the issues involved in choosing genetic testing. The robust methods used to calculate risk estimates are useful on their own, but also ensure that each DCE is realistic and relevant in its field.

An 'opt out' option has been included since this most accurately reflects the clinical reality of genetic testing. In a German study of patient and physician opinions of pharmacogenetic testing for asthma, 95.9% of patients would accept testing (134). A similarly high percentage of patients (94.4%) thought it valuable to know their own genetic disposition. Patients were hopeful that pharmacogenetics could aid in finding the correct drug, dosage, and minimise side effects. I hypothesise that similar rates of uptake will be found in the results of this DCE.

I knew that I wished to focus on level of evidence in this DCE, but I was unsure if patients would use this as a factor in decision making. Qualitative work found that the level of evidence for a genetic test was one of the most important attributes to patients and healthcare professionals. It was less important to focus group participants, but running further focus groups with more discussion focussed on this would have potentially yielded more useful insights.

By making level of evidence an intrinsic part of the survey, with randomisation, having to explain the difference between 'high' and 'low' evidence to participants is avoided. This will decrease the cognitive burden to each participant. Instead, by randomising patients between DCEs, responses between 'high' and 'low' evidence scenarios in the same therapeutic area can be compared. Participants will receive slightly different scenarios depending on which evidence level they are randomised to. In a 'high' evidence scenario, they will be told that the genetic test is recommended by regulatory authorities. In the 'low' evidence scenario, they will be told that the test is not currently widely recommended for use. It will be interesting to see how participants value tests with varying levels of evidence. I hypothesise that participants will be willing to accept higher costs if a test has high levels of evidence.

6.4.1 Limitations

Although my qualitative work included different populations and methods, it suffered from small sample sizes. This was limited by resources, and later, the COVID-19 pandemic. Although I believe that the existing qualitative work is sufficient for this current DCE, larger sample sizes and more extensive testing would have further improved the DCE design.

The limited time available in a PhD project did not allow me to perform full systematic reviews and meta-analyses of the evidence behind each gene-drug-ADR combination. There will therefore be some error in these estimates. I believe the mini-reviews conducted were sufficient for my purposes, but if this project is to be replicated, full reviews could be conducted to further refine the estimate. Further, although I tried to only include data from UK or European populations (since the final survey will mostly be in this group), this was not always possible. Some estimates may therefore be less relevant. This will also be a limitation of the final analysis of utility.

If a patient undergoes a pharmacogenetic test that results in a recommendation for them to be prescribed an alternative drug, clinicians may also have to consider that the alternative drug has its own associated ADRs. This concept was not specifically

addressed in the DCE design, to avoid further participant burden. This scenario may be useful for further research and could inform policy around incidental findings in genetic testing.

Including the 'cost to the NHS' attribute limits the generalisability of the DCE. It may be difficult to apply the results in places with different models of healthcare. However, the relevance of the results in the UK is enhanced, with the potential for large impacts on genetic testing policy within the NHS. It also allows for possible future calculation of WTP, which is an important outcome for policy makers (21, 23).

Finally, only a section of the survey was pilot tested, rather than the full survey. Although the final survey was broadly easy to complete, this may have been improved through the use of pilot testing and informational interviews.

6.5 Conclusion

DCEs are a powerful method for quantifying preferences. However, they are complex to design and implement, and require extensive qualitative work to ensure their relevance. I have conducted several different types of qualitative work, including a systematic review (Chapter 5), in order to inform the DCE. The final design was informed by this work, and offers a highly relevant and effective way to measure the preferences of the general public for genetic testing in the prevention of ADRs. The implementation and results of the DCE are presented in Chapter 7.

6.6 References

1. Evans D. Patient and public involvement in research in the English NHS: a documentary analysis of the complex interplay of evidence and policy. Evidence & Policy: A Journal of Research, Debate and Practice. 2014;10(3):361-77.
2. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? Expert Rev Pharmacoecon Outcomes Res. 2017;17(6):531-42.
3. Medical Device Innovation Consortium. A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology 2015 [Available from: http://mdic.org/wp-content/uploads/2015/05/MDIC_PCBR_Framework_Web1.pdf.
4. Epstein RM, Street RL, Jr. The values and value of patient-centered care. Ann Fam Med. 2011;9(2):100-3.

5. Barwell J, Snape K, Wedderburn S. The new genomic medicine service and implications for patients. Clinical medicine (London, England). 2019;19(4):273-7.
6. Johnson SB, Slade I, Giubilini A, Graham M. Rethinking the ethical principles of genomic medicine services. Eur J Hum Genet. 2020;28(2):147-54.
7. Caudle KE, Klein TE, Hoffman JM, Muller DJ, Whirl-Carrillo M, Gong L, et al. Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process. Current drug metabolism. 2014;15(2):209-17.
8. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ. 2004;329(7456):15-9.
9. Veeren JC, Weiss M. Trends in emergency hospital admissions in England due to adverse drug reactions: 2008-2015. Journal of Pharmaceutical Health Services Research. 2017;8(1):5-11.
10. Issa AM, Tufail W, Atehortua N, McKeever J. A national study of breast and colorectal cancer patients' decision-making for novel personalized medicine genomic diagnostics. Per Med. 2013;10(3):245-56.
11. Faulkner E, Holtorf AP, Walton S, Liu CY, Lin H, Biltaj E, et al. Being Precise About Precision Medicine: What Should Value Frameworks Incorporate to Address Precision Medicine? A Report of the Personalized Precision Medicine Special Interest Group. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2020;23(5):529-39.
12. Ryan M. Discrete choice experiments in health care. BMJ. 2004;328(7436):360-1.
13. Lancsar E, Fiebig DG, Hole AR. Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software. Pharmacoeconomics. 2017;35(7):697-716.
14. Bryan S, Dolan P. Discrete choice experiments in health economics. For better or for worse? Eur J Health Econ. 2004;5(3):199-202.
15. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics. 2008;26(8):661-77.
16. Bridges JFP, Kinter ET, Kidane L, Heinzen RR, McCormick C. Things are Looking up Since We Started Listening to Patients. The Patient: Patient-Centered Outcomes Research. 2008;1(4):273-82.
17. Food and Drug Administration (US). Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian

Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling: Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders 2016 [Available from: https://www.fda.gov/media/92593/download.

18. Powell G, Holmes EA, Plumpton CO, Ring A, Baker GA, Jacoby A, et al. Pharmacogenetic testing prior to carbamazepine treatment of epilepsy: patients' and physicians' preferences for testing and service delivery. British journal of clinical pharmacology. 2015;80(5):1149-59.

19. Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Muhlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013;16(1):3-13.

20. Ryan M, Gerard K, Amaya-Amaya M. Discrete Choice Experiments in a Nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care. The Economics of Non-Market Goods and Resources. Dordrecht: Springer Netherlands; 2008. p. 13-46.

21. Ryan M, Gerard K, Amaya-Amaya M. Using discrete choice experiments to value health and health care: Springer Science & Business Media; 2007.

22. Ryan M, Kolstad JR, Rockers PC, Dolea C. How to conduct a discrete choice experiment for health workforce recruitment and retention in remote and rural areas: a user guide with case studies. The World Bank; 2012.

23. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883-902.

24. Coast J, Al-Janabi H, Sutton EJ, Horrocks SA, Vosper AJ, Swancutt DR, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. Health Econ. 2012;21(6):730-41.

25. Crocker JC, Ricci-Cabello I, Parker A, Hirst JA, Chant A, Petit-Zeman S, et al. Impact of patient and public involvement on enrolment and retention in clinical trials: systematic review and meta-analysis. BMJ. 2018;363:k4738.

26. Street DJ, Burgess L, Viney R, Louviere J. Designing Discrete Choice Experiments for Health Care. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care. The Economics of Non-Market Goods and Resources. Dordrecht: Springer Netherlands; 2008. p. 47-72.

27. Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. Journal of health services research & policy. 2007;12(1):25-30.
28. Louviere JJ. What You Don't Know Might Hurt You: Some Unresolved Issues in the Design and Analysis of Discrete Choice Experiments. Environmental and Resource Economics. 2006;34(1):173-88.
29. Louviere JJ, Hensher DA, Swait JD. Stated choice methods: analysis and applications: Cambridge university press; 2000.
30. Vass C, Rigby D, Payne K. The Role of Qualitative Research Methods in Discrete Choice Experiments. Medical decision making : an international journal of the Society for Medical Decision Making. 2017;37(3):298-313.
31. Kløjgaard ME, Bech M, Søgaard R. Designing a Stated Choice Experiment: The Value of a Qualitative Process. Journal of Choice Modelling. 2012;5(2):1-18.
32. SurveyMonkey Inc. SurveyMonkey San Mateo, California, USA2021 [Available from: www.surveymonkey.com.
33. Johnson D, Jorgensen A, Hughes D, Pirmohamed SM, editors. A discrete choice experiment to measure public preferences for genetic testing to prevent adverse drug reactions. TRIALS; 2019: BMC CAMPUS, 4 CRINAN ST, LONDON N1 9XW, ENGLAND.
34. National Institute for Health Research U. Briefing notes for researchers - public involvement in NHS, health and social care research 2021 [Available from: https://www.nihr.ac.uk/documents/briefing-notes-for-researchers-public-involvement-in-nhs-health-and-social-care-research/27371.
35. Frost J, Gibson A, Harris-Golesworthy F, Harris J, Britten N. Patient involvement in qualitative data analysis in a trial of a patient-centred intervention: Reconciling lay knowledge and scientific method. Health expectations : an international journal of public participation in health care and health policy. 2018;21(6):1111-21.
36. Morris D. A Patient Led Project: Building a Research Community 2017 [Available from: https://www.nwcscnsenate.nhs.uk/files/6915/0600/7280/NWC_Clinical_Research_Network_project_pack.pdf.
37. Guest G, Namey E, Taylor J, Eley N, McKenna K. Comparing focus groups and individual interviews: findings from a randomized study. International Journal of Social Research Methodology. 2017;20(6):693-708.

38. Onwuegbuzie AJ, Dickinson WB, Leech NL, Zoran AG. A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research. International Journal of Qualitative Methods. 2009;8(3):1-21.
39. Ltd. QIP. NVivo qualitative data analysis software. 12 ed2018.
40. Issa AM, Hutchinson JF, Tufail W, Fletcher E, Ajike R, Tenorio J. Provision of personalized genomic diagnostic technologies for breast and colorectal cancer: an analysis of patient needs, expectations and priorities. Per Med. 2011;8(4):401-11.
41. Herbild L. Designing a DCE to outlay patients' and the publics' preferences for genetic screening in the treatment of depression. 2007.
42. Bombard Y, Rozmovits L, Trudeau ME, Leighl NB, Deal K, Marshall DA. Patients' perceptions of gene expression profiling in breast cancer treatment decisions. Current oncology (Toronto, Ont). 2014;21(2):e203-11.
43. Bombard Y, Rozmovits L, Trudeau M, Leighl NB, Deal K, Marshall DA. Access to personalized medicine: factors influencing the use and value of gene expression profiling in breast cancer treatment. Current oncology (Toronto, Ont). 2014;21(3):e426-33.
44. Illumina. Infinium Global Screening Array-24 Kit 2020 [Available from: https://emea.illumina.com/products/by-type/microarray-kits/infinium-global-screening.html?langsel=/gb/.
45. Barbarino JM, Whirl-Carrillo M, Altman RB, Klein TE. PharmGKB: A worldwide resource for pharmacogenomic information. Wiley Interdiscip Rev Syst Biol Med. 2018;10(4):e1417.
46. PharmGKB. Clinical Annotation Levels of Evidence [Available from: https://www.pharmgkb.org/page/clinAnnLevels.
47. DescTools. BinomCI: Confidence Intervals for Binomial Proportions. 0.99.42 ed.
48. Shapiro M, Ward KM, Stern JJ. A near-fatal hypersensitivity reaction to abacavir: case report and literature review. The AIDS reader. 2001;11(4):222-6.
49. Clay PG. The abacavir hypersensitivity reaction: a review. Clin Ther. 2002;24(10):1502-14.
50. Hetherington S, McGuirk S, Powell G, Cutrell A, Naderer O, Spreen B, et al. Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. Clin Ther. 2001;23(10):1603-14.
51. Sousa-Pinto B, Pinto-Ramos J, Correia C, Goncalves-Costa G, Gomes L, Gil-Mata S, et al. Pharmacogenetics of abacavir hypersensitivity: A systematic

review and meta-analysis of the association with HLA-B*57:01. J Allergy Clin Immunol. 2015;136(4):1092-4 e3.
52. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, et al. HLA-B*5701 screening for hypersensitivity to abacavir. The New England journal of medicine. 2008;358(6):568-79.
53. Bodey GP, Buckley M, Sathe YS, Freireich EJ. Quantitative relationships between circulating leukocytes and infection in patients with acute leukemia. Annals of internal medicine. 1966;64(2):328-40.
54. Caggiano V, Weiss RV, Rickert TS, Linde-Zwirble WT. Incidence, cost, and mortality of neutropenia hospitalization associated with chemotherapy. Cancer-Am Cancer Soc. 2005;103(9):1916-24.
55. Chen J, Wang J, Xu T. Comparison of efficacy and safety of S-1 and capecitabine in patients with metastatic colorectal carcinoma: A systematic review and meta-analysis. Medicine. 2019;98(30):e16667.
56. Lu Y, Zhao Z, Wang J, Lv W, Lu L, Fu W, et al. Safety and efficacy of combining capecitabine and temozolomide (CAPTEM) to treat advanced neuroendocrine neoplasms: A meta-analysis. Medicine. 2018;97(41):e12784.
57. Madi A, Fisher D, Maughan TS, Colley JP, Meade AM, Maynard J, et al. Pharmacogenetic analyses of 2183 patients with advanced colorectal cancer; potential role for common dihydropyrimidine dehydrogenase variants in toxicity to chemotherapy. European journal of cancer (Oxford, England : 1990). 2018;102:31-9.
58. Dean L, Kane M. Capecitabine Therapy and DPYD Genotype. In: Pratt VM, Scott SA, Pirmohamed M, Esquivel B, Kane MS, Kattman BL, et al., editors. Medical Genetics Summaries. Bethesda (MD): National Center for Biotechnology Information (US); 2012.
59. Caudle KE, Thorn CF, Klein TE, Swen JJ, McLeod HL, Diasio RB, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. Clinical pharmacology and therapeutics. 2013;94(6):640-5.
60. Deenen MJ, Meulendijks D, Cats A, Sechterberger MK, Severens JL, Boot H, et al. Upfront Genotyping of DPYD*2A to Individualize Fluoropyrimidine Therapy: A Safety and Cost Analysis. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2016;34(3):227-34.
61. Plumpton CO, Yip VL, Alfirevic A, Marson AG, Pirmohamed M, Hughes DA. Cost-effectiveness of screening for HLA-A*31:01 prior to initiation of carbamazepine in epilepsy. Epilepsia. 2015;56(4):556-63.

62. Marson AG, Al-Kharusi AM, Alwaidh M, Appleton R, Baker GA, Chadwick DW, et al. The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial. Lancet. 2007;369(9566):1000-15.
63. McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperaviciute D, Carrington M, et al. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. The New England journal of medicine. 2011;364(12):1134-43.
64. Mushiroda T, Takahashi Y, Onuma T, Yamamoto Y, Kamei T, Hoshida T, et al. Association of HLA-A*31:01 Screening With the Incidence of Carbamazepine-Induced Cutaneous Adverse Reactions in a Japanese Population. JAMA Neurol. 2018;75(7):842-9.
65. Hylek EM. Complications of oral anticoagulant therapy: bleeding and nonbleeding, rates and risk factors. Semin Vasc Med. 2003;3(3):271-8.
66. Johnson JA, Cavallari LH. Warfarin pharmacogenetics. Trends Cardiovasc Med. 2015;25(1):33-41.
67. Pirmohamed M. Warfarin: almost 60 years old and still causing problems. British journal of clinical pharmacology. 2006;62(5):509-11.
68. Pirmohamed M, Burnside G, Eriksson N, Jorgensen AL, Toh CH, Nicholson T, et al. A randomized trial of genotype-guided dosing of warfarin. The New England journal of medicine. 2013;369(24):2294-303.
69. Anderson JL, Horne BD, Stevens SM, Woller SC, Samuelson KM, Mansfield JW, et al. A randomized and clinical effectiveness trial comparing two pharmacogenetic algorithms and standard care for individualizing warfarin dosing (CoumaGen-II). Circulation. 2012;125(16):1997-2005.
70. Jorgensen AL, FitzGerald RJ, Oyee J, Pirmohamed M, Williamson PR. Influence of CYP2C9 and VKORC1 on patient response to warfarin: a systematic review and meta-analysis. PLoS One. 2012;7(8):e44064.
71. Kimmel SE, French B, Kasner SE, Johnson JA, Anderson JL, Gage BF, et al. A pharmacogenetic versus a clinical algorithm for warfarin dosing. The New England journal of medicine. 2013;369(24):2283-93.
72. Caraco Y, Blotnick S, Muszkat M. CYP2C9 genotype-guided warfarin prescribing enhances the efficacy and safety of anticoagulation: a prospective randomized controlled study. Clinical pharmacology and therapeutics. 2008;83(3):460-70.
73. Gage BF BA, Lin H, Woller SC, Stevens SM, Al-Hammadi N, Li J, Rodríguez T, Miller JP, McMillin GA, Pendleton RC. Effect of Genotype-Guided Warfarin

Dosing on Clinical Events and Anticoagulation Control Among Patients Undergoing Hip or Knee Arthroplasty: The GIFT Randomized Clinical Trial. JAMA. 2017;318(12):1115-24.
74. David S, Hamilton JP. Drug-induced Liver Injury. US Gastroenterol Hepatol Rev. 2010;6:73-80.
75. Ostapowicz G, Fontana RJ, Schiodt FV, Larson A, Davern TJ, Han SH, et al. Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States. Annals of internal medicine. 2002;137(12):947-54.
76. Hoofnagle JH, Bjornsson ES. Drug-Induced Liver Injury - Types and Phenotypes. The New England journal of medicine. 2019;381(3):264-73.
77. Kullak-Ublick GA, Andrade RJ, Merz M, End P, Benesic A, Gerbes AL, et al. Drug-induced liver injury: recent advances in diagnosis and risk assessment. Gut. 2017;66(6):1154-64.
78. Desta Z, Gammal RS, Gong L, Whirl-Carrillo M, Gaur AH, Sukasem C, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2B6 and Efavirenz-Containing Antiretroviral Therapy. Clinical pharmacology and therapeutics. 2019;106(4):726-33.
79. Yimer G, Amogne W, Habtewold A, Makonnen E, Ueda N, Suda A, et al. High plasma efavirenz level and CYP2B6*6 are associated with efavirenz-based HAART-induced liver injury in the treatment of naive HIV patients from Ethiopia: a prospective cohort study. Pharmacogenomics J. 2012;12(6):499-506.
80. Benichou C. Criteria of drug-induced liver disorders. Report of an international consensus meeting. J Hepatol. 1990;11(2):272-6.
81. Yimer G, Gry M, Amogne W, Makonnen E, Habtewold A, Petros Z, et al. Evaluation of patterns of liver toxicity in patients on antiretroviral and anti-tuberculosis drugs: a prospective four arm observational study in ethiopian patients. PLoS One. 2014;9(4):e94271.
82. Mugusi S, Ngaimisi E, Janabi M, Minzi O, Bakari M, Riedel KD, et al. Liver enzyme abnormalities and associated risk factors in HIV patients on efavirenz-based HAART with or without tuberculosis co-infection in Tanzania. PLoS One. 2012;7(7):e40180.
83. Zhou Y, Ingelman-Sundberg M, Lauschke VM. Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. Clinical pharmacology and therapeutics. 2017;102(4):688-700.
84. Bailly C. Irinotecan: 25 years of cancer treatment. Pharmacol Res. 2019;148:104398.

85. Wulaningsih W, Wardhana A, Watkins J, Yoshuantari N, Repana D, Van Hemelrijck M. Irinotecan chemotherapy combined with fluoropyrimidines versus irinotecan alone for overall survival and progression-free survival in patients with advanced and/or metastatic colorectal cancer. The Cochrane database of systematic reviews. 2016;2(2):CD008593.
86. Steventon G. Uridine diphosphate glucuronosyltransferase 1A1. Xenobiotica. 2020;50(1):64-76.
87. Hoskins G, Williams B, Abhyankar P, Donnan P, Duncan E, Pinnock H, et al. Achieving Good Outcomes for Asthma Living (GOAL): mixed methods feasibility and pilot cluster randomised controlled trial of a practical intervention for eliciting, setting and achieving goals for adults with asthma. Trials. 2016;17(1):584.
88. Yang Y, Zhou M, Hu M, Cui Y, Zhong Q, Liang L, et al. UGT1A1*6 and UGT1A1*28 polymorphisms are correlated with irinotecan-induced toxicity: A meta-analysis. Asia Pac J Clin Oncol. 2018;14(5):e479-e89.
89. Gallop K. Review article: phenytoin use and efficacy in the ED. Emerg Med Australas. 2010;22(2):108-18.
90. Lonjou C, Borot N, Sekula P, Ledger N, Thomas L, Halevy S, et al. A European study of HLA-B in Stevens-Johnson syndrome and toxic epidermal necrolysis related to five high-risk drugs. Pharmacogenetics and genomics. 2008;18(2):99-107.
91. Thorn CF, Whirl-Carrillo M, Leeder JS, Klein TE, Altman RB. PharmGKB summary: phenytoin pathway. Pharmacogenetics and genomics. 2012;22(6):466-70.
92. Shi YW, Min FL, Zhou D, Qin B, Wang J, Hu FY, et al. HLA-A*24:02 as a common risk factor for antiepileptic drug-induced cutaneous adverse reactions. Neurology. 2017;88(23):2183-91.
93. Rodriguez-Novoa S, Garcia-Gasco P, Blanco F, Gonzalez-Pardo G, Castellares C, Moreno V, et al. Value of the HLA-B*5701 allele to predict abacavir hypersensitivity in Spaniards. AIDS Res Hum Retroviruses. 2007;23(11):1374-6.
94. Chung SJ, Ahn KM, Oh JH, Shim JS, Park HW. Incidence rates of severe cutaneous adverse reactions due to antiseizure medication: A nationwide study using health claims data in Korea. Epilepsia. 2021;62(1):250-7.
95. Mockenhaupt M, Messenheimer J, Tennis P, Schlingmann J. Risk of Stevens-Johnson syndrome and toxic epidermal necrolysis in new users of antiepileptics. Neurology. 2005;64(7):1134-8.
96. Sedgwick P. Relative risks versus odds ratios. Bmj. 2014;348(feb07 2):g1407-g.

97. Gonzalez-Galarza FF, McCabe A, Santos E, Jones J, Takeshita L, Ortega-Rivera ND, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res. 2020;48(D1):D783-D8.
98. Stroes ES, Thompson PD, Corsini A, Vladutiu GD, Raal FJ, Ray KK, et al. Statin-associated muscle symptoms: impact on statin therapy-European Atherosclerosis Society Consensus Panel Statement on Assessment, Aetiology and Management. Eur Heart J. 2015;36(17):1012-22.
99. Alfirevic A, Neely D, Armitage J, Chinoy H, Cooper RG, Laaksonen R, et al. Phenotype standardization for statin-induced myotoxicity. Clinical pharmacology and therapeutics. 2014;96(4):470-6.
100. Staffa JA, Chang J, Green L. Cerivastatin and reports of fatal rhabdomyolysis. The New England journal of medicine. 2002;346(7):539-40.
101. Kashani A, Phillips CO, Foody JM, Wang Y, Mangalmurti S, Ko DT, et al. Risks associated with statin therapy: a systematic overview of randomized clinical trials. Circulation. 2006;114(25):2788-97.
102. Romaine SP, Bailey KM, Hall AS, Balmforth AJ. The influence of SLCO1B1 (OATP1B1) gene polymorphisms on response to statin therapy. Pharmacogenomics J. 2010;10(1):1-11.
103. Carr DF, Francis B, Jorgensen AL, Zhang E, Chinoy H, Heckbert SR, et al. Genomewide Association Study of Statin-Induced Myopathy in Patients Recruited Using the UK Clinical Practice Research Datalink. Clinical pharmacology and therapeutics. 2019;106(6):1353-61.
104. Tonk ECM, Gurwitz D, Maitland-van der Zee AH, Janssens A. Assessment of pharmacogenetic tests: presenting measures of clinical validity and potential population impact in association studies. Pharmacogenomics J. 2017;17(4):386-92.
105. Stewart A. SLCO1B1 Polymorphisms and Statin-Induced Myopathy. PLoS Curr. 2013;5.
106. Pedro-Botet J, Climent E, Benaiges D. Muscle and statins: from toxicity to the nocebo effect. Expert Opin Drug Saf. 2019;18(7):573-9.
107. Tobert JA, Newman CB. Statin tolerability: In defence of placebo-controlled trials. Eur J Prev Cardiol. 2016;23(8):891-6.
108. Riaz H, Khan AR, Khan MS, Rehman KA, Alansari SAR, Gheyath B, et al. Meta-analysis of Placebo-Controlled Randomized Controlled Trials on the Prevalence of Statin Intolerance. Am J Cardiol. 2017;120(5):774-81.
109. Santos PCJL, Gagliardi ACM, Miname MH, Chacra AP, Santos RD, Krieger JE, et al. SLCO1B1 haplotypes are not associated with atorvastatin-induced

myalgia in Brazilian patients with familial hypercholesterolemia. European journal of clinical pharmacology. 2012;68(3):273-9.
110. Zhang L, Lv H, Zhang Q, Wang D, Kang X, Zhang G, et al. Association of SLCO1B1 and ABCB1 Genetic Variants with Atorvastatin-induced Myopathy in Patients with Acute Ischemic Stroke. Curr Pharm Des. 2019;25(14):1663-70.
111. Gaist D, Rodriguez LA, Huerta C, Hallas J, Sindrup SH. Lipid-lowering drugs and risk of myopathy: a population-based follow-up study. Epidemiology. 2001;12(5):565-9.
112. Voora D, Shah SH, Spasojevic I, Ali S, Reed CR, Salisbury BA, et al. The SLCO1B1*5 genetic variant is associated with statin-induced side effects. J Am Coll Cardiol. 2009;54(17):1609-16.
113. Yebyo HG, Aschmann HE, Kaufmann M, Puhan MA. Comparative effectiveness and safety of statins as a class and of specific statins for primary prevention of cardiovascular disease: A systematic review, meta-analysis, and network meta-analysis of randomized trials with 94,283 participants. Am Heart J. 2019;210:18-28.
114. Phan L, Jin Y, Zhang H, Qiang W, Shekhtman E, Shao D, et al. ALFA: Allele Frequency Aggregator. National Center for Biotechnology Information, US National Library of Medicine Available online: www ncbi nlm nih gov/snp/docs/gsr/alfa/(accessed on 10 March 2020). 2020.
115. National Library of Medicine, National CEnter for Biotechnology Information. Reference SNP (rs) Report, rs4149056 2020 [Available from: https://www.ncbi.nlm.nih.gov/snp/rs4149056#frequency_tab.
116. Xiang Q, Chen SQ, Ma LY, Hu K, Zhang Z, Mu GY, et al. Association between SLCO1B1 T521C polymorphism and risk of statin-induced myopathy: a meta-analysis. Pharmacogenomics J. 2018;18(6):721-9.
117. Brunham LR, Lansberg PJ, Zhang L, Miao F, Carter C, Hovingh GK, et al. Differential effect of the rs4149056 variant in SLCO1B1 on myopathy associated with simvastatin and atorvastatin. Pharmacogenomics J. 2012;12(3):233-7.
118. Hubáček JA, Dlouhá D, Adámková V, Zlatohlavek L, Viklický O, Hrubá P, et al. SLCO1B1 polymorphism is not associated with risk of statin-induced myalgia/myopathy in a Czech population. Medical science monitor: international medical journal of experimental and clinical research. 2015;21:1454.
119. Patel J, Superko HR, Martin SS, Blumenthal RS, Christopher-Stine L. Genetic and immunologic susceptibility to statin-related myopathy. Atherosclerosis. 2015;240(1):260-71.

120. Wilke RA, Ramsey LB, Johnson SG, Maxwell WD, McLeod HL, Voora D, et al. The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. Clinical pharmacology and therapeutics. 2012;92(1):112-7.
121. PharmGKB. Atorvastatin 2016 [Available from: https://www.pharmgkb.org/chemical/PA448500/clinicalAnnotation/1043880818.
122. British National Formulary. Adverse reactions to drugs 2020 [Available from: https://bnf.nice.org.uk/guidance/adverse-reactions-to-drugs.html.
123. Rodriguez-Martin S, Martin-Merino E, Lerma V, Rodriguez-Miguel A, Gonzalez O, Gonzalez-Herrada C, et al. Incidence of Stevens-Johnson syndrome/toxic epidermal necrolysis among new users of different individual drugs in a European population: a case-population study. Eur J Clin Pharmacol. 2019;75(2):237-46.
124. Braunack-Mayer A, Fabrianesi B, Street J, O'Shaughnessy P, Carter SM, Engelen L, et al. Sharing Government Health Data With the Private Sector: Community Attitudes Survey. J Med Internet Res. 2021;23(10):e24200.
125. Kimpton JE, Carey IM, Threapleton CJD, Robinson A, Harris T, Cook DG, et al. Longitudinal exposure of English primary care patients to pharmacogenomic drugs: An analysis to inform design of pre-emptive pharmacogenomic testing. British journal of clinical pharmacology. 2019;85(12):2734-46.
126. Orme B. Sample size issues for conjoint analysis studies. Sawthooth Software Research paper Series Squim, WA, USA: Sawthooth Software Inc. 1998.
127. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. Patient. 2015;8(5):373-84.
128. Louviere JJ, Hensher DA, Swait JD, Adamowicz W. Stated Choice Methods : Analysis and Applications. Cambridge, UNITED KINGDOM: Cambridge University Press; 2000.
129. Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA. Measuring patient preferences for colorectal cancer screening using a choice-format survey. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2007;10(5):415-30.
130. Linley WG, Hughes DA. Decision-makers' preferences for approving new medicines in Wales: a discrete-choice experiment with assessment of external validity. Pharmacoeconomics. 2013;31(4):345-55.
131. Johnson FR, Backhouse M. Eliciting stated preferences for health-technology adoption criteria using paired comparisons and recommendation

judgments. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2006;9(5):303-11.
132. ChoiceMetrics. Ngene 1.1.1. Australia2012.
133. Louviere JJ, Islam T, Wasi N, Street D, Burgess L. Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? Journal of Consumer Research. 2008;35(2):360-75.
134. Rogausch A, Prause D, Schallenberg A, Brockmoller J, Himmel W. Patients' and physicians' perspectives on pharmacogenetic testing. Pharmacogenomics. 2006;7(1):49-59.

Chapter 7: DCE results

7.1 Introduction

With tens of thousands of hospital admissions every year, adverse drug reactions (ADRs) are a major cause of mortality and morbidity (1, 2). One method that may help prevent some of these is the use of pharmacogenetics. Genetic testing can predict an individual's risk of an ADR in response to certain medicines. If the predicted risk is high, an alternative drug or reduced dose can be prescribed. This has been used successfully in the case of abacavir and associated abacavir hypersensitivity syndrome (see Chapter 1) (3-5) and capecitabine, *DPYD,* and toxicity (6, 7). The use of pharmacogenetics is increasing and is likely to become even more widespread in the near future (8).

However, patient acceptance is important for the wider implementation of pharmacogenetics, and this may be assessed by their preferences for medicines and testing services (9, 10). However, measuring these preferences can be difficult, particularly in complex situations like those involved in pharmacogenetics, which involve choices/decisions concerning the medicine, the disease being managed, the sharing of genetic information and the risk of ADR. Incorporating the views of patients is becoming increasingly important to regulators and funders, and is an invaluable tool to policy makers and healthcare professionals (9, 11, 12). It can potentially lead to improved uptake and adherence to treatment among patients (13, 14). Another element to consider is the level of evidence that matters to patients and the general public. There is little research in this area. Level of evidence is a complex concept to convey to a lay audience, but knowledge of preferences is important for those deciding whether to implement a genetic test (9, 12).

Genetic testing is moving into an era of panel testing (15, 16). Genetic panel tests, including 25+ genes, have been proposed for use in pharmacogenetics (15-23). These may be preferred to single gene tests, as the results can be stored and used to inform the prescribing of future medications the patient may require. They often do not cost significantly more than single gene tests and can even be more cost-effective per gene tested (15). However, panel tests are normally more complex, requiring validation of each included variant (22), and raise questions about data security and storage (15, 24, 25). Another issue is the level of evidence. Before a variant can be included in a panel test, or its results shared with prescribers for

clinical decision making, there should be sufficient evidence behind it for its clinical utility and validity (26). The forms that this evidence should take is debated (see Chapter 2 for a wider exploration of this issue).

The discrete choice experiment (DCE) is a method well-suited for measuring stated preferences in complex scenarios (27-32). It has been widely utilised in pharmacogenetic scenarios (Chapter 5) and healthcare more generally (12, 33-37). A DCE allows the measurement of stated preferences for the utility of a good or service (28, 38). They are based on Lancaster's Theory of Economic Value, which states how goods and services can be described by their attributes and the overall utility of the good or service as a function of its attributes (38). I have chosen this method to evaluate the preferences of the general public for genetic testing. A complex parameter, level of evidence, was incorporated into the experiment as an independent variable by randomising participants between two DCEs in the same disease area. One DCE was presented in the context of a 'high evidence' scenario, e.g. a well established gene-drug-ADR connection, highly rated on the Pharmacogenomics Knowledgebase (PharmGKB), and the other in the context of a 'low evidence' scenario, e.g. a newer or less studied gene-drug-ADR connection, lower rated on PharmGKB (see Chapter 6 for further discussion). I am not aware of any other DCE that evaluates participant preferences including level of evidence as a comparator.

This work has been informed by a systematic review of previous DCEs in ADRs (Chapter 5), and extensive qualitative research (Chapter 6). This chapter focusses on the analysis and implications of the DCE results (Figure 7.1).

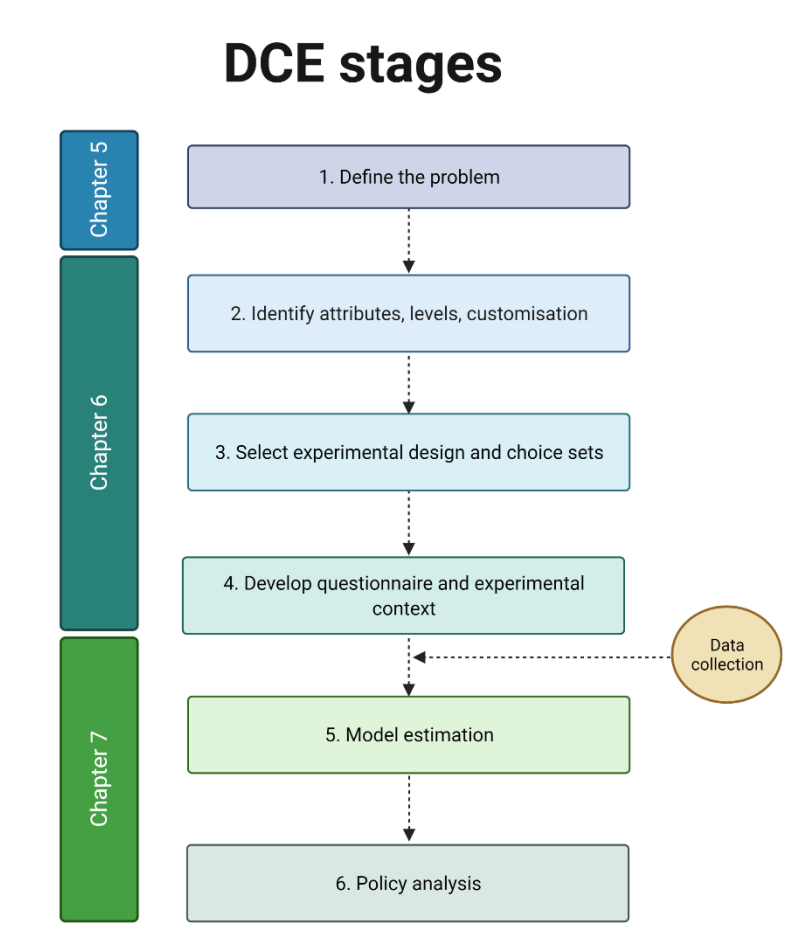


Figure 7.1 - DCE stages as defined by Street, et al. 2008 (39), with reference to the chapters of this thesis that correspond to each stage. Created using BioRender.com.

The aims of this experiment were to:

- Conduct a DCE in the general public
- Compare the general public's preferences for pharmacogenetic testing in high and low evidence scenarios
- Examine preferences for other aspects of genetic testing, including privacy and cost

7.2 Methods

7.2.1 Participants and administration

The survey was distributed to an age (18 and over) and gender representative sample of the UK population. This was completed through a market research company (Bilendi, London, UK (40)). Participants were compensated by Bilendi for their time. Safeguarding information (including researchers' contact details, and

contact information for mental health and disease-related charities) was provided at the start and end of each survey. Full details are shown in Appendix 6.3.

In order to measure participant preferences for level of evidence, participants were randomised between two different DCE scenarios – one with a 'high' level of evidence, and one with a 'low' level of evidence. This was repeated across 4 different clinical indications (HIV, cancer, epilepsy, and cardiovascular disease), for a total of 8 DCEs (Figure 7.2). The effect of the level of evidence supporting a genetic test may then be implied, based on the differences between utility for each scenario.

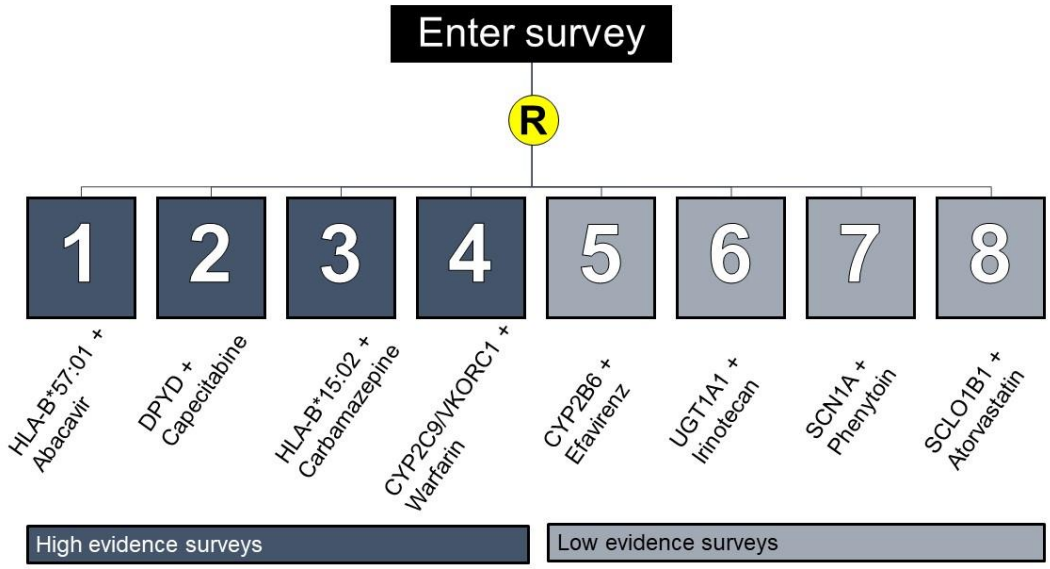


Figure 7.2 - Structure of the 8 DCE experiment, comparing participant preferences for genetic testing across high and low evidence. Participants enter the survey and are randomised to one of eight surveys, each of which contains a different scenario relating to genetic testing and adverse drug reactions. R = randomisation

Surveys were uploaded to the Jisc survey platform (London, UK) (41). Randomisation was set up by University of Liverpool Computing Services Desk. Participants were randomised upon indicating their consent to participate in the study (from https://ctrc.liv.ac.uk/indevelopment/dce).

Ethical approval for this study was granted by the University of Liverpool Health and Life Sciences Research Ethics Committee (Human participants, tissues and databases), reference number 4736.

7.2.2 Attribute and level selection

The development of attributes and levels was informed by a systematic review (Chapter 5) and a set of qualitative work that included surveys of experts and patients, focus groups with the general public, and pilot testing (Chapter 6).

The final list of attributes and levels is shown in Table 7.1.

Attribute	Levels
Risk of ADR from this medicine *	Level 0
	Level 1
	Level 2
Use of your data for further research	No
	Yes, but no contact (anonymous)
	Yes, and they can contact me (linked to medical record)
Number of medicines the test can be used to inform	1 (corresponds to single gene test)
	25 (corresponds to smaller panel test)
	50 (corresponds to larger panel test)
Cost of the test to the NHS (£)	10
	30
	50
Risk of serious ADR from any medicine over the next 10 years, excluding this medicine	1 in 5 (0.2)
	1 in 20 (0.05)
	1 in 50 (0.02)

*Table 7.1 - attributes and levels of the discrete choice experiment. *The levels for this attribute differ across each of the eight DCEs. ADR = adverse drug reaction.*

7.2.3 Experimental design

The DCE used a fractional factorial design, maximising D-efficiency with a multinomial logit (MNL) design. The design was generated in Ngene software (42) Full details of the DCE design are provided in Chapter 6.

All 8 DCEs had the same structure, and only differed on the introductory scenario shown to participants. Once randomised to a DCE, participants were given an introduction into genetic testing to prevent a specific gene-drug-ADR in that disease area. The level of evidence for genetic testing was highlighted (Figure 7.3A).

Participants were then given explanations of each attribute with accompanying pictograms (Figure 7.3B).

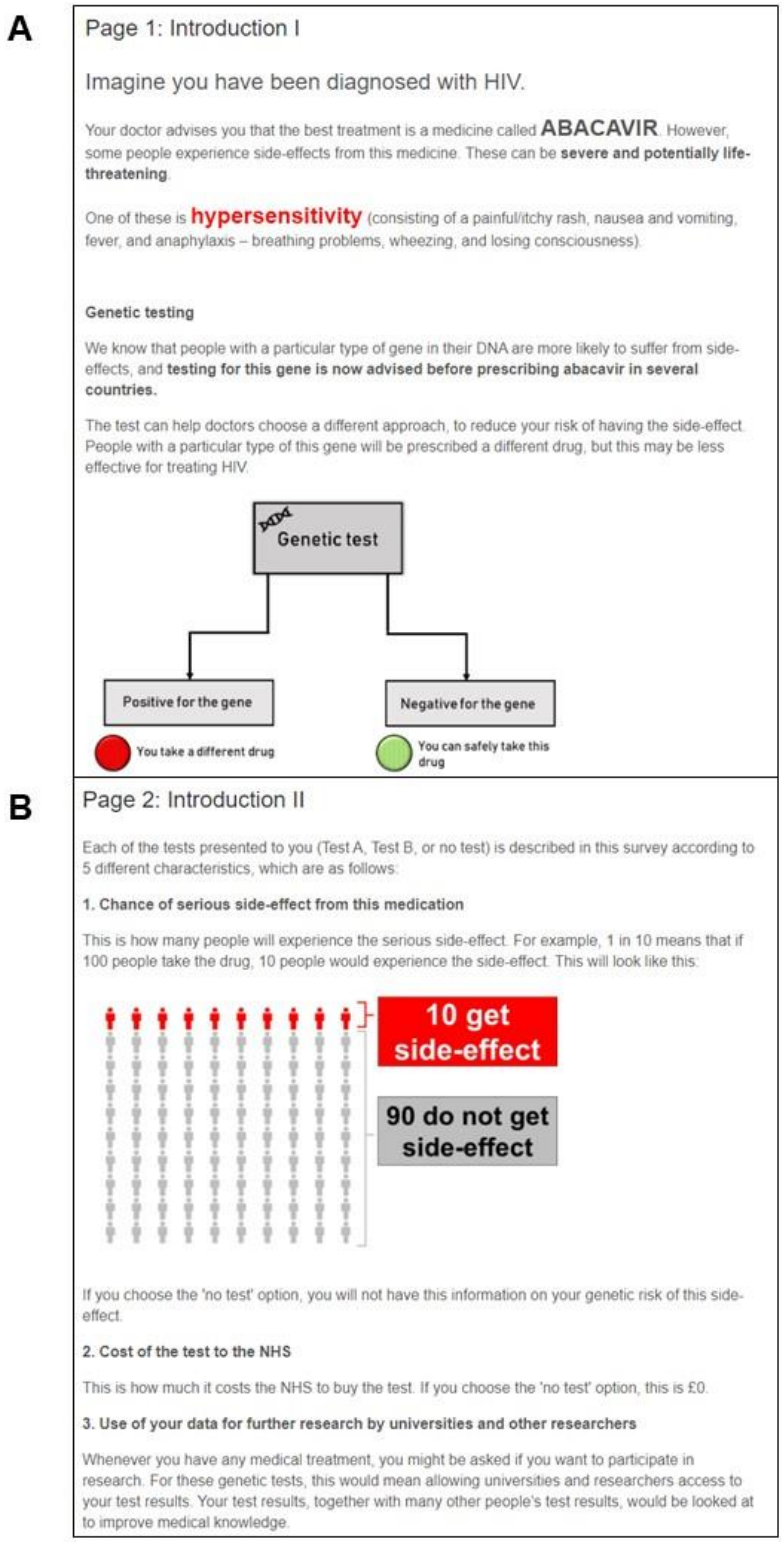


Figure 7.3 – Screenshots showing a portion of the DCE as participants saw it. A) First page of the abacavir DCE showing the disease scenario (HIV), drug (abacavir), and the adverse drug reaction (hypersensitivity). This is an example of a genetic test with 'high' levels of evidence, and this is represented to participants as 'testing for this gene is now advised before prescribing abacavir in several countries.' Genetic tests with 'low' levels of evidence were represented to participants as 'there isn't enough evidence to recommend testing anyone who needs to take [drug]' . B) Second page of the

abacavir DCE showing the start of the explanation of each attribute of the DCE, and how risks are represented in pictogram form.

Following these explanations, participants began the choice sets section, consisting of 15 forced choice questions. In the choice sets, participants were asked to choose between two different genetic tests, shown as ‘Test A’ or ‘Test B’.

Participants could also choose a ‘no test’ option, indicating that they would not choose either of the tests. This option was provided to more accurately reflect current clinical practice of genetic testing (true for all gene-drug-ADR combinations except abacavir) (43). ‘No test’ was associated with the highest risk of ADR, both from the current medicine and in the future. Levels for these were not shown to participants in the DCE, but discussed as part of the scenario shown to participants (see Appendix 6.3). The levels shown were used in the analysis of each DCE. It was assumed that the level for ‘Risk of ADR from this medicine’ would be the highest risk level of each ADR without genetic testing, as identified in Chapter 6. The level chosen for ‘Risk of serious ADR from any medicine over the next 10 years, excluding this medicine’ was specified as the highest level for that attribute (0.2), It was assumed that those who choose not to be tested may be more likely to suffer a future ADR than those who do get tested. Patients who do have a test will have pharmacogenetic data in their medical record, and this may be used in future prescriptions. Those without this data will be prescribed without this data in their record, so may be more likely to suffer an ADR.

Tests were not labelled to participants as genetic panel tests or single gene tests. These were modelled from the data in the later analyses.

Participants completed 15 choice tasks each. This number was chosen based on striking a balance between data collection and appropriate participant burden (31, 44, 45).

At the end of the survey, participants were asked to complete details of their age group and gender. They were also asked if they had previously had a genetic test, and if they had previously suffered from the disease in their scenario. Finally, participants were asked to rate the difficulty of completing the questionnaire, on a scale of 1 (not difficult at all) to 10 (almost impossible). All these questions were optional. Participants were also given the option to provide any additional feedback in a free text question at the end of the survey.

7.2.4 Analysis

7.2.4.1 Data coding

Data was downloaded from the Jisc platform and a data matrix was prepared for each DCE using Microsoft Excel. This contains every piece of data and is coded as per Table 7.2. The matrix contains multiple observations for each individual, as each respondent answers more than one discrete choice question. It also contains multiple observations for each choice set, since the choice sets presented to individuals contain 3 alternatives (Test A, Test B, no test) (32).

Attribute (name in code)	Levels	Levels design code	Expected direction of effect	Coding
Risk of ADR from this medicine (adr_today)	15 in 100 (0.15)	Level 0	Least preferred	Effects coding
	5 in 100 (0.05)	Level 1	Mid preferred +	
	3 in 100 (0.03)	Level 2	Most preferred ++	
Use of your data for further research (privacy)	No	Level 0	Mid preferred +	Effects coding
	Yes, no contact (anonymous)	Level 1	Most preferred ++	
	Yes, can contact (linked)	Level 2	Least preferred	
Number of medicines the test can be used to inform (medsno)	1	1	Least preferred	Continuous
	25	25	Mid preferred +	
	50	50	Most preferred ++	
Cost of the test to the NHS (cost)	£10	10	Most preferred ++	Continuous
	£30	30	Mid preferred +	
	£50	50	Least preferred	
Risk of serious ADR in the next 10 years (future_adr)	20 in 100 (0.2)	Level 0	Least preferred	Effects coding
	5 in 100 (0.05)	Level 1	Mid preferred +	
	2 in 100 (0.02)	Level 2	Most preferred ++	

Table 7.2 - data matrix coding DCEs. Levels for adr_today for abacavir are used here as an example.

Risk (both today and future) and privacy variables were coded using effects coding. Effects coding takes -1, 0, or 1 and effects are uncorrelated with the intercept (46). The reference level is -1 and was assigned to the highest risk categories. Level 1 risks were assigned to 0 and Level 2 (lowest risks) were assigned to 1. Privacy was coded similarly, with no further use your data assigned as the reference (-1) level and anonymous contact assigned to 0, and data linked to your medical record assigned to 1.

The remaining variables (medsno and cost) were coded as continuous in all DCEs.

Random effects logistic regression was used to estimate the parameters of the utility model given by:

$$V = \beta_{ASC} + \beta_{adr_{today_1}} + \beta_{privacy} + \beta_{medsno} + \beta_{cost} + \beta_{future\_adr} + \beta_{const} + \varepsilon_i$$

Where V is the utility derived from a given choice, ε refers to the error term, and all other variables are defined as attributes (βs are coefficients) (31, 32, 39, 47). β_{ASC} is an alternative specific constant (ASC) that captures differences in the mean of the distribution of the unobserved effects between the 'no test' and the other alternatives (48). ach parameter estimates the marginal utility of a change in that outcome, e.g., the utility of an increase of 1 in the number of medicines the test can be used to inform (49).

The regression was conducted in Stata version 14 (StataCorp, College Station TX, USA) (50). Bootstrapping was used to calculate confidence intervals, with 1000 replications. Under effects coding, the value of the omitted variable was given by (46, 49, 51):

$$\beta_0 = -1 * (\beta_1 + \beta_2)$$

where β_0 is the coefficient of the omitted variable, and β_1 and β_2 are the coefficients of the included variables.

7.2.4.2 Demographics and written responses

Details of participant ages and genders were collected and summarising plots and tables were produced in RStudio (52). The same methods were used to plot whether participants had previously suffered from the illness described in the survey, and if they had ever had a genetic test before.

Written responses, an optional free text question at the end of the survey, were collected and checked for identifying information. Any entry containing identifying information was removed before further analysis. No formal qualitative analysis was undertaken, but responses were read and some themes are summarised below.

7.2.4.3 Stata coding

Random effects logit (command `xtlogit`) was used in this analysis (53). This allows for multiple observations per participant and per choice set. A bootstrapped sample was computed in order to be able to calculate confidence intervals for utility estimates.

The code used in Stata 14 for the analysis can be found in the Appendix 7.1.

7.2.4.4 Calculation of a preference-weighted utility model

The utility of each test type was calculated by weighting the 1000 bootstrapped results of the regression against base case assumptions of outcomes with panel testing, single gene testing, and not testing. The base case is the likely 'real-world' testing scenario, to which changes can be made to estimate their impact (32). Base cases for each gene-drug-ADR combination were constructed using estimates from the systematic review (Chapter 5) and from the literature (Table 7.3).

	DCE name	Attribute	Base case value	Rationale
Panel test and single gene test	Abacavir	adr_today	Level 2 (smallest risk)	Lowest level for each DCE type. See Chapter 6 for further details of this calculation.
	Efavirenz	adr_today		
	Capecitabine	adr_today		
	Irinotecan	adr_today		
	Carbamazepine	adr_today		
	Phenytoin	adr_today		
	Warfarin	adr_today		
	Atorvastatin	adr_today		
Panel test	All DCEs	privacy	Anonymous data sharing (Level 1)	Current practice and likely future practice of data sharing in the NHS (54-56)

		medsno	25	The NHSEI and Genomics England Pharmacogenomics Working Group completed an evidence review (literature up to 2019) and defined an initial priority shortlist of 29 drug-gene pairs with potential clinical utility for pharmacogenomic testing (17)
		cost	50 (Level 2)	D Hughes (personal communication) Based on prices given by Illumina for a global screening array kit (as of December 2020) (57).
		future_adr	(Level 0)	Consulting expert opinion, based on Kimpton, *et al.* (2019) (18) and D Hughes (personal communication) (17).
Single gene test	All DCEs	privacy	Anonymous data sharing (Level 1)	Current practice and likely future practice of data sharing in the NHS (54-56)
		medsno	1	Definition of single gene test
		cost	30 (Level 1)	D Hughes (personal communication) Based on prices given by Illumina for a global screening array kit (as of December 2020) (57).

		future_adr	(Level 1)	Consulting expert opinion, based on Kimpton, *et al.* (2019) (18) and D Hughes (personal communication) (17).
No test	All DCEs	adr_today	Highest level in each DCE (Level 2)	
		privacy	No data sharing (Level 0)	Nothing to share
		medsno	0	Cannot inform any medicines
		cost	0	No test costs nothing
		future_adr	(Level 2)	Based on highest level

Table 7.3 - Base case assumptions for each type of test modelled in the analysis – a panel test of multiple genes, a single gene test, and no test. Base case values for risk represent the highest level of available risks for each DCE type.

Evidence indicates that full understanding of very rare risks by the general public is limited (58-61). For these reasons, risks of very rare ADRs (in carbamazepine, phenytoin, warfarin, and atorvastatin DCEs) were presented in categories (Uncommon, Rare, etc). It was assumed that more common ADR risks (in abacavir, efavirenz, capecitabine, and irinotecan DCEs) would be more easily comprehended by participants.

In the analysis, treating more common ADRs as continuous variables led to non-linearities on the ADR attributes. This meant that it was most appropriate the model all risks using effects coding. Consequently, the utility model used the category that most closely matched the actual risk for each gene-drug-ADR scenario.

The total utility for each test type was given by:

$$\sum (\beta_{adr_{today}} * eventrate_{adrtoday}) + (\beta_{privacy} * eventrate_{privacy}) + (\beta_{medsno} * eventrate_{medsno}) + (\beta_{cost} * eventrate_{cost}) + (\beta_{adr_{future}} * eventrate_{adrfuture}) + \beta_{const} + \beta_{ASC} + \varepsilon_i$$

This was repeated across all eight DCEs, using the full range of bootstrapped values obtained from Stata. This allowed for the calculation of 95% confidence

intervals. This analysis was completed in RStudio (1.4.1106, RStudio Team, Boston MA) (52).

Base case numbers were multiplied by β coefficients generated in Stata.

Due to heterogeneity in experiment design, the significance of the difference between DCEs was compared narratively.

7.2.4.5 Other statistics

The rate of participants choosing 'no test' was calculated by dividing the total number of 'no test' responses by the total number of individual responses for that survey. The rate of skipping choice sets in one DCE (atorvastatin) was calculated in the same way.

7.3 Results

7.3.1 Demographics

A total of 2,019 responses were collected, evenly distributed across the 8 DCEs (Table 7.4). There were similar numbers of male and female participants, and a small number of other genders (49.7% female, 48.7% male, 0.4% another gender). Older age groups were more highly represented in the sample, with the largest sample coming from the over 65s (24.8%). This is comparable to the UK general population (62).

Participants were asked to rate the difficulty of the DCE on a scale of 1 (not difficult at all) to 10 (almost impossible). Most participants rated the difficulty as 1, with a mean difficulty rating of 3.5.

	Abacavir	Capecitabine	Carbamazepine	Warfarin	Efavirenz	Irinotecan	Phenytoin	Atorvastatin	**Total**
n	232	245	242	265	251	248	260	276	**2019**
Gender									
Female (%)	115 (49.6)	143 (58.4)	115 (47.5)	126 (47.5)	124 (49.4)	120 (48.4)	124 (47.7)	137 (49.6)	**1004 (49.7)**
Male (%)	111 (47.8)	100 (40.8)	121 (50.0)	136 (51.3)	124 (49.4)	123 (49.6)	133 (51.2)	136 (49.3)	**984 (48.7)**
Another (%)	2 (0.9)	0 (0)	1 (0.4)	2 (0.8)	2 (0.8)	0 (0)	0 (0)	1 (0.4)	**8 (0.4)**
Prefer not to answer (%)	4 (1.7)	2 (0.8)	5 (2.1)	1 (0.4)	1 (0.4)	5 (2.0)	3 (1.2)	2 (0.7)	**23 (1.1)**
Age group									
18-24 (%)	25 (10.8)	14 (5.7)	13 (5.4)	17 (6.4)	15 (6.0)	20 (8.1)	18 (6.9)	24 (8.7)	**146 (7.2)**
25-34 (%)	38 (16.4)	35 (14.3)	36 (14.9)	37 (14.0)	42 (16.7)	35 (14.1)	36 (13.8)	32 (11.6)	**291 (14.4)**
35-44 (%)	43 (18.5)	35 (14.3)	38 (15.7)	43 (16.2)	54 (21.5)	49 (19.8)	40 (15.4)	48 (17.4)	**350 (17.3)**
45-54 (%)	43 (18.5)	39 (15.9)	46 (19.0)	49 (18.5)	51 (20.3)	45 (18.1)	62 (23.8)	61 (22.1)	**396 (19.6)**
55-64 (%)	33 (14.2)	50 (20.4)	40 (16.5)	49 (18.5)	29 (11.6)	37 (14.9)	46 (17.7)	43 (15.6)	**327 (16.2)**
65+	50 (21.6)	72 (29.4)	67 (27.7)	70 (26.4)	59 (23.5)	61 (24.6)	57 (21.9)	65 (23.6)	**501 (24.8)**
Prefer not to answer (%)	0 (0)	0 (0)	2 (0.8)	0 (0)	1 (0.4)	1 (0.4)	1 (0.4)	3 (1.1)	**8 (0.4)**
Difficulty									
Mean difficulty rating	3.5	3.5	3.3	3.4	3.4	3.5	3.4	3.6	**3.5**
Not answered (%)	1 (0.4)	1 (0.4)	1 (0.4)	3 (1.1)	4 (1.6)	1 (0.4)	2 (0.8)	3 (1.1)	**16 (0.8)**

Table 7.4 - demographics of DCE sample including gender, age, and difficulty rating for each of the 8 DCEs.

Participants were also asked if they had previously had a genetic test, or if they had previously suffered from the illness described in their survey. Only a small number had previously had a genetic test (6.4%) (Table 7.5A), and even fewer had previously suffered from the illness (overall 3.6%) (Table 7.5B). This was highest for the cardiovascular DCEs, warfarin and atorvastatin (5.3% and 6.9%).

A

	abacavir (N=232)	capecitabine (N=245)	carbamazepine (N=242)	warfarin (N=265)	efavirenz (N=251)	irinotecan (N=248)	phenytoin (N=260)	atorvastatin (N=276)	Overall (N=2019)
Had a genetic test before?									
Yes	14 (6.0%)	17 (6.9%)	16 (6.6%)	17 (6.4%)	21 (8.4%)	12 (4.8%)	19 (7.3%)	13 (4.7%)	129 (6.4%)
No	209 (90.1%)	211 (86.1%)	210 (86.8%)	225 (84.9%)	215 (85.7%)	223 (89.9%)	223 (85.8%)	253 (91.7%)	1769 (87.6%)
Don't know	5 (2.2%)	11 (4.5%)	12 (5.0%)	17 (6.4%)	12 (4.8%)	8 (3.2%)	14 (5.4%)	4 (1.4%)	83 (4.1%)
Prefer not to answer	0 (0%)	2 (0.8%)	0 (0%)	1 (0.4%)	0 (0%)	3 (1.2%)	1 (0.4%)	1 (0.4%)	8 (0.4%)
Missing	4 (1.7%)	4 (1.6%)	4 (1.7%)	5 (1.9%)	3 (1.2%)	2 (0.8%)	3 (1.2%)	5 (1.8%)	30 (1.5%)

B

	abacavir (N=232)	capecitabine (N=245)	carbamazepine (N=242)	warfarin (N=265)	efavirenz (N=251)	irinotecan (N=248)	phenytoin (N=260)	atorvastatin (N=276)	Overall (N=2019)
Had the disease before?									
Yes	4 (1.7%)	8 (3.3%)	10 (4.1%)	14 (5.3%)	6 (2.4%)	3 (1.2%)	8 (3.1%)	19 (6.9%)	72 (3.6%)
No	221 (95.3%)	228 (93.1%)	225 (93.0%)	239 (90.2%)	238 (94.8%)	233 (94.0%)	246 (94.6%)	243 (88.0%)	1873 (92.8%)
Don't know	2 (0.9%)	3 (1.2%)	3 (1.2%)	8 (3.0%)	1 (0.4%)	5 (2.0%)	4 (1.5%)	6 (2.2%)	32 (1.6%)
Prefer not to answer	0 (0%)	3 (1.2%)	0 (0%)	0 (0%)	1 (0.4%)	4 (1.6%)	0 (0%)	2 (0.7%)	10 (0.5%)
Missing	5 (2.2%)	3 (1.2%)	4 (1.7%)	4 (1.5%)	5 (2.0%)	3 (1.2%)	2 (0.8%)	6 (2.2%)	32 (1.6%)

Table 7.5 – A) Participant answers to 'Have you ever had a genetic test before?', categorised by DCE type. B) Participant answers to 'Have you had the illness mentioned in this survey before?', categorised by DCE type.

To calculate the rate at which participants chose each option, the total number of each response type was summed and divided by the total number of individual responses. Rates of choosing Test A and Test B were similar across DCE types.

The rates of choosing ‘no test’ were lowest for irinotecan (0.097) and highest in the phenytoin survey (0.174) (Table 7.6).

DCE type	Test A	Test B	No test rate
Abacavir	0.405	0.454	0.141
Capecitabine	0.424	0.428	0.147
Carbamazepine	0.416	0.447	0.14
Warfarin	0.411	0.444	0.15
Efavirenz	0.434	0.453	0.124
Irinotecan	0.452	0.451	0.097
Phenytoin	0.406	0.419	0.175
Atorvastatin*	0.388	0.455	0.154

*Table 7.6 - rates of each response for each DCE. *Some questions int he atorvastatin DCE were not made mandatory due to researcher error.*

7.3.2 Written responses

Many participants chose to provide feedback in an optional question at the end of the survey.

The issue of cost was the deciding factor for some participants:

Truth is, NHS needs to save money... And life threatening side effects on a (eventually) terminal illness is not such a concern in my mind...

I felt it was of more value to test for multiple drugs rather than just one, whatever the cost. However the cost needs to be considered whenever possible.

My main concern was the cost to the NHS.

Many also commented on privacy issues, with some strong feelings on data sharing and data privacy:

I am used to volunteering for this type of testing, I have MS [multiple sclerosis] *so am well aware of how volunteering for this and offering results to studies help, if people don’t take part then researchers don’t make leaps forward in treatments and new medicines.*

I mostly choose the options where my data would be used for research as this might lead to further medicines being developed

is result of the genetic test held securely? Can an insurance company, for example, get a hold of it? Hackers? Sways my answers.

While many positive comments were received, some participants responses raised issues with potential misunderstandings of the DCE:

i thought i was going to get a genetic test from this. shame.

i really didn't understand why having a test would have any bearing on the chance of potential serious side effects of other medications you might take over the next 10 years, it wasn't properly explained and needs some work.

Nobody is likely to take a test that may result in severe burns

7.3.3 Results by disease area

Data matrices for each DCE can be found in Appendix 7.2. Full β-coefficients and details of utility modelling can be found in Appendix 7.3 and 7.4.

7.3.3.1 HIV: Abacavir and Efavirenz

7.3.3.1.1 Preferences

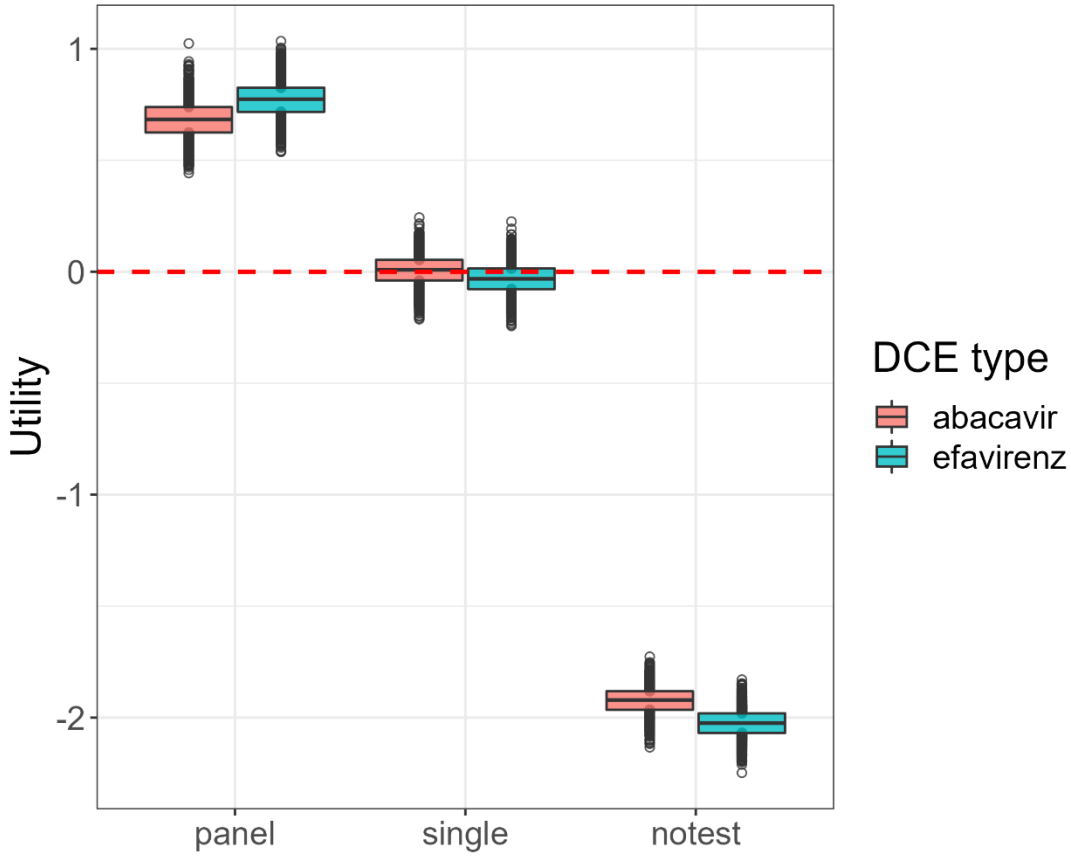
Preferences in both abacavir and efavirenz experiments followed the expected direction of effect in terms of risk, with lower risk preferred in terms of both the risk of ADR from the current drug, and the future risk of ADR. The expected direction of effect was also seen for the number of medicines (positive, greater numbers preferred) and the cost of the test (negative, lower prices preferred). Anonymous data sharing was the most preferred option regarding privacy of data. Full results can be seen in Table 7.7.

Attribute	Abacavir-*HLA-B*57:01*-AHS	Efavirenz-*CYP2B6*-DILI
Risk of ADR today 0 (highest)	-0.663*	-0.640*
Risk of ADR today 1 (mid)	0.226 (0.155 to 0.297) ***p=4.51E-10***	0.216 (0.144 to 0.287) ***p=3.00E-09***
Risk of ADR today 2 (lowest)	0.435 (0.363 to 0.507) ***p=1.68E-32***	0.422 (0.350 to 0.494) ***p=1.77E-30***
Privacy – no data sharing	-0.113*	-0.078*
Privacy – anonymous data sharing	0.099 (0.025 to 0.174) ***p=0.009***	0.196 (0.122 to 0.270) ***p=2.03E-07***
Privacy – sharing linked to medical records	0.014 (-0.059 to 0.087) *p=0.713*	-0.120 (-0.194 to -0.047) ***p=0.001***
Number of medicines	0.009 (0.006 to 0.013) ***p=6.06E-09***	0.014 (0.011 to 0.017) ***p=1.14E-18***
Cost of test	-0.007 (-0.011 to -0.003) ***p=0.0003***	-0.009 (-0.012 to -0.005) ***p=5.61E-06***
Risk of future ADR 0	-0.597*	-0.532*
Risk of future ADR 1	0.004 (-0.068 to 0.077) *p=0.904*	-0.049 (-0.122 to 0.023) *p=0.183*
Risk of future ADR 2	0.591 (0.510 to 0.671) ***p=2.63E-47***	0.580 (0.500 to 0.660) ***p=4.94E-46***
Constant	-0.336 (-0.444 to -0.229) ***p=8.09E-10***	-0.352 (-0.459 to -0.244) ***p=1.41E-10***
ASC	-0.216 (-0.397 to -0.034) ***p=0.020***	-0.426 (-0.608 to -0.244) ***p=4.45E-06***
No. of observations	10,440	10,440
No. of groups	232	232
Sigma (Standard deviation)	0.0007 (4.86E-08 to 9.661)	0.0007 (6.32E-08 to 7.914)
Rho	1.43e-07 (7.19e-16 to 0.966)	1.52E-07 (1.21E-15 to 0.950)

*Table 7.7 – abacavir and efavirenz attribute beta coefficients, 95% confidence intervals (generated by 1000 bootstrap replications), and p-values. *This variable was omitted from the regression model as per effects coding (see Bech and Gyrd-Hansen (2005) (45) for more detail on this) and calculated from the other two values in that class. ADR – adverse drug reaction. AHS = abacavir hypersensitivity syndrome. ASC = alternative specific constant. DILI = drug induced liver injury.*

7.3.3.1.2 Utility modelling

The utility of each test type was compared between high evidence (abacavir) and low evidence (efavirenz) indications. The ranges of utilities calculated across panel, single gene, and no test scenarios shows considerable overlap, indicating there is likely to be no difference in utility between high and low evidence scenarios (Figure 7.4).



	Abacavir	Efavirenz
Panel test	0.684 (0.443 to 1.024)	0.773 (0.538 to 1.035)
Single gene test	0.008 (-0.213 to 0.244)	-0.031 (-0.243 to 0.225)
No test	-1.923 (-2.133 to -1.726)	-2.026 (-2.247 to -1.829)

Figure 7.4 - Total utility of panel, single gene, and no test scenarios in high evidence (abacavir) and low evidence (efavirenz) DCEs.

7.3.3.2 Cancer: Capecitabine and Irinotecan

7.3.3.2.1 Preferences

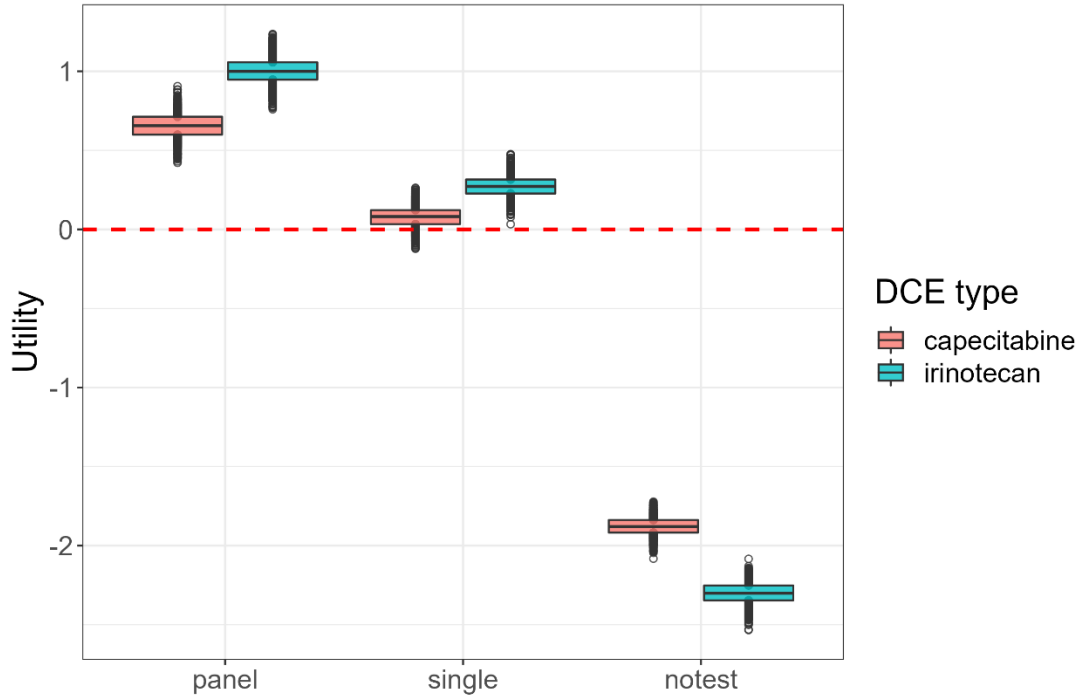
Preferences followed the expected direction of effect in terms of risk, with lower risk preferred in both the risk of ADR from the current drug, and the future risk of ADRs. This was also correct for the number of medicines (greater numbers preferred) and the cost of the test (lower prices preferred). Anonymous data sharing was the most preferred option regarding privacy of data. Full results can be seen in Table 7.8.

Attribute	Capecitabine – *DPYD* – Neutropenia	Irinotecan – UGT1A1 – Neutropenia
Risk of ADR today 0 (highest)	-0.574*	-0.574*
Risk of ADR today 1 (mid)	0.0004 (-0.068 to 0.069) *p=0.992*	0.022 (-0.045 to 0.088) *p=0.520*
Risk of ADR today 2 (lowest)	0.572 (0.503 to 0.641) ***p=1.76E-59***	0.584 (0.516 to 0.651) ***p=6.35E-65***
Privacy – no data sharing	-0.122*	-0.062
Privacy – anonymous data sharing	0.119 (0.048 to 0.191) ***p=0.001***	0.178 (0.103 to 0.252) ***p=3.07E-06***
Privacy – sharing linked to medical records	0.002 (-0.069 to 0.072) *p=0.965*	-0.116 (-0.186 to -0.046) ***p=0.001***
Number of medicines	0.010 (0.007 to 0.013) ***p=3.61E-10***	0.010 (0.007 to 0.013) [checked and they are truly the same – differences found with more precision] ***p=5.20E-11***
Cost of test	-0.006 (-0.009 to -0.002) ***p=0.001***	-0.006 (-0.009 to -0.002) ***p=0.002***
Risk of future ADR 0	-0.293*	-0.293*
Risk of future ADR 1	-0.085 (-0.155 to -0.014) ***p=0.018***	-0.053 (-0.122 to 0.016) *p=0.134*
Risk of future ADR 2	0.376 (0.299 to 0.453) ***p=6.23E-22***	0.549 (0.471 to 0.627) ***p=4.31E-43***
Constant	-0.362 (-0.465 to -0.259) ***p=6.08E-12***	-0.273 (-0.373 to -0.172) ***p=9.62E-08***
ASC	-0.528 (-0.698 to -0.359) ***p=1.08E-09***	-0.863 (-1.038 to -0.688) ***p=4.44E-22***
No. of observations	11,025	11,160
No. of groups	245	248
Sigma (Standard deviation)	0.0007 (4.48E-08 to 9.655)	0.0004 (6.26E-06 to 0.022)
Rho	1.32E-07 (6.10E-16 to 0.966)	4.19E-08 (1.19E-11 to 0.0001)

*Table 7.8 - capecitabine and irinotecan attribute beta coefficients, 95% confidence intervals (generated by 1000 bootstrap replications), and p-values. *This variable was omitted from the regression model as per effects coding (see Bech and Gyrd-Hansen (2005) (45) for more detail on this) and calculated from the other two values in that class. ADR = adverse drug reaction. ASC = alternative specific constant.*

7.3.3.2.2 Utility modelling

The utility of each test type was compared between high evidence (capecitabine) and low evidence (efavirenz) indications. The ranges of utilities calculated across panel, single gene, and no test scenarios show some overlap, indicating there is likely to be little difference in utility between high and low evidence scenarios Figure 7.5.



	Capecitabine	Irinotecan
Panel test	0.656 (0.421 to 0.906)	1.002 (0.760 to 1.235)
Single gene test	0.079 (-0.122 to 0.263)	0.273 (0.033 to 0.476)
No test	-1.879 (-2.082 to -1.722)	-2.301 (-2.534 to -2.084)

Figure 7.5 – Total utility of panel, single gene, and no test scenarios in high evidence (capecitabine) and low evidence (irinotecan) DCEs.

7.3.3.3 Epilepsy: Carbamazepine and Phenytoin

7.3.3.3.1 Preferences

Preferences in carbamazepine and phenytoin DCEs followed the expected direction of effect. Lower risk was preferred in both the risk of ADR from the current drug, and the risk of future ADRs. The expected direction of effect was also seen for the

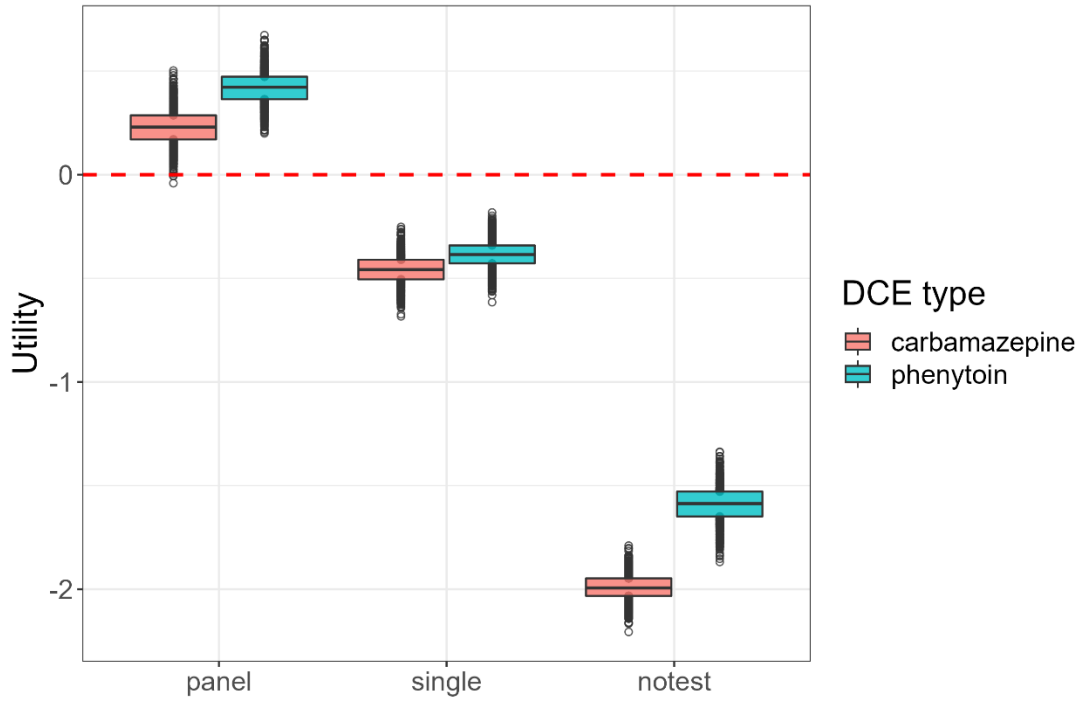
number of medicines the test could inform (positive, greater numbers preferred), and the cost of the test (negative, lower prices preferred). Anonymous data sharing was the most preferred option regarding privacy of data. Full results can be seen in Table 7.9.

Attribute	Carbamazepine – *HLA-B*15:02* – SJS/TEN	Phenytoin – *HLA-A*24:02* – SJS/TEN
Risk of ADR today 0 (highest)	-0.247*	-0.393*
Risk of ADR today 1 (mid)	0.018 (-0.050 to 0.086) *p=0.608*	0.041 (-0.023 to 0.106) *p=0.210*
Risk of ADR today 2 (lowest)	0.228 (0.159 to 0.297) ***p=8.71E-11***	0.350 (0.284 to 0.417) ***p=6.31E-25***
Privacy – no data sharing	-0.168*	-0.174*
Privacy – anonymous data sharing	0.131 (0.060 to 0.202) ***p=0.0003***	0.133 (0.064 to 0.203) ***p=0.0002***
Privacy – sharing linked to medical records	0.036 (-0.033 to 0.106) *p=0.30780295*	0.040 (-0.029 to 0.108) *p=0.254*
Number of medicines	0.019 (0.016 to 0.022) ***p=8.99E-36***	0.018 (0.015 to 0.021) ***p=2.09E-32***
Cost of test	-0.011 (-0.015 to -0.007) ***p=1.42E-09***	-0.007 (-0.010 to -0.004) ***p=0.00001***
Risk of future ADR 0	-0.228*	-0.408*
Risk of future ADR 1	-0.109 (-0.179 to -0.039) ***p=0.002***	-0.056 (-0.123 to 0.011) *p=0.102*
Risk of future ADR 2	0.337 (0.261 to 0.413) ***p=3.57E-18***	0.462 (0.389 to 0.536) ***p=3.31E-35***
Constant	-0.400 (-0.502 to -0.297) ***p=2.39E-14***	-0.618 (-0.702 to -0.535) ***p=1.26E-47***
ASC	-0.944 (-1.112 to -0.776) ***p=3.18E-28***	-0.002 (-0.108 to 0.104) *p=0.970*
No. of observations	10,890	11,700
No. of groups	242	260
Sigma (Standard deviation)	0.0007 (3.88E-08 to 10.931)	0.0008 (6.84E-07 to 0.907)
Rho	1.29E-07 (4.56E-16 to 0.973)	1.88E-07 (1.42E-13 to 0.200)

*Table 7.9 - carbamazepine and phenytoin attribute beta coefficients, 95% confidence intervals (generated by 1000 bootstrap replications), and p-values. *This variable was omitted from the regression model as per effects coding (see Bech and Gyrd-Hansen (2005) (45) for more detail on this) and calculated from the other two values in that class. ADR = adverse drug reaction. ASC = alternative specific constant. SJS/TEN = Stevens-Johnson syndrome/toxic epidermal necrolysis.*

7.3.3.3.2 Utility modelling

The utility of each test type was compared between high evidence (carbamazepine) and low evidence (phenytoin) DCEs. The range of utilities calculated across panel test, single gene, and no test scenarios shows overlap, indicating there is likely to be no difference in utility between high and low evidence scenarios (Figure 7.6). There was a small difference in utility of in the no test scenario, although these were both still negative.



	Carbamazepine	Phenytoin
Panel test	0.229 (-0.041 to 0.503)	0.420 (0.200 to 0.673)
Single gene test	-0.458 (-0.683 to -0.252)	-0.386 (-0.615 to -0.182)
No test	-1.991 (-2.206 to -1.789)	-1.590 (-1.868 to -1.337)

Figure 7.6 – Total utility of panel, single gene, and no test scenarios in high evidence (carbamazepine) and low evidence (phenytoin) DCEs.

7.3.3.4 Cardiovascular disease: Warfarin and Atorvastatin

Due to researcher error some of the choice sets in the atorvastatin DCE were made optional and were subsequently not completed by some participants. These responses (a total of 14/4140 total responses, 0.34%) were removed before final analysis.

7.3.3.4.1 Preferences

Coefficients in the warfarin and atorvastatin experiments followed expected direction of effect. Lower risk was preferred in terms of both risk from the current drug, and the future risk of ADR. The expected direction of effect was also seen for the number of medicines the test could inform (positive, greater numbers preferred) and the cost of the test (negative, lower prices preferred). Anonymous data sharing was the most preferred option in the warfarin experiment, but full data sharing was the most preferred in the atorvastatin experiment. However, these values in the atorvastatin experiment were not statistically significant. Full details of results can be found in Table 7.10.

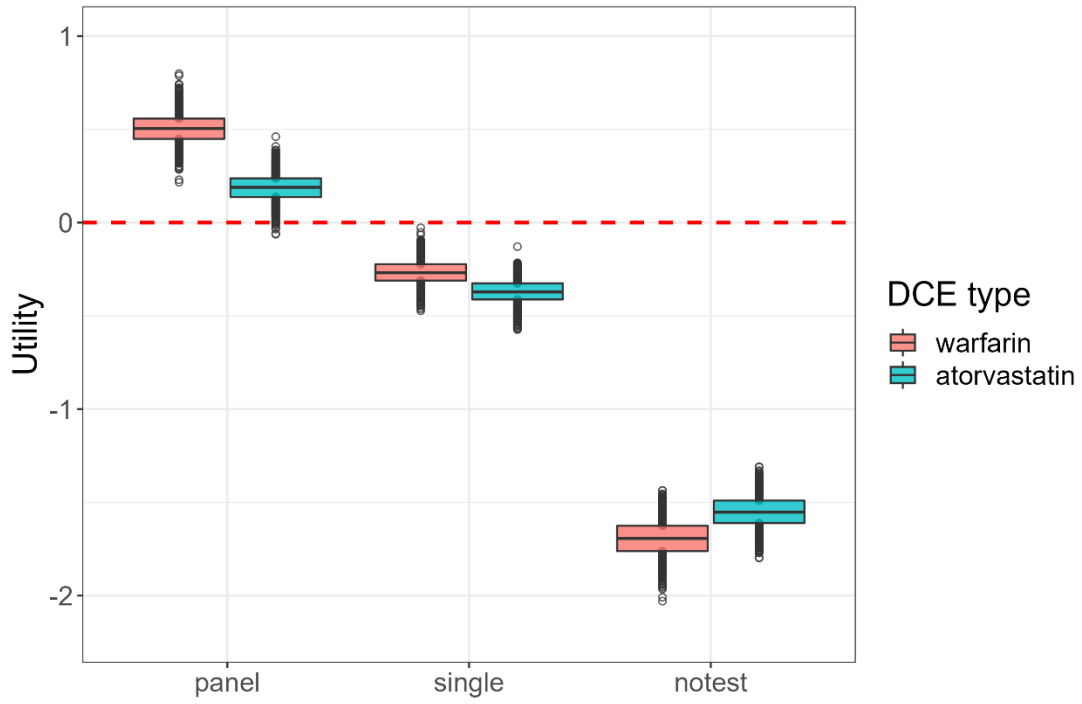
Attribute	Warfarin – *CYP2C9/VKORC1* – bleeding	Atorvastatin – *SLCO1B1* – muscle ADRs
Risk of ADR today 0 (highest)	-0.551*	-0.417*
Risk of ADR today 1 (mid)	0.149 (0.085 to 0.213) ***p=4.52E-06***	0.110 (0.047 to 0.172) ***p=5.44E-04***
Risk of ADR today 2 (lowest)	0.400 (0.335 to 0.466) ***p=1.09E-32***	0.306 (0.242 to 0.370) ***p=8.35E-21***
Privacy – no data sharing	-0.124*	-0.123*
Privacy – anonymous data sharing	0.099 (0.030 to 0.168) ***p=5.15E-03***	0.057 (-0.010 to 0.124) *p=0.096*
Privacy – sharing linked to medical records	0.024 (-0.044 to 0.092) *p=0.490*	0.064 (-0.002 to 0.130) *p=0.056*
Number of medicines	0.017 (0.014 to 0.020) ***p=1.38E-28***	0.016 (0.013 to 0.019) ***p=7.00E-29***
Cost of test	-0.007 (-0.010 to -0.004) ***p=8.78E-06***	-0.006 (-0.009 to -0.003) **p=0.0001**
Risk of future ADR 0	-0.483*	-0.403*
Risk of future ADR 1	-0.014 (-0.081 to 0.052) *p=0.676*	0.057 (-0.007 to 0.122) *p=0.081*
Risk of future ADR 2	0.496 (0.424 to 0.569) ***p=6.52E-41***	0.345 (0.275 to 0.416) ***p=9.66E-22***
Constant	-0.549 (-0.632 to -0.466) ***p=1.89E-38***	-0.626 (-0.706 to -0.545) ***p=2.40E-52***
ASC	0.008 (-0.106 to 0.122) *p=0.890*	0.016 (-0.092 to 0.123) *p=0.776*
No. of observations	11,925	12,420
No. of groups	265	276
Sigma (Standard deviation)	0.0008 (1.05E-06 to 0.647)	0.0008 (8.85E-07 to 0.710)
Rho	2.05E-07 (3.32E-13 to 0.113)	1.91E-07 (2.38E-13 to 0.133)

*Table 7.10 - warfarin and atorvastatin attribute beta coefficients, 95% confidence intervals (generated by 1000 bootstrap replications), and p-values. *This variable was omitted from the regression model as per effects coding (see Bech and Gyrd-Hansen (2005) (45) for more detail on this) and calculated from the other two values in that class. ADR = adverse drug reaction. ASC = alternative specific constant.*

7.3.3.4.2 Utility modelling

The utility of each test type was compared between high evidence (warfarin) and low evidence (atorvastatin) indications. The ranges of utilities across panel, single

gene, and no test scenarios showed considerable overlap, indicating there is likely to be no difference in utility between high and low evidence scenarios (Figure 7.7).



	Warfarin	Atorvastatin
Panel test	0.504 (0.217 to 0.799)	0.187 (-0.062 to 0.461)
Single gene test	-0.269 (-0.473 to -0.028)	-0.371 (0.574 to -0.129)
No test	-1.695 (-2.030 to -1.436)	-1.551 (-1.798 to -1.309)

Figure 7.7 - Total utility of panel, single gene, and no test scenarios in high evidence (warfarin) and low evidence (atorvastatin) DCEs.

7.4 Discussion

This final chapter details the successful implementation of 8 DCEs that measured preferences for genetic testing in the general public. I tested the general public's preferences for genetic panel tests vs single gene tests vs no test, for a number of clinical contexts and tests associated with varying levels of supporting evidence. There did not appear to be a difference in utility between 'high' and 'low' evidence scenarios. This difference was not formally assessed but overlapping confidence intervals of utility strongly indicate this conclusion.

Panel tests were the most preferred option across all 8 DCEs, followed by single gene tests. The utility for each test type (and no test) was broadly consistent across

all test types, indicating good face validity in the description of the tests. In the context of the presented scenarios, single gene tests were associated with negative or small positive utility. Not testing was associated with strongly negative utilities in all DCEs. Participants preferred any test to no test.

The results of these DCEs showed clear preferences for reduced risks of ADRs, both from the 'current' medication discussed in the scenario, and any future ADRs. This direction of effect was as expected *a priori* and provides further evidence that participants clearly understood the choice tasks.

The attributes 'number of medicines the test can inform' and 'cost of test' had small β-coefficients but were associated with p-values <0.05, indicating they did have small impact on participant choices. It is important to note that the cost of genetic testing continues to fall, and the price of panel testing can be comparable to the cost of single gene testing in some cases (15, 21).

Including a monetary attribute in a DCE allows for the estimation of the willingness to pay of the other attributes (32). However, the range of costs for genetic testing were not as wide as I initially expected when it was decided to include a cost attribute. The estimate of these costs was based on the cost of one sequencing technology. Further work should involve work with NHS trusts to determine 'real world' costs that include equipment, consumables, and staff time on testing and interpretation. I did not calculate willingness to pay at this stage of analysis, although the data is available to do this for future work. However, the inclusion of the cost attribute is still important for policy decisions, providing an indication of the importance placed on cost by the general public.

The privacy attribute provides interesting results. The largest positive β-coefficients of each of the privacy options were associated with 'anonymous data sharing', in all DCEs. This result was hypothesised due to the content of the focus group discussions (see Chapter 6). Written responses also show the willingness of participants to participate in research. Not sharing data was associated with a negative β-coefficient, proving that this is not just willingness, but a clear preference. This has important implications for policy making in this area. The scenario shown to participants indicated that data sharing would be with 'universities and other researchers.' It was not specified that data would not be shared with companies in the private sector, but participants would likely conclude this from the wording of the scenario. The preferences for privacy in this survey would likely differ if private companies' access to health data was discussed (63).

Participants were asked to rate the DCE on a scale of 1 to 10 (where 1 is 'not difficult at all' and 10 is 'very difficult'). Most participants rated the difficulty as a 1, indicating that the participants understood the task well enough to complete their preferences accurately. This is supported by many of the written responses, an optional extra question collected at the end of the choice tasks.

Data were coded using effects coding. This was chosen over the alternative (dummy coding) as it allows the effect of all levels to be estimated (46). Additionally, perceptions of risk are notoriously unreliable, more so when very low risks are involved (58-61). Effects coding splits risks so they are relative to each other, removing the need for participants to fully appreciate the difference between e.g., 1 in 1,000 and 1 in 10,000.

There is little previous research on how patients and the general public regard level of evidence. This may be due to the phenomenon of publishing bias – negative results are less likely to be published than positive ones. Studies finding no effect of level of evidence on patient preferences may have been completed but not published. It could also be the case that this research question has not been previously investigated.

7.4.1 Limitations

This is likely to be the first DCE that examines preferences for level of evidence in the general public, and as such there are several adaptations to the method that could be made for future experiments.

Participants may have struggled to retain the large amount of information contained at the start of the survey. This was indicated in some of the written responses. One solution used by another DCE was allowing participants to view descriptions again by positioning their cursor over text where required (64). Unfortunately, this was not available within the JISC software. This would also allow the further emphasis of the level of evidence for each test.

The use of animations to explain scenarios to patients significantly reduces the number of random choices, improving choice consistency (65). This is something that was considered, but ultimately decided against due to budgetary constraints. With hindsight, and considering the written feedback from participants, this would have been a useful addition to the DCE.

The age groups represented in the survey appear to be broadly representative of the UK population (62), however this was not formally assessed.

Some technical aspects of DCE design and implementation may have been improved upon. For example, levels for each gene-drug combination's risk of ADRs were chosen to include the confidence intervals around these estimates. There is therefore some uncertainty in these estimates, and therefore also around the levels chosen for the 'risk of ADR from this medicine' attribute. However, this uncertainty is minimised by the use of effects coding, effectively converting risks into 'high', 'medium', and 'low' in the utility model.

In one of the DCEs (atorvastatin), researcher error meant that some choice sets were not made compulsory. Only a small number of participants subsequently skipped choice sets (0.34% of total answers). This error is therefore unlikely to have a large impact on the results of this DCE.

From the optional written responses section of the DCE, some feedback was received that indicated some misunderstanding of the survey. While this was a small proportion of the overall sample, there is interesting learning here around how complex scenarios can be explained to participants. For example, some participants believed that the genetic test itself would cause the stated ADR. One way that other DCEs have combatted this is to insert a comprehension question before the start of choice tasks, to ensure participants fully understand the scenario given (66-69). Analysis can then be undertaken including or excluding participants that fail this comprehension question. I decided against including this in order to minimise participant burden. I would consider adding one in any future DCEs to guard against misunderstandings. The fact that most participants rated the difficulty of the DCE as 'not difficult at all' indicates a reasonable level of understanding, but there is room for improvement on this.

Another way of doing this is to explicitly define a dominant choice and define participants' understanding by whether they chose the option with the highest utility (30). In this case, the dominant choice was difficult to define. While it was assumed that participants would always choose the lowest risk of ADRs, the direction of effect for privacy and cost attributes was less clear *a priori.* More extensive piloting may have resolved this issue and allowed the inclusion of a dominant choice.

7.5 Conclusion

The increasing use of pharmacogenetics to prevent ADRs necessitates policy changes that need to be informed by the preferences of the general public. Incorporating these preferences is essential, not only morally, but may also increase uptake and adherence to treatments (13, 14).

This successful application of the DCE method shows a general positive response from the general public on the topic of genetic testing to prevent ADRs. Panel tests were preferred, and utility modelling indicated willingness to share data for research. However, the level of evidence for a test did not affect the total utility of testing among the general public, a finding that held true across test types and DCE scenarios.

7.6 References

1. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ. 2004;329(7456):15-9.
2. Veeren JC, Weiss M. Trends in emergency hospital admissions in England due to adverse drug reactions: 2008-2015. Journal of Pharmaceutical Health Services Research. 2017;8(1):5-11.
3. Ma JD, Lee KC, Kuo GM. HLA-B*5701 testing to predict abacavir hypersensitivity. PLoS Curr. 2010;2:RRN1203.
4. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, et al. HLA-B*5701 screening for hypersensitivity to abacavir. The New England journal of medicine. 2008;358(6):568-79.
5. Martin MA, Hoffman JM, Freimuth RR, Klein TE, Dong BJ, Pirmohamed M, et al. Clinical Pharmacogenetics Implementation Consortium Guidelines for HLA-B Genotype and Abacavir Dosing: 2014 update. Clinical pharmacology and therapeutics. 2014;95(5):499-500.
6. Caudle KE, Thorn CF, Klein TE, Swen JJ, McLeod HL, Diasio RB, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. Clinical pharmacology and therapeutics. 2013;94(6):640-5.
7. Lunenburg CA, van Staveren MC, Gelderblom H, Guchelaar HJ, Swen JJ. Evaluation of clinical implementation of prospective DPYD genotyping in 5-fluorouracil- or capecitabine-treated patients. Pharmacogenomics. 2016;17(7):721-9.
8. Rollinson V, Turner R, Pirmohamed M. Pharmacogenomics for Primary Care: An Overview. Genes (Basel). 2020;11(11):1337.
9. Bridges JFP, Kinter ET, Kidane L, Heinzen RR, McCormick C. Things are Looking up Since We Started Listening to Patients. The Patient: Patient-Centered Outcomes Research. 2008;1(4):273-82.
10. Whichello C, Bywall KS, Mauer J, Stephen W, Cleemput I, Pinto CA, et al. An overview of critical decision-points in the medical product lifecycle: Where to

include patient preference information in the decision-making process? Health policy (Amsterdam, Netherlands). 2020;124(12):1325-32.
11. Food and Drug Administration (US). Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling: Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders 2016 [Available from: https://www.fda.gov/media/92593/download.
12. Bourke SM, Plumpton CO, Hughes DA. Societal Preferences for Funding Orphan Drugs in the United Kingdom: An Application of Person Trade-Off and Discrete Choice Experiment Methods. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2018;21(5):538-46.
13. Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. JAMA. 2008;300(4):436-8.
14. Blackwood J, Armstrong MJ, Schaefer C, Graham ID, Knaapen L, Straus SE, et al. How do guideline developers identify, incorporate and report patient preferences? An international cross-sectional survey. BMC health services research [Internet]. 2020 2020/05//; 20(1):[458 p.]. Available from: http://europepmc.org/abstract/MED/32448198

https://doi.org/10.1186/s12913-020-05343-x

https://europepmc.org/articles/PMC7247137

https://europepmc.org/articles/PMC7247137?pdf=render.
15. Lynce F, Isaacs C. How Far Do We Go With Genetic Evaluation? Gene, Panel, and Tumor Testing. Am Soc Clin Oncol Educ Book. 2016;35:e72-8.
16. Chinthapalli K. NHS pilots new genetic testing service for cancer patients. BMJ. 2013;346:f3323.
17. Hughes D. In: Johnson D, editor. Number of genes in a proposed NHS gene panel (29) ed2021.
18. Kimpton JE, Carey IM, Threapleton CJD, Robinson A, Harris T, Cook DG, et al. Longitudinal exposure of English primary care patients to pharmacogenomic drugs: An analysis to inform design of pre-emptive pharmacogenomic testing. British journal of clinical pharmacology. 2019;85(12):2734-46.
19. Genelex. Genelex PGx Testing [Available from: https://www.genelex.com/test-menu/.
20. Yorczyk A, Robinson LS, Ross TS. Use of panel tests in place of single gene tests in the cancer genetics clinic. Clinical genetics. 2015;88(3):278-82.
21. Taylor A, Brady AF, Frayling IM, Hanson H, Tischkowitz M, Turnbull C, et al. Consensus for genes to be included on cancer panel tests offered by UK genetics services: guidelines of the UK Cancer Genetics Group. J Med Genet. 2018;55(6):372-7.
22. Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, et al. Gene-panel sequencing and the prediction of breast-cancer risk. The New England journal of medicine. 2015;372(23):2243-57.
23. Robson ME, Bradbury AR, Arun B, Domchek SM, Ford JM, Hampel HL, et al. American Society of Clinical Oncology Policy Statement Update: Genetic and Genomic Testing for Cancer Susceptibility. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2015;33(31):3660-7.
24. Robson ME, Storm CD, Weitzel J, Wollins DS, Offit K, American Society of Clinical O. American Society of Clinical Oncology policy statement update: genetic and genomic testing for cancer susceptibility. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(5):893-901.

25. Frizzo-Barker J, Chow-White PA, Charters A, Ha D. Genomic big data and privacy: challenges and opportunities for precision medicine. Computer Supported Cooperative Work (CSCW). 2016;25(2-3):115-36.
26. Gillis NK, Innocenti F. Evidence required to demonstrate clinical utility of pharmacogenetic testing: the debate continues. Clinical pharmacology and therapeutics. 2014;96(6):655-7.
27. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883-902.
28. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics. 2008;26(8):661-77.
29. Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Muhlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013;16(1):3-13.
30. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Applied health economics and health policy. 2003;2(1):55-64.
31. Ryan M, Gerard K, Amaya-Amaya M. Discrete Choice Experiments in a Nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care. The Economics of Non-Market Goods and Resources. Dordrecht: Springer Netherlands; 2008. p. 13-46.
32. Ryan M, Kolstad JR, Rockers PC, Dolea C. How to conduct a discrete choice experiment for health workforce recruitment and retention in remote and rural areas: a user guide with case studies. The World Bank; 2012.
33. Bien DR, Danner M, Vennedey V, Civello D, Evers SM, Hiligsmann M. Patients' Preferences for Outcome, Process and Cost Attributes in Cancer Treatment: A Systematic Review of Discrete Choice Experiments. Patient. 2017;10(5):553-65.
34. Boeri M, McMichael AJ, Kane JPM, O'Neill FA, Kee F. Physician-Specific Maximum Acceptable Risk in Personalized Medicine: Implications for Medical Decision Making. Medical decision making : an international journal of the Society for Medical Decision Making. 2018;38(5):593-600.
35. Guerra RL, Castaneda L, de Albuquerque RCR, Ferreira CBT, Correa FM, Fernandes RRA, et al. Patient Preferences for Breast Cancer Treatment Interventions: A Systematic Review of Discrete Choice Experiments. Patient. 2019;12(6):559-69.
36. Powell G, Holmes EA, Plumpton CO, Ring A, Baker GA, Jacoby A, et al. Pharmacogenetic testing prior to carbamazepine treatment of epilepsy: patients' and physicians' preferences for testing and service delivery. British journal of clinical pharmacology. 2015;80(5):1149-59.
37. Muhlbacher A, Johnson FR. Choice Experiments to Quantify Preferences for Health and Healthcare: State of the Practice. Applied health economics and health policy. 2016;14(3):253-66.
38. Lancaster KJ. A New Approach to Consumer Theory. Journal of Political Economy. 1966;74(2):132-57.
39. Street DJ, Burgess L, Viney R, Louviere J. Designing Discrete Choice Experiments for Health Care. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care. The Economics of Non-Market Goods and Resources. Dordrecht: Springer Netherlands; 2008. p. 47-72.
40. Bilendi. Bilendi 2018 [Available from: https://www.bilendi.co.uk/.

41. Jisc. Online surveys (formerly BOS) 2021 [Available from: https://www.onlinesurveys.ac.uk/.
42. ChoiceMetrics. Ngene 1.1.1. Australia2012.
43. Hill M, Johnson JA, Langlois S, Lee H, Winsor S, Dineley B, et al. Preferences for prenatal tests for Down syndrome: an international comparison of the views of pregnant women and health professionals. Eur J Hum Genet. 2016;24(7):968-75.
44. Orme B. Sample size issues for conjoint analysis studies. Sawthooth Software Research paper Series Squim, WA, USA: Sawthooth Software Inc. 1998.
45. Louviere JJ, Hensher DA, Swait JD, Adamowicz W. Stated Choice Methods : Analysis and Applications. Cambridge, UNITED KINGDOM: Cambridge University Press; 2000.
46. Bech M, Gyrd-Hansen D. Effects coding in discrete choice experiments. Health Econ. 2005;14(10):1079-83.
47. Ryan M, Gerard K, Amaya-Amaya M. Using discrete choice experiments to value health and health care: Springer Science & Business Media; 2007.
48. Vass CM, Barton A, Payne K. Towards Personalising the Use of Biologics in Rheumatoid Arthritis: A Discrete Choice Experiment. Patient. 2022;15(1):109-19.
49. Holmes EAF, Plumpton C, Baker GA, Jacoby A, Ring A, Williamson P, et al. Patient-Focused Drug Development Methods for Benefit-Risk Assessments: A Case Study Using a Discrete Choice Experiment for Antiepileptic Drugs. Clinical pharmacology and therapeutics. 2019;105(3):672-83.
50. StataCorp. Stata Statistical Softare: Release 14. College Station, TX: StataCorp LP; 2015.
51. Payne K, Fargher EA, Roberts SA, Tricker K, Elliott RA, Ratcliffe J, et al. Valuing pharmacogenetic testing services: a comparison of patients' and health care professionals' preferences. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2011;14(1):121-34.
52. R Studio Team. RStudio: Integrated Development for R. RStudio, Inc. Boston, MA2016.
53. StataCorp. xtlogit — Fixed-effects, random-effects, and population-averaged logit models.
54. Scobie S, Castle-Clarke S. Implementing learning health systems in the UK NHS: Policy actions to improve collaboration and transparency and support innovation and better use of analytics. Learn Health Syst. 2020;4(1):e10209.
55. Evans D. Patient and public involvement in research in the English NHS: a documentary analysis of the complex interplay of evidence and policy. Evidence & Policy: A Journal of Research, Debate and Practice. 2014;10(3):361-77.
56. Evans H. Using data in the NHS: the implications of the opt-out and GDPR. King's Fund website www kingsfund org uk/publications/using-data-nhs-gdpr. 2018.
57. Illumina. Infinium Global Screening Array-24 Kit 2020 [Available from: https://emea.illumina.com/products/by-type/microarray-kits/infinium-global-screening.html?langsel=/gb/.
58. Schmälzle R, Renner B, Schupp HT. Health Risk Perception and Risk Communication. Policy Insights from the Behavioral and Brain Sciences. 2017;4(2):163-9.
59. Ferrer R, Klein WM. Risk perceptions and health behavior. Curr Opin Psychol. 2015;5:85-9.
60. Frewer L. The public and effective risk communication. Toxicol Lett. 2004;149(1-3):391-7.
61. Waters EA, Fagerlin A, Zikmund-Fisher BJ. Overcoming the Many Pitfalls of Communicating Risk. In: Diefenbach MA, Miller-Halegoua S, Bowen DJ, editors. Handbook of Health Decision Science. New York, NY: Springer New York; 2016. p. 265-77.

62. Office for National Statistics U. Overview of the UK population: January 2021 2021 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/january2021.
63. Braunack-Mayer A, Fabrianesi B, Street J, O'Shaughnessy P, Carter SM, Engelen L, et al. Sharing Government Health Data With the Private Sector: Community Attitudes Survey. J Med Internet Res. 2021;23(10):e24200.
64. Smith ML, White CB, Railey E, Sledge GW, Jr. Examining and predicting drug preferences of patients with metastatic breast cancer: using conjoint analysis to examine attributes of paclitaxel and capecitabine. Breast Cancer Res Treat. 2014;145(1):83-9.
65. Vass CM, Davison NJ, Vander Stichele G, Payne K. A Picture is Worth a Thousand Words: The Role of Survey Training Materials in Stated-Preference Studies. Patient. 2020;13(2):163-73.
66. Mansfield C, Poulos C, Boeri M, Hauber B. Performance of a Comprehension Question in Discrete-Choice Experiment Surveys. ISPOR Europe 2019; Copenhagen, Denmark2019.
67. Chan SL, Wen Low JJ, Lim YW, Finkelstein E, Farooqui MA, Chia KS, et al. Willingness-to-pay and preferences for warfarin pharmacogenetic testing in Chinese warfarin patients and the Chinese general public. Per Med. 2013;10(2):127-37.
68. Pearce A, Harrison M, Watson V, Street DJ, Howard K, Bansback N, et al. Respondent Understanding in Discrete Choice Experiments: A Scoping Review. Patient. 2021;14(1):17-53.
69. Pearce AM, Mulhern BJ, Watson V, Viney RC. How Are Debriefing Questions Used in Health Discrete Choice Experiments? An Online Survey. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2020;23(3):289-93.

8: Conclusion

It has long been known that individual patients can respond very differently to the same medications. Standard drug choice and dosing does not work for all patients. The rise of personalised medicine has enabled greater precision in choosing drugs and doses to maximise drug efficacy and minimise the risk of harmful adverse drug reactions (ADRs). A notable success story is the one of *HLA-B*57:01*, HIV drug abacavir, and abacavir hypersensitivity syndrome (AHS). The link between this gene, abacavir, and AHS was first reported in 2002 (1), and since then genetic testing prior to abacavir prescription has become mandatory in many jurisdictions (2-5). This has resulted in the near elimination of AHS where genetic testing is available (6), improving HIV treatment for all patients.

Since one of the earliest drugs specifically designed with a pharmacogenetic component (trastuzumab) was released in 1998, there has also been a large increase in the number of pharmacogenetic drugs prescribed in the UK (7). Chapter 1 of this thesis discusses how pharmacogenetics is known to reduce the risk of ADRs, improve drug efficacy, and can improve the process of drug development. However, several impediments exist to the wider implementation of these pharmacogenetic advances.

Aside from the cost of genetic testing (which does continue to fall), and technical issues (such as adapting old systems and health records), the main issue impeding the progress of pharmacogenetics is a lack of suitable evidence (8, 9). The evidence required for clinical use is often at least one randomised controlled trial (RCT). This presents three problems, that will be addressed here.

Firstly, regulatory agencies' guidelines for the evidence required for the approval of a pharmacogenetic test are complex and often out-dated. Regulation is required in order to protect patients and the general public but the rapid speed of development in the field of pharmacogenetics makes it difficult to keep regulatory guidelines relevant. The US Food and Drug Administration (FDA), the world's largest drug regulatory authority, acknowledged that these new technologies place a 'strain' on its regulatory process (10).

The regulatory process includes assessment of analytic validity, clinical utility, and clinical validity. Several frameworks have been proposed to assess these parameters. Chapter 2 of this thesis discusses some of these frameworks (ACCE (11), PhRMA (12), Personalised Precision Medicine Special Interest Group (13)),

alongside guidelines issued by the UK Medicines and Healthcare products Regulatory Authority (MHRA) and the US FDA. I also contacted the MHRA for details of upcoming updates to the guidelines for medical devices and in vitro diagnostic medical devices (the category for most pharmacogenetic tests). New trial designs and statistical analysis plans are also needed at this new frontier. Previous work by this group has begun to address this problem (14). Another issue confronted by regulatory authorities is the need for patient and public involvement in setting regulatory standards. Not only is this correct from a moral standpoint, inclusion can also improve the benefit-risk profile of decision making (15).
It is clear that there is no overall standard for the evidence required for pharmacogenetic testing or implementation. The main conclusion from Chapter 2 is that there is a need for a unified set of standards, work that could be led by existing evidence collectors such as the Pharmacogenomics Knowledgebase (PharmGKB) and the Clinical Pharmacogenetics Implementation Consortium (CPIC) (see below).

This issue is explored further in Chapter 3. Parts of this chapter were published in a 2019 paper (16). The lack of regulatory guidance has led to many different interpretations of the level of evidence required before including a pharmacogenetic biomarker to guide treatment in an RCT. Five such RCTs were evaluated and I explored the types and strength of evidence that each one used to justify the inclusion of the tested biomarker within the trial. I also discussed the timing of evidence compared to the start of, or publication of, the trial. Although labour intensive, this was a useful exercise that demonstrated that there is no standard approach to collating such evidence, with many different types included. Although a previous RCT is the ‘gold standard’ of evidence, this is not possible in many cases. Based on the literature review, three recommendations for future work were made. First, as explored in Chapter 2, more guidance from regulatory authorities is needed. This would ensure rigorous evidence standards are adhered to both before implementing pharmacogenetic testing in a trial, and later in clinical practice. Second, future work should consider that pharmacogenetic interventions require validation in groups from diverse ancestries. Current research is predominantly in white and Western populations (17, 18), making the evidence for some pharmacogenetic interventions weaker in those from other populations. Finally, it was recommended that a systematic review is undertaken before the start of an RCT with a pharmacogenetic component. Although labour intensive, this would ensure that all the available evidence relating to a pharmacogenetic intervention has been evaluated prior to the trial. This review may even form its own evidence,

negating the need for a trial. This concept, of observational evidence providing output and precision similar to that of a prospective trial, is discussed further in Chapter 4.

A second issue relating to availability of evidence to support pharmacogenetic interventions is that RCTs (the evidence 'gold standard') can be difficult to perform in very rare or complex conditions (19-22). The chapter is based on three key papers (Concato, *et al.* 2000 (23), Golder, *et al.* 2011 (24), and Benson & Hartz, 2000 (25)) that show that effect estimates obtained from observational evidence can be similar to, and as precise as, those obtained from prospective studies. The other key paper of this chapter is Tonk, *et al.* (2016) (26), who showed that the sensitivity, specificity, positive predictive value and negative predictive value of a genetic test can be calculated using observational studies by incorporating the frequency of the genetic variant and the prevalence of the event it has been designed to predict (e.g. the ADR). These important pieces of evidence for the clinical implementation of a pharmacogenetic test were calculated alongside effect estimates in the case of *HLA-B*15:02* and carbamazepine-induced Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN). Through a systematic review of the literature and subsequent meta-analysis (much required in this field, as the last that could be located was published in 2014) data was pooled from 437 cases and 1,717 controls. Estimates obtained from the meta-analysis were compared to the effect estimates and estimates of clinical validity from a prospective trial conducted in 2011 (27). The precision of these estimates was much higher in the observational data than the prospective data, indicating the potential usefulness of observational data in evidence gathering. This analysis was taken further with a novel systematic review and meta-analysis of *HLA-A*31:01* and carbamazepine-induced SJS/TEN. This less-studied risk allele had less literature associated with the association, and no prospective study. Therefore, estimates from pooled observational data (196 cases and 1,677 controls) were compared to a simulated prospective clinical trial, with similar parameters to the pivotal Chen, *et al.* trial in *HLA-B*15:02*. Again, I found that the precision of the effect estimates and measures of clinical validity were greater in the observational dataset. These analyses demonstrate a novel way of evidence gathering in cases where prospective trials cannot be performed, and suggest that in some cases, they do not need to be performed. Wider use of observational data in this manner could greatly impact the regulatory and implementation challenges of pharmacogenetics. Both regulatory authorities and

clinicians should consider the use of observational data for the approval of pharmacogenetic testing.

The third and final issue related to evidence on pharmacogenetic tests is that the level of evidence required for a test to be accepted by the general public is not known. As the public are the potential primary users of pharmacogenetic testing, this information is essential for policy making. The inclusion of these preferences can also potentially increase acceptance and test uptake when they are used in clinic (15). I chose to illustrate this with examples of pharmacogenetic tests used to prevent ADRs.

This issue was explored using the discrete choice experiment (DCE) method. To gain an overview of the field, I first conducted a systematic review of existing DCEs in pharmacogenetics. The scope of the review was narrowed to only include DCEs with an ADR component, to make the review more relevant to my aims. This systematic review was also useful for learning more about the DCE method, one primarily used in health economics contexts (28, 29). Chapter 5 presents an overview of 13 DCEs located in a systematic search of the literature. After summarising the findings, I began examining the design, methods, and results of each DCE and extracted useful lessons from each. The qualitative methods of each DCE, details of their administration, and how they presented information to participants, were particularly useful.

Details of the qualitative work performed by other DCEs influenced Chapter 6, which summarises the qualitative work that underpins my own DCE. An extensive program of qualitative work was undertaken that included surveys, focus groups, and pilot testing. For example, the most common qualitative work used by DCEs of the systematic review was gathering expert opinion. I therefore began my own qualitative work with a survey of healthcare professionals and academics (n=17). Some valuable insights were gained from those on this side of pharmacogenetic testing. Written feedback from participants also provided good starting points for further qualitative work, beginning with a survey delivered to patients. Although this was a small patient group (n=20), it provided a valuable perspective for the development of a DCE that would be accessible and relevant to the general public and policy makers. The final DCE conducted was in the general public, but I felt it important to also gain the perspective of people with more extensive public experience of the healthcare system.

From this, I conducted focus groups with my target population, the general public, to gauge their current understanding of pharmacogenetic research and their views on potential future testing. Although cut short due to the start of the COVID-19 pandemic, these produced many insights not previously considered, e.g., altruistic motivations for sharing healthcare data for research. These groups also highly valued data privacy and anonymisation, aspects that were taken forward into the final DCE.

A small pilot study in the general public (n=16) confirmed that participants both understood the attributes of the DCE and were willing to trade on them. This also provided informal feedback that some potential participants may find the subject matter (e.g., imagine you have been diagnosed with a disease) upsetting. From this, I ensured to warn participants of this wording in a landing page before they consented to participate, and provided information and links to helpful resources.

The end of Chapter 6 shows how this qualitative work was combined to produce a DCE design that indirectly tests how the level of evidence for a pharmacogenetic intervention to prevent ADRs affects participant preferences and test utility. Eight different drug-gene-ADR combinations were chosen across four indications (HIV, cancer, epilepsy, and cardiovascular disease) and a DCE was created incorporating each. Combinations with ‘high’ and ‘low’ levels of evidence for their clinical utility were paired within each indication. The risk of each ADR was then calculated with and without genetic testing, in order to choose levels for this attribute that would accurately reflect real world scenarios. An online system was then used to randomise participants to one of the 8 DCEs. I also included the choice of genetic panel tests, single gene tests, and compared them to not testing, as this is a currently highly relevant policy issue. Genetic testing is currently moving into an era of panel testing, since in many cases panel tests do not cost significantly more than single gene tests while providing more information (30, 31).

The DCE was launched in May 2021, and Chapter 7 presents the results of these experiments. A UK-representative sample of 2,019 participants was recruited. Most (87.6%) had never had a genetic test before. The coefficients calculated in a random effects model showed the expected direction of effect for all attributes across all eight DCEs. Utility was calculated using a preference-weight utility model for each DCE pair. Comparing the utility of testing in scenarios with ‘high’ and ‘low’ evidence of clinical utility showed considerable overlap, indicating there was likely to be no difference in utility between them. Not testing was always associated with

negative utility, showing that the general public views genetic testing to prevent ADRs favourably. As suggested by the qualitative work, participants' most preferred option for data sharing and privacy was for their data to be shared anonymously for research. This was preferred over full data sharing, and also over not sharing any data. This is a highly positive finding for policy makers and researchers in the future. This DCE did not find any difference in the utility at differing levels of evidence. This may indicate that the level of evidence is not as significant to the general public's views of genetic testing as issues of privacy and ADR risk reductions. However, this result needs to be confirmed by directly asking participants about level of evidence (see below).

This thesis has explored the issue of evidence gathering in pharmacogenetics from three angles – from the perspective of regulatory bodies, the use of observational data as evidence, and from the perspective of the general public. I have identified issues within the field and recommended improvements and solutions for the future. This work provides guidance to policy makers and other stakeholders that will be valuable as the use of pharmacogenetics continues to grow.

8.1 Future directions

While the field of pharmacogenetics continues to grow rapidly, implementation of new technologies at the clinical level has been slow. As one the main reasons for this is a lack of suitable evidence (8, 9), several important pieces of future research detailed below have the potential to improve the level of implementation of pharmacogenetics at the patient level.

Chapters 2 and 3 detail how a lack of regulatory guidance on evidence has led to a patchwork of recommendations, guidelines, and frameworks on the issue, along with varied justifications for biomarker use by trials. A thorough systematic review of regulatory agencies' policies on pharmacogenetics would provide a valuable overview, allowing researchers to see what is missing in evidentiary standards. This research should include the input of regulatory specialists and that of policy makers. A qualitative component of this review, interviewing these stakeholders for their views on evidence, would be a useful exercise that could identify further relevant issues. The work in this thesis focussed on the systems of the UK and USA, but with sufficient funding, this research could also include the regulatory bodies of many more jurisdictions.

In addition to this, the work of PharmGKB and CPIC should be continued to further refine and consolidate existing evidence on pharmacogenetic biomarkers. These bodies are similarly well-placed to participate in an evidence-setting process with regulatory agencies. A unified set of evidence standards, with input from patients, the general public, academics, clinicians, policy makers, payers, and regulatory bodies would be a ground-breaking addition to the field of pharmacogenetics. These standards would inform the development of new pharmacogenetic drugs and accompanying tests. Drug companies would have an idea of the costs and duration of pharmacogenetic drug development, and both clinicians and patients could be reassured that their products have met minimum evidentiary standards.

A final piece of work that would validate the findings of the DCE is another DCE where the general public are directly asked about their preferences for the level of evidence for a pharmacogenetic intervention. The 'high' and 'low' evidence gene-drug-ADR combinations were compared indirectly, by randomising patients between combinations. Although my findings suggest that there is little difference in utility between combinations, this should be confirmed with a follow-up survey with direct questioning. Extensive planning and qualitative work are required for this project, to ensure the concept of 'level of evidence' is adequately explained without too high of a participant burden.

As a final point to this thesis, this section presents details of an ongoing trial that I believe will have a large impact on future pharmacogenetic practice. This trial, by the European Ubiquitous Pharmacogenomics Consortium (U-PGx) is examining the effect of implementing pharmacogenetic testing in 7 countries (32). The trial is recruiting patients who receive a prescription for one of 41 drugs, chosen because of their inclusion in Dutch Pharmacogenomics Working Group guidelines. The trial uses a cross-over design, with countries as the unit of randomisation. Each country is randomised to use pharmacogenetic-informed prescribing or standard of care for 18 months, then crossed-over. The primary outcome is the occurrence of ADRs, with quality of life, adherence, and health expenditure among some of the secondary outcomes. This trial is part of a wider initiative to better integrate pharmacogenetics into clinical use, including improving the use of electronic health records, clinical decision support systems, and examining the ethical, legal, and societal implications of pharmacogenetics (33). This initiative provides a blueprint for a complete approach to improve the implementation of pharmacogenetics, while also integrating a massive trial with a novel design.

8.2 References

1. Hetherington S, Hughes AR, Mosteller M, Shortino D, Baker KL, Spreen W, et al. Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. Lancet. 2002;359(9312):1121-2.
2. British National Formulary. ABACAVIR [Available from: https://bnf.nice.org.uk/drug/abacavir.html#preTreatmentScreeningInformation.
3. Medicines and Healthcare products Regulatory Agency. Summary of Product Characteristics: Abacavir 2019 [Available from: https://mhraproducts4853.blob.core.windows.net/docs/449e73eeb73d9e299df5ecb747d94ed3bd702036.
4. Martin MA, Hoffman JM, Freimuth RR, Klein TE, Dong BJ, Pirmohamed M, et al. Clinical Pharmacogenetics Implementation Consortium Guidelines for HLA-B Genotype and Abacavir Dosing: 2014 update. Clinical pharmacology and therapeutics. 2014;95(5):499-500.
5. Phillips E, Mallal S. Successful translation of pharmacogenetics into the clinic: the abacavir example. Mol Diagn Ther. 2009;13(1):1-9.
6. Mounzer K, Hsu R, Fusco JS, Brunet L, Henegar CE, Vannappagari V, et al. HLA-B*57:01 screening and hypersensitivity reaction to abacavir between 1999 and 2016 in the OPERA((R)) observational database: a cohort study. AIDS Res Ther. 2019;16(1):1.
7. Kimpton JE, Carey IM, Threapleton CJD, Robinson A, Harris T, Cook DG, et al. Longitudinal exposure of English primary care patients to pharmacogenomic drugs: An analysis to inform design of pre-emptive pharmacogenomic testing. British journal of clinical pharmacology. 2019;85(12):2734-46.
8. McKinnon RA, Ward MB, Sorich MJ. A critical analysis of barriers to the clinical implementation of pharmacogenomics. Ther Clin Risk Manag. 2007;3(5):751-9.
9. Klein ME, Parvez MM, Shin JG. Clinical Implementation of Pharmacogenomics for Personalized Precision Medicine: Barriers and Solutions. J Pharm Sci. 2017;106(9):2368-79.
10. Food and Drug Administration (US). Optimizing FDA's Regulatory Oversight of Next Generation Sequencing Diagnostic Tests—Preliminary Discussion Paper 2014 [Available from: https://www.fda.gov/media/90403/download.
11. Centers for Disease Control and Prevention (US). ACCE Model Process for Evaluating Genetic Tests 2010 [Available from: https://www.cdc.gov/genomics/gtesting/acce/index.htm.
12. Altar CA, Amakye D, Bounos D, Bloom J, Clack G, Dean R, et al. A prototypical process for creating evidentiary standards for biomarkers and diagnostics. Clinical pharmacology and therapeutics. 2008;83(2):368-71.
13. Faulkner E, Holtorf AP, Walton S, Liu CY, Lin H, Biltaj E, et al. Being Precise About Precision Medicine: What Should Value Frameworks Incorporate to Address Precision Medicine? A Report of the Personalized Precision Medicine Special Interest Group. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2020;23(5):529-39.
14. Antoniou M, Jorgensen AL, Kolamunnage-Dona R. Biomarker-guided trial designs (BiGTeD): An online tool to help develop personalised medicine [Available from: http://www.bigted.org/.
15. Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. JAMA. 2008;300(4):436-8.

16. Johnson D, Hughes D, Pirmohamed M, Jorgensen A. Evidence to Support Inclusion of Pharmacogenetic Biomarkers in Randomised Controlled Trials. J Pers Med. 2019;9(3):42.
17. Drozda K, Wong S, Patel SR, Bress AP, Nutescu EA, Kittles RA, et al. Poor warfarin dose prediction with pharmacogenetic algorithms that exclude genotypes important for African Americans. Pharmacogenetics and genomics. 2015;25(2):73-81.
18. Johnson D, Wilke MA, Lyle SM, Kowalec K, Jorgensen A, Wright GE, et al. A systematic review and analysis of the use of polygenic scores in pharmacogenomics. Clinical pharmacology and therapeutics. 2021;n/a(n/a).
19. Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. Statistics in medicine. 2014;33(24):4186-201.
20. Antoniou M, Kolamunnage-Dona R, Wason J, Bathia R, Billingham C, Bliss JM, et al. Biomarker-guided trials: Challenges in practice. Contemp Clin Trials Commun. 2019;16:100493.
21. Billingham L, Malottki K, Steven N. Small sample sizes in clinical trials: a statistician's perspective. Clinical Investigation. 2012;2(7):655-7.
22. Miller F, Zohar S, Stallard N, Madan J, Posch M, Hee SW, et al. Approaches to sample size calculation for clinical trials in rare diseases. Pharm Stat. 2018;17(3):214-30.
23. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. The New England journal of medicine. 2000;342(25):1887-92.
24. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. PLoS Med. 2011;8(5):e1001026.
25. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. The New England journal of medicine. 2000;342(25):1878-86.
26. Tonk ECM, Gurwitz D, Maitland-van der Zee AH, Janssens A. Assessment of pharmacogenetic tests: presenting measures of clinical validity and potential population impact in association studies. Pharmacogenomics J. 2017;17(4):386-92.
27. Chen P, Lin JJ, Lu CS, Ong CT, Hsieh PF, Yang CC, et al. Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan. The New England journal of medicine. 2011;364(12):1126-33.
28. Ryan M, Gerard K, Amaya-Amaya M. Discrete Choice Experiments in a Nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care. The Economics of Non-Market Goods and Resources. Dordrecht: Springer Netherlands; 2008. p. 13-46.
29. Bryan S, Dolan P. Discrete choice experiments in health economics. For better or for worse? Eur J Health Econ. 2004;5(3):199-202.
30. Lynce F, Isaacs C. How Far Do We Go With Genetic Evaluation? Gene, Panel, and Tumor Testing. Am Soc Clin Oncol Educ Book. 2016;35:e72-8.
31. Chinthapalli K. NHS pilots new genetic testing service for cancer patients. BMJ. 2013;346:f3323.
32. van der Wouden CH, Cambon-Thomsen A, Cecchin E, Cheung KC, Davila-Fajardo CL, Deneer VH, et al. Implementing Pharmacogenomics in Europe: Design and Implementation Strategy of the Ubiquitous Pharmacogenomics Consortium. Clinical pharmacology and therapeutics. 2017;101(3):341-58.
33. Manson LE, van der Wouden CH, Swen JJ, Guchelaar HJ. The Ubiquitous Pharmacogenomics consortium: making effective treatment optimization accessible to every European citizen. Pharmacogenomics. 2017;18(11):1041-5.