Economic evaluations of pharmacogenetic and pharmacogenomic screening tests: a systematic review. Second update of the literature

Elizabeth J.J. Berm, Margot de Looff, Bob Wilffert, Cornelis Boersma, Lieven Annemans, Stefan Vegter, Job F.M. van Boven, Maarten J. Postma

ABSTRACT

Objective: Due to extended application of pharmacogenetic and pharmacogenomic screening (PGx) tests it is important to assess whether they provide good value for money. This review provides an update of the literature.

Methods: A literature search was performed in PubMed and papers published between August 2010 and September 2014, investigating the cost-effectiveness of PGx screening tests, were included. Papers from 2000 until July 2010 were included via two previous systematic reviews. Studies’ overall quality was assessed with the Quality of Health Economic Studies (QHES) instrument.

Results: We found 38 studies, which combined with the previous 42 studies resulted in a total of 80 included studies. An average QHES score of 76 was found. Since 2010, more studies were funded by pharmaceutical companies. Most recent studies performed cost-utility analysis, univariate and probabilistic sensitivity analyses, and discussed limitations of their economic evaluations. Most studies indicated favorable cost-effectiveness. Majority of evaluations did not provide information regarding the intrinsic value of the PGx test. There were considerable differences in the costs for PGx testing. Reporting of the direction and magnitude of bias on the cost-effectiveness estimates as well as motivation for the chosen economic model and perspective were frequently missing.

Conclusions: Application of PGx tests was mostly found to be a cost-effective or cost-saving strategy. We found that only the minority of recent pharmacoeconomic evaluations assessed the intrinsic value of the PGx tests. There was an increase in the number of studies and in the reporting of quality associated characteristics. To improve future evaluations, scenario analysis including a broad range of PGx tests costs and equal costs of comparator drugs to assess the intrinsic value of the PGx tests, are recommended. In addition, robust clinical evidence regarding PGx tests’ efficacy remains of utmost importance.
4.1.1. INTRODUCTION

Pharmacogenetics and pharmacogenomics investigate the influence of genetic and genomic variations on drug response in individuals (1). The term pharmacogenetics covers the study of single genes, whereas pharmacogenomics is used to describe the study of several genes (1). The abbreviation PGx is used to cover both pharmacogenetics and pharmacogenomics. PGx tests offer great opportunities for personalised medicine, by combining genetic information and corresponding phenotypes (2). Ideally, by applying PGx, the most optimal, tailored pharmacotherapy can be determined, thereby increasing the treatment’s overall efficacy and decreasing the incidence of adverse events (1). In the field of oncology it has been shown that for certain therapies the specific genetic variations in cancer cells can affect the drug efficacy and/or adverse events (2). Hence, patients may benefit from the PGx tests by utilising an alternative therapy, or changing the drug dosage (3,4). Therefore, PGx is nowadays often used as a synonym for personalised medicine, although personalized medicine is a much broader concept (2).

It is likely that an increasing amount of patient-specific genomic information will be available in the near future and this may result in an increased usage of PGx tests which needs evaluation of effects, but also cost effectiveness (5). PGx has the potential to reduce the costs associated with inappropriate, expensive drug treatments and/or serious adverse drug reactions, in particular those that require hospitalisation (3). Therefore, next to optimising health outcomes, PGx tests might be cost-saving (1,5). However, in order to obtain valid, accurate, and relevant estimates of cost-effectiveness, reliable economic studies are required.

Economic evaluations of PGx tests entail some specific difficulties. Often there is no hard clinical evidence regarding the effects of the PGx test on the clinical utility and it is unlikely that such evidence will be available for the use of every genetic variant (6). Furthermore, for PGx tests, compliance and adherence of clinicians to the test results might have an effect on the effectiveness of PGx tests which is hard to incorporate in a cost-effectiveness analysis (7). Differences in costs for the PGx test can be substantial between countries, or even laboratories, and therefore it is advised to include different costs in scenario analysis (7). In addition, the sensitivity and specificity of a PGx test can vary due to different ethnicities studied, or genetic variations analysed.

In the last decade, several reviews investigated economic evaluations of genetic tests (1,8-14). These reviews showed that the level of consistency and quality could be improved. Many original studies lacked a thorough sensitivity analysis and moreover, in general a poor quality of methodology was noticed (9,11). Inconsistencies mainly resulted from e.g. lack of clinical evidence, different methodologies as well as statistics and modest heterogeneity among study designs and patient populations (3,9). However, these different methodologies have not been in detail dealt with in previous systematic
reviews. Recent developments have led to a bifurcation in the nature of the economic evaluations of PGx testing in, on the one hand, studies assessing the intrinsic value of a test and, on the other hand, studies assessing the value of the test in combination with an active compound. For example in colorectal cancer, the economic value of testing for KRAS as compared to no testing could be considered the “intrinsic value” of the PGx test. KRAS testing before treatment with cetuximab, is found to be dominant (i.e. cost-saving and better) as compared to no prior testing and therefore it is recommended before administration of cetuximab (15). By its uptake in clinical guidelines, a shift in the comparator for future economic evaluations took place, as the combination of cetuximab and KRAS testing became usual care. In future evaluations, the intrinsic value of the KRAS test itself will no longer be assessed, but rather the combination of a drug and its test as compared with a new treatment option. Following this development, a distinction between the nature of economic evaluations of PGx tests is important for a fair comparison of studies.

The objective of this study was to give recommendations for improvement and an update of the literature about PGx tests, taking into account the difference between the intrinsic value of tests themselves and tests embedded into economic evaluations as usual care or best current care. Our new findings were placed in perspective with respect to findings from our previous reviews (1,9). As such, our study links together a period from 2000-September 2014 of PGx testing and pharmacoeconomics.

4.1.2. METHODS

A search in PubMed was performed using combinations of the following terms [PubMed search: all fields] and their thesauri variants: [‘cost-effectiveness [including MeSH]’ OR ‘cost-utility’ OR ‘cost-benefit’ OR ‘cost-minimization’ OR ‘pharmacoeconomics [including MeSH]’] AND [‘pharmaco-genetics’ OR ‘pharmacogenomics [including MeSH]’ OR ‘genotyping’ OR ‘genetic screening’ OR ‘genetic testing [including MeSH]’ OR ‘genotyped’ OR ‘polymorphism screening’]. These search terms were in line with the terms that were used in previous reviews, performed in 2008 and 2010 (1,9). The search was last updated in October 2014 and studies were included if they were: published between August 2010 and September 2014, peer reviewed, performed on a genetic screening method of the human genome, evaluating economic outcomes, written in English, and the genetic or genomic variations were shown to influence the drug efficacy or drug safety. Articles were first screened on title. If the title was not informative enough to form a decision with respect to these criteria, abstracts were assessed. Additional articles were identified through reference tracking.
From the selected studies, the following data were extracted: (I) area of disease or patient population, (II) gene(s) analysed by the pharmacogenetic test, (III) the costs of the pharmacogenetic test, (IV) pharmaceutical compound influenced by the genetic variation, (V) type of economic analysis, (VI) type of sensitivity analysis, (VII) time horizon, (VIII) discounting, (IX) perspective, (X) the outcome measurements, and (XI) the funding body. For interpretation of the outcome measure (i.e. cost-effectiveness), the conclusions as reported by the authors were used. Furthermore, an assessment of the papers ‘discussion on the study limitations’ was made. In this assessment, all limitations mentioned by the authors were captured to look for common and uncommon themes. As stated in the previous review, assessment of these points is expected to provide good information for an adequate interpretation of the studies design, reporting, robustness, methodologies used, and statistical analyses performed (9). In addition to these points, which were assessed in our previous reviews, we added (XII) reporting of analytical validity of the PGx test, (XIII) the cost-effectiveness threshold, (XIV) the country which was used for the perspective of the economic evaluation, and (XV) a weighted quality assessment for the studies included by this update. To assess the quality, the Quality of Health Economic Studies (QHES) instrument was used (16). This instrument was used to improve generalisability of the results with respect to other reviews performed in the same field (10,13). According to the QHES checklist a score between 0 and 100 was generated. A score of ≥ 75 was considered as a high quality score (11). Two reviewers assessed the quality of the included studies, if results were different, consensus was reached through discussion.

Data from before 2008, and for the period 2008 - July 2010 were retrieved from the two previous reviews by Vegter et al. (2008, 2010) (1,9).

4.1.3. RESULTS

Included studies

Results of the search strategy are provided in a PRISMA flow chart (fig. 1) (17). The PubMed search yielded 4408 hits. Duplicates were removed and out of the remaining 733 articles, 160 were selected for full text assessment. Three articles were identified through reference tracking. After inspection of full texts, 122 studies were excluded, resulting in 38 included studies (15,18-54). Main reasons for exclusion were the type of genetic tests studied (i.e. not related to pharmacogenetics) and review papers. Subsequently, 42 studies published before July 2010 were added based on the two reviews by Vegter et al. (1,9,55-96), resulting in a final inclusion of 80 studies. Fig. 2A provides an overview of the total of 80 studies, published from 2000 until September 2014, sorted by the type of pharmacoeconomic analysis performed. Most studies were published in 2012 (16%), with a total of 13 publications.
**Type of analysis: intrinsic value or combination with new drug**

Cost utility analysis (CUA) was the technique mostly applied, namely in 54 studies (68%). Cost-effectiveness analysis (CEA) was performed in 20 studies (25%), cost minimization analysis (CMA) was used 5 times (6%). Note that for the sake of clarity we explicitly differentiate between CUA (with results expressed in cost per QALY) and CEA (with other parameters for effectiveness). Before 2008, CEA was the most frequently applied study type. Since 2008, CUA was performed in most of the publications. Some studies directly assessed the intrinsic value of the PGx test (15,18-22,24-28), while other studies applied scenarios in which equal costs and efficacy were assumed for the drug related to the PGx tests and the alternative treatment (23,29). Both were considered cost-effectiveness estimates that provide an indication of the intrinsic value of the PGx test (table 1A).
Economic evaluations of pharmacogenetic tests.

However, we found that the majority of the newly included (i.e. since 2010) CEAs incorporated a PGx test strategy in combination with a drug and compared this combination to another drug (table 1B). As such, the intrinsic value of the PGx test itself was not assessed. For example, Handorf et al. compared therapy for non-small cell lung cancer with a platinum combination to PGx selected treatment with erlotinib (44). In this analysis, the cost-effectiveness of the PGx test itself was not assessed, but the combination of the PGx test and erlotinib. This makes the outcome mainly dependent on the price of erlotinib. The incremental cost-effectiveness ratios (ICER) of base case scenarios are shown in table 1A and B.

Costs of PGx tests

When comparing the costs of the PGx tests, considerable differences between the costs of the tests were observed. This is not surprising, since technology of genetic testing is developing and costs are likely to be further reduced in the future. This was demonstrated by two studies of which one was performed in 2012 (41) and one in 2013 (40). The costs for this particular screening test, US$72 and £20 respectively, were considerably lower compared to screening costs in studies performed in 2009 and 2010, which ranged from US$175 until US$575, respectively (63,78,96). Another substantial difference was seen in the costs for the CYP2C9 and VKORC1 tests, which ranged from £20 to US$575 (40,78).

Sensitivity analysis

Until 2008, only 5 out of the 31(16%) studies performed both a univariate and probabilistic sensitivity analysis, whereas since 2008 this were 35 out of the 49 (71%) studies (fig. 2B). One of the major uncertainty factors was the lack of robust clinical evidence for clinical utility of the PGx test itself. Hence, almost all authors had to define assumptions which were sometimes solely based on expert opinion. Other frequently mentioned uncertainty factors were the costs of the PGx tests and their real-world utility and performance. As a consequence of the uncertainty around the included genotyping costs, several studies included a costs range in their analysis. For example, Schackman et al. (2013) demonstrated that at a test cost of US$107, genetic testing was not cost-effective. However, at a price of US$10 per test, the PGx test was cost-effective (29). Although variance in genotyping costs was frequently found to have a major impact on the ICER, 11 out of the 38 studies (30%) assessed in this update did not include a range of genotype costs in their sensitivity analysis (S1 Table). With respect to drug costs, some studies showed that drug driven costs did not influence the study’s outcome. For example, Crespin et al. (2011) showed in their sensitivity analysis that even if the costs of the drugs guided by PGx testing dropped substantially, the non-PGx-test-guided drug remained cost-effective (37).
Table 1. Outcomes and funding sources of pharmaco-economic PGx studies on the intrinsic value of PGx test (A) or treatment comparisons involving PGx testing (B), published between August 2010 and September 2014. Note that numbers were rounded towards hundreds.

<table>
<thead>
<tr>
<th>1st Author (reference)</th>
<th>PGx test</th>
<th>Analytical validity PGx test reported</th>
<th>Outcome Measure</th>
<th>Quantitative outcome or ICER (US$)</th>
<th>Cost effectiveness threshold ($/QALY)</th>
<th>Conclusion based on outcome</th>
<th>Funding</th>
<th>Country</th>
<th>QHES score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Klang 2010 (18)</td>
<td>21 gene assay</td>
<td>No</td>
<td>QALY</td>
<td>$10,800</td>
<td>No number</td>
<td>Cost-effective</td>
<td>Teva Pharmaceutical Industries Ltd.</td>
<td>Israel</td>
<td>83</td>
</tr>
<tr>
<td>Bacchi 2010 (19)</td>
<td>21 gene assay</td>
<td>No</td>
<td>Costs</td>
<td>$800 saved per patient</td>
<td>n.a.</td>
<td>Cost-saving</td>
<td>Unknown</td>
<td>Brazil</td>
<td>34</td>
</tr>
<tr>
<td>Hall 2012 (20)</td>
<td>21 gene assay</td>
<td>No</td>
<td>QALY</td>
<td>$8,900</td>
<td>£20,000-30,000 ($31,200-46,700)</td>
<td>Cost-effective, though substantial uncertainty</td>
<td>Academic resources</td>
<td>UK</td>
<td>90</td>
</tr>
<tr>
<td>Vanderlaan 2011 (21)</td>
<td>21 gene assay</td>
<td>No</td>
<td>QALY</td>
<td>$4,4009 saved per patient per year</td>
<td>n.a.</td>
<td>Dominant</td>
<td>Genomic Health, Inc.</td>
<td>USA</td>
<td>31</td>
</tr>
<tr>
<td>Verhoef 2013 (22)</td>
<td>CYP2C9 and VKORC1</td>
<td>No</td>
<td>QALY</td>
<td>€2,700 ($3200)</td>
<td>€20,000 ($24,200)</td>
<td>Cost-effective</td>
<td>European grant</td>
<td>Netherlands</td>
<td>87</td>
</tr>
<tr>
<td>Dong 2012 (23)</td>
<td>HLA-B *1502</td>
<td>Yes</td>
<td>QALY</td>
<td>$29,800*</td>
<td>$50,000</td>
<td>Cost-effective for Singaporean Chinese and Malays, but not for Singaporean Indians.</td>
<td>Academic resources</td>
<td>Singapore</td>
<td>63</td>
</tr>
<tr>
<td>Tiamkao 2013 (24)</td>
<td>HLA-B *1502</td>
<td>Yes</td>
<td>Costs</td>
<td>98,600 baht ($3,000) saved per 100 cases</td>
<td>n.a.</td>
<td>Cost-effective</td>
<td>None</td>
<td>Thailand</td>
<td>48</td>
</tr>
<tr>
<td>1st Author (reference)</td>
<td>PGx test</td>
<td>Analytical validity PGx test reported</td>
<td>Outcome Measure</td>
<td>Quantitative outcome or ICER (US$)</td>
<td>Cost effectiveness threshold ($/QALY)</td>
<td>Conclusion based on outcome</td>
<td>Funding</td>
<td>Country</td>
<td>QHES score</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------</td>
<td>-------------------------------------</td>
<td>-----------------</td>
<td>------------------------------------</td>
<td>---------------------------------------</td>
<td>-------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>Shiroiwa 2010 (15)</td>
<td>KRAS</td>
<td>No</td>
<td>QALY and LYG</td>
<td>Dominant</td>
<td>n.a.</td>
<td>Dominant</td>
<td>Roche Diagnostics K.K.</td>
<td>Japan</td>
<td>90</td>
</tr>
<tr>
<td>Blank 2011 (25)</td>
<td>KRAS and BRAF</td>
<td>Yes</td>
<td>QALY</td>
<td>KRAS and BRAF saves €3,300 ($2,500) per patient</td>
<td>n.a.</td>
<td>Cost saving</td>
<td>Academic resources</td>
<td>Switzerland</td>
<td>83</td>
</tr>
<tr>
<td>Behl 2012 (26)</td>
<td>KRAS and/or BRAF</td>
<td>No</td>
<td>LYS</td>
<td>KRAS alone saves $7,500 per patient Additional BRAF testing saves $1,000 per patient</td>
<td>n.a.</td>
<td>Cost saving</td>
<td>Academic resources</td>
<td>USA</td>
<td>84</td>
</tr>
<tr>
<td>Shiffman 2012 (27)</td>
<td>LPA</td>
<td>No</td>
<td>QALY</td>
<td>$25,000</td>
<td>No number</td>
<td>Could be cost-effective</td>
<td>Berkeley HeartLab, Inc.</td>
<td>USA</td>
<td>34</td>
</tr>
<tr>
<td>Donnan 2011 (28)</td>
<td>TPMT</td>
<td>Yes</td>
<td>Life months</td>
<td>- Cost no test CAN$ 700 ($600) per patient - Cost genetic test CAN$ 1,100 ($900) per patient - With test no LY gained</td>
<td>n.a.</td>
<td>Not cost-effective</td>
<td>Academic resources</td>
<td>Canada</td>
<td>77</td>
</tr>
<tr>
<td>Schackman 2013 (29)</td>
<td>UGT1A1</td>
<td>Yes</td>
<td>QALY</td>
<td>$2,000,000*</td>
<td>$100,000</td>
<td>Not cost-effective unless assay cost are low</td>
<td>Academic resources</td>
<td>USA</td>
<td>83</td>
</tr>
</tbody>
</table>

*Authors assumed equal costs and efficacy for different pharmaceuticals

BRAF, v-Raf murine sarcoma viral oncogene homolog B1; HLA, human leukocyte antigen; HSR, hyper sensitivity reaction; ICER, incremental cost-effectiveness ratio; KRAS, Kirsten rat sarcoma viral oncogene homolog; LPA, lipoprotein-a; LYG, life-years-gained; LYS, life-years-saved; PGx, pharmacogenetic; QALY, quality-adjusted-life-year; TPMT, thiopurine S-methyltransferase; UGT, UDP-glucuronosyltransferase.
ICER = ICER is not related to PGx scenario

<table>
<thead>
<tr>
<th>1st Author (reference)</th>
<th>PGx test</th>
<th>Analytical validity PGx test reported (y/n)</th>
<th>Outcome Measure</th>
<th>Quantitative outcome or ICER (US$)</th>
<th>Cost-effectiveness threshold ($/QALY)</th>
<th>Conclusion based on outcome</th>
<th>Funding</th>
<th>Country</th>
<th>QHES score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olgiati 2012 (30)</td>
<td>5-HTTLPR</td>
<td>No</td>
<td>QALW</td>
<td>- Euro A $1,100</td>
<td>&lt; 3 times the GDP per capita</td>
<td>Cost-effective in high-income countries</td>
<td>Unknown</td>
<td>Europe</td>
<td>61</td>
</tr>
<tr>
<td>Serretti 2011 (31)</td>
<td>5-HTTLPR</td>
<td>No</td>
<td>QALW</td>
<td>$2,900</td>
<td>$50,000</td>
<td>Not cost-effective</td>
<td>Unknown</td>
<td>Italy</td>
<td>87</td>
</tr>
<tr>
<td>Reed 2011 (32)</td>
<td>8-14 risk alleles</td>
<td>No</td>
<td>QALY</td>
<td>$-98,100</td>
<td>$100,000</td>
<td>Cost-effective</td>
<td>National Cancer Institute</td>
<td>USA</td>
<td>83</td>
</tr>
<tr>
<td>Djalalov 2012 (33)</td>
<td>APOE ε4</td>
<td>Yes</td>
<td>QALY</td>
<td>CAN$38,000 ($32,700)</td>
<td>No number</td>
<td>May be economically attractive</td>
<td>Academic resources</td>
<td>Canada</td>
<td>90</td>
</tr>
<tr>
<td>Kazi 2014 (34)</td>
<td>CYP2C19</td>
<td>No</td>
<td>QALY</td>
<td>- Extendedly dominated;</td>
<td>$50,000</td>
<td>may improve cost effectiveness</td>
<td>American Heart Association and academic resources</td>
<td>USA</td>
<td>90</td>
</tr>
<tr>
<td>Reese 2012 (35)</td>
<td>CYP2C19</td>
<td>No</td>
<td>CVE avoided</td>
<td>- Cost saving</td>
<td>No number</td>
<td>Dominant</td>
<td>Unknown</td>
<td>USA</td>
<td>64</td>
</tr>
<tr>
<td>Sorich 2013 (36)</td>
<td>CYP2C19</td>
<td>No</td>
<td>QALY</td>
<td>- AUSS6,300 ($5,200)</td>
<td>AUSS30,00-50,000 ($24,500-40,800)</td>
<td>Cost-effective</td>
<td>Heart Foundation of Australia</td>
<td>Australia</td>
<td>75</td>
</tr>
<tr>
<td>Crespin 2011 (37)</td>
<td>CYP2C19 *2</td>
<td>Yes</td>
<td>QALY</td>
<td>$10,100</td>
<td>$50,000</td>
<td>Cost-effective</td>
<td>Academic resources</td>
<td>USA</td>
<td>93</td>
</tr>
<tr>
<td>Lala 2013 (38)</td>
<td>CYP2C19 *2</td>
<td>Yes</td>
<td>QALY</td>
<td>- Dominant</td>
<td>- Dominant</td>
<td>n.a.</td>
<td>Dominant</td>
<td>Academic resources</td>
<td>USA</td>
</tr>
</tbody>
</table>
B (continued)

<table>
<thead>
<tr>
<th>1st Author (reference)</th>
<th>PGx test</th>
<th>Analytical validity PGx test reported (y/n)</th>
<th>Outcome Measure</th>
<th>Quantitative outcome or ICER (US$)</th>
<th>Cost-effectiveness threshold ($/QALY)</th>
<th>Conclusion based on outcome</th>
<th>Funding</th>
<th>Country</th>
<th>QHES score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panattoni 2012 (39)</td>
<td>CYP2C19 *2</td>
<td>No</td>
<td>QALY</td>
<td>NZ$ 24,600 ($19,200)</td>
<td>NZ$ 50,000 ($39,000)</td>
<td>Cost-effective</td>
<td>Academic resources</td>
<td>New Zealand</td>
<td>75</td>
</tr>
<tr>
<td>Pink 2013 (40)</td>
<td>CYP2C9 and VKORC1</td>
<td>No</td>
<td>QALY</td>
<td>£13,200 ($20,600)</td>
<td>£20,000-30,000 ($31,200-46,700)</td>
<td>Cost-effective</td>
<td>Academic resources</td>
<td>UK</td>
<td>93</td>
</tr>
<tr>
<td>You 2012 (41)</td>
<td>CYP2C9 and VKORC1</td>
<td>No</td>
<td>QALY</td>
<td>Dominated by genotype–guided approach; $13,800 per QALY; Dominated by dabigatran 150 mg</td>
<td>$50,000</td>
<td>Dabigatran 150 mg seems to be cost-effective</td>
<td>No funding</td>
<td>USA</td>
<td>84</td>
</tr>
<tr>
<td>You 2014 (42)</td>
<td>CYP2C9 and VKORC1</td>
<td>No</td>
<td>QALY</td>
<td>$2,800 per QALY</td>
<td>$50,000</td>
<td>Cost-effective</td>
<td>Research Grants Council of the Hong Kong special administrative Region, China</td>
<td>USA</td>
<td>70</td>
</tr>
<tr>
<td>de Lima Lopes 2012 (43)</td>
<td>EGFR</td>
<td>No</td>
<td>QALY</td>
<td>Dominant</td>
<td>n.a.</td>
<td>Dominant</td>
<td>AstraZeneca Pte Ltd.</td>
<td>Singapore</td>
<td>87</td>
</tr>
<tr>
<td>Handorf 2012 (44)</td>
<td>EGFR</td>
<td>No</td>
<td>QALY</td>
<td>$110,600</td>
<td>$100,000</td>
<td>Cost-effective</td>
<td>OSI Pharmaceuticals/Genentech</td>
<td>USA</td>
<td>90</td>
</tr>
<tr>
<td>Zhu 2013 (45)</td>
<td>EGFR</td>
<td>No</td>
<td>QALY and LYG</td>
<td>$57,000 per QALY; $35,300 per LYG</td>
<td>&lt; 3 times the GDP per capita of China ($16,300)</td>
<td>Not cost-effective*</td>
<td>Shanghai Health Bureau</td>
<td>China</td>
<td>90</td>
</tr>
<tr>
<td>1st Author (reference)</td>
<td>PGx test</td>
<td>Analytical validity PGx test reported (y/n)</td>
<td>Outcome Measure</td>
<td>Quantitative outcome or iCER (US$)</td>
<td>Cost-effectiveness threshold ($/QALY)</td>
<td>Conclusion based on outcome</td>
<td>Funding</td>
<td>Country</td>
<td>QHES score</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------</td>
<td>------------------------------------------</td>
<td>-----------------</td>
<td>---------------------------------</td>
<td>-----------------------------------</td>
<td>---------------------------</td>
<td>---------</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>Kauf</strong> 2010 (46)</td>
<td>HLA-B*5701</td>
<td>No</td>
<td>HSR avoided</td>
<td>$300 (60 days’ time horizon)</td>
<td>no number.</td>
<td>Cost-effective</td>
<td>Glaxo Smith Kline, Inc.</td>
<td>USA</td>
<td>87</td>
</tr>
<tr>
<td><strong>Rattanavipapong</strong> 2013 (47)</td>
<td>HLA-B*1502</td>
<td>Yes</td>
<td>QALY</td>
<td>Epilepsy patients:</td>
<td></td>
<td></td>
<td>Academic resources</td>
<td>Thailand</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- THB220,000 ($6,700)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- THB23,522,000 ($984,800)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neuropathic pain patients:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- THB130,000 ($3,900)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- THB35,877,00 ($108,600)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Liu</strong> 2012 (48)</td>
<td>IL-28B</td>
<td>No</td>
<td>QALY</td>
<td>$50,400</td>
<td>No number.</td>
<td>Not clear</td>
<td>Academic and governmental</td>
<td>USA</td>
<td>83</td>
</tr>
<tr>
<td><strong>Greeley</strong> 2011 (49)</td>
<td>KCNJ11 and ABCB</td>
<td>Yes</td>
<td>QALY</td>
<td>Dominant</td>
<td>n.a.</td>
<td>Dominant</td>
<td>Academic resources</td>
<td>USA</td>
<td>56</td>
</tr>
<tr>
<td><strong>Parthan</strong> 2013 (50)</td>
<td>KIF6</td>
<td>No</td>
<td>QALY</td>
<td>$45,000</td>
<td>$100,000</td>
<td>May be cost effective</td>
<td>Celera corporation</td>
<td>USA</td>
<td>83</td>
</tr>
<tr>
<td><strong>Vijayaraghavan</strong> 2012 (51)</td>
<td>KRAS</td>
<td>Yes</td>
<td>LYS</td>
<td>- Cost saving:</td>
<td>No number.</td>
<td>Cost-saving in both US and Germany</td>
<td>Roche Molecular Systems, Inc.</td>
<td>USA and Germany</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Cost saving:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- $35,500 per LYS.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hagaman</strong> 2010 (53)</td>
<td>TPMT</td>
<td>No</td>
<td>QALY</td>
<td>$29,700</td>
<td>$50,000</td>
<td>Cost-effective</td>
<td>Unknown</td>
<td>USA</td>
<td>64</td>
</tr>
<tr>
<td><strong>Thompson</strong> 2014 (52)</td>
<td>TPMT</td>
<td>No</td>
<td>QALY</td>
<td>Negative ICER</td>
<td>n.a.</td>
<td>Cost-saving but also health loss</td>
<td>Department of Health UK</td>
<td>UK</td>
<td>88</td>
</tr>
<tr>
<td>1st Author (reference)</td>
<td>PGx test</td>
<td>Analytical validity</td>
<td>Outcome measure</td>
<td>Quantitative outcome or ICER (US$)</td>
<td>Cost-effectiveness threshold ($) or QALY</td>
<td>Conclusion based on outcome</td>
<td>Funding</td>
<td>Country</td>
<td>QHES score</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>----------------------------------------</td>
<td>---------------------------</td>
<td>---------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>Pichereau 2010 (54)</td>
<td>UGT1A1*28</td>
<td>No</td>
<td>neutropenia avoided</td>
<td>€900-1,100 ($1100-1300)</td>
<td>No number</td>
<td>Cost-effective</td>
<td>No specific financial support for this study</td>
<td>France</td>
<td>84</td>
</tr>
</tbody>
</table>

ABCC, ATP-binding cassette transporter sub family C; APOE, apolipoprotein-E; CYP, cytochrome P-450; EGFR, epidermal growth factor receptor; GPD: gross domestic product; HLA, human leukocyte antigen; HTTLPR, serotonin-transporter-linked polymorphic region; ICER, incremental cost-effectiveness ratio; KCNJ, Potassium inwardly-rectifying channel, subfamily J; KRAS, Kirsten rat sarcoma viral oncogene homolog; LYG, life-years-gained; LYS, life-years-saved; PGx, pharmacogenetic; QALW, quality-adjusted-life-week; QALY, quality-adjusted-life-year; THB, Thai Baht; TPMT, thiopurine S-methyltransferase; UGT, UDP-glucuronosyltransferase; UK, United Kingdom; VKORC, Vitamin K epoxide reductase complex

* With the gefitinib patient assistance program (sponsored therapy after first six months) it might be a cost-effective treatment option.
Figure 2. Type of outcome analysis (A) and sensitivity analysis (B) of PGx studies from 2000- September 2014. Two studies performed both a CEA and a CUA.

* two studies performed both a CEA and a CUA

Genes investigated and analytical validity of PGx test

In the period from 2000 until 2014, the majority of studies investigated the *TPMT* gene (17 studies). Among the studies presented in this update (i.e. from August 2010 onwards), most investigations concerned the *CYP2C19* screening tests (fig. 3A). We noticed that the *CYP2C19* gene was studied in many different scenarios. Some studies assessed if the use of a *CYP2C19* independent drug like prasugrel or ticagrelor would be a cost-effective treatment option (35,36,38). Similar results were found for *CYP2C9* and *VKORC1* testing in combination with warfarin treatment, which was compared with *CYP2C9* and *VKORC1* independent novel oral anticoagulants, like dabigatran.(41).

In general the analytical validity of PGx tests is high (>95%), nevertheless variation in the analytical validity will in combination with the prevalence of a certain genetic trait determine the predictive value of a test (97). This can be of influence on the cost-effectiveness and was nicely demonstrated by Kauf et al., who identified the negative predictive value as an important input parameter for the costs-effectiveness of *HLA-B*<sup>*S701*</sup> genotyping in their sensitivity analysis (46). Nevertheless, underlying assumptions about the analytical validity of the test were not reported. It was found that only 11 out of the 38 (30%) studies included since 2010, did report underlying analytical validity of the PGx test (table1).

Outcome of the studies

There were 21 out of 80 (26%) studies which concluded that PGx testing was dominant (i.e. resulting in clinical benefits as well as cost-savings). From 2010 onwards, most authors concluded that PGx testing was cost-effective, while only four studies concluded that it was not cost-effective (fig. 3B). In the period from 2010 until 2014, several studies provided the specific conditions at which genetic testing might become cost-effective (table 1A and B). For example, Dong et al. (2012) and Rattanavipapong et al. (2013) showed that the PGx tests could be cost-effective depending on either the patient population or the disease (23,47). Due to an imperfect capability of PGx tests to differentiate between carriers of a genetic variant, some patients might be misclassified and receive suboptimal treatment. As a result, two studies assessing *KRAS* and *BRAF* testing and one study assessing *TMPT* testing found the PGx testing strategy was cost-saving, but with a small health loss compared to the non-testing strategy (25,26,52).

Time horizon and discounting

To capture all costs, savings and effects of an intervention, a lifelong time horizon often seems the best time horizon, although in some scenarios it may be argued that a shorter period is acceptable. Before 2008, 12 out of 26 studies (46%) applied a time horizon of 12 months, and only five (19%) studies applied a lifelong time window. Among the studies published since 2010, there was a broad range in the applied time horizons, from
two weeks until lifetime (S1 Table). In addition, out of the 38 studies 17 (45%) applied a lifelong time horizon, six (16%) studies applied a 30 year time horizon of which two studies also used a lifetime horizon. To deal with uncertainty around the appropriated time horizon, different time horizons can be used. We found that six out of 38 studies (16%) varied their time horizons (18,20,26,38,46,49). In this way, insight into short- and long-term outcomes was given.

![Figure 3. Genes analysed (A) and study outcomes (B) of papers published between August 2010 and September 2014.](image)
Discounting was applied for almost all studies published after 2006 that applied a
time horizon longer than one year. The majority of the studies used discounting at 3%
anually for costs and effects similarly (S1 Table).

Discussion of limitations
The limitations and uncertainties of an economic analysis have to be acknowledged
in order to judge the study on its merits. Nearly all of the newly included papers (i.e.
published between August 2010 and October 2014) discussed their limitations, uncer-
tainties and possible shortcomings of their economic analysis (S2 Table). The topic most
commonly discussed was the lack of solid clinical evidence. As a result, many studies
had to make assumptions or relied on experts’ opinions. However, as mentioned be-
fore, most authors did include different efficacy scenarios in their sensitivity analysis.
Not all papers gave clear information about assumptions concerning sensitivity and
specificity of the PGx tests, and only three out of 38 (8%) mentioned these assumptions
as a limitation (45,49,51). Another typical limitation for the effects of PGx tests is the
time in which test results will become available for clinical decision making. In current
analyses, test results were often assumed to be immediately available. However, this
might be an unrealistic assumption and only one out of 38 studies explicitly mentioned
this assumption as a limitation (23). More general limitations were the lack of data with
respect to heterogeneity in patient populations, hampering extrapolation of results to
patients of different ethnicities, subpopulations and/or country specific populations.
Moreover, several papers mentioned the difficulties in extrapolating long-term clinical
utility results from the short-term clinical trials and lab studies.

Role of funding
Out of the 78 selected studies, 11 (14%) were funded by pharmaceutical companies.
Before 2008, none of the studies was (directly) funded by pharmaceutical companies. In
2008 and 2009 only two (12.5%) out of 16 studies published in that period were funded
by a pharmaceutical company (70,71). Between August 2010 and September 2014, 9
(24%) out of the 38 selected studies were funded by pharmaceutical companies (table
1A and B). Regarding outcomes and conclusions all of these studies concluded that PGx
tests were dominant, cost-saving, or cost-effective. Among the remaining studies which
were funded by other resources, 14% concluded PGx tests were not cost-effective.

Quality assessment
The studies included through this update received a quality score according to the QHES.
There was 2% disagreement between the reviewers for which consensus was reached
through discussion. The average quality score was 76. The score which was given to each
study is shown in table 1A and B. The majority of the studies (71%) were of high quality.
On average, studies concerning testing for \textit{EGFR} received the highest rating and studies about the 21-gene assay received the lowest rating. Some items were scored negative for the majority of the studies. One of them was the item concerning the perspective of the analysis (i.e. societal, health care payers, etcetera). Most studies did not explain why the perspective of the analysis was chosen. In addition, low scoring was received for separate reporting of the short-and long-term outcomes. Most studies did not make such a distinction. Lastly, the direction and magnitude of potential biases were often not discussed.

\textbf{4.1.4. DISCUSSION}

\textbf{Principal findings}

Since 2004, there is an increase in the number of studies evaluating the economic value of PGx tests and this increase accelerated from 2008 onwards. There were not that many economic evaluations of PGx tests available as one might expect given the unravelling of the whole human genome in 2003, the clinical possibilities, and the fast development and decreasing costs of genetic tests (98). This could be related to limited implementation of pharmacogenetic knowledge into daily clinical practice (5). Reasons for this are the uncertainty about clinical relevance, concerns about the availability of genetic data and considerable differences in cost-effectiveness which are found, in particular between countries (5,99,100).

Many studies included through this update did not assess the intrinsic value of the PGx test itself, but compared a PGx test treatment combination with an alternative treatment. For example, the tests for \textit{CYP2C19} or, both \textit{CYP2C9} and \textit{VKORC1} were incorporated in several models as the current treatment option in combination with clopidogrel or warfarin treatment, respectively. The alternative treatment options, which were assessed in the included studies, were independent of pharmacogenetics and were found to be cost-effective. Therefore, in cost-effectiveness analysis of antithrombotic therapy, there seems an ongoing movement away from pharmacogenetic testing, towards treatment options with compounds that are, so far, considered to be independent of pharmacogenetics.

Before 2008, most analyses were cost-effectiveness analyses, however since 2008 there has been a trend towards the use of cost-utility analyses. Cost-utility analysis is currently considered as the preferred type of analysis for health care choices as is advised in several national guidelines, although other types can be suitable depending on the specific study (101,102). Before 2008, most studies performed only a simple univariate sensitivity analysis, if any (9). We found that more recent studies performed both univariate \textit{and} probabilistic sensitivity analyses. This combination is also advised in
Economic evaluations of pharmacogenetic tests.

several national guidelines for pharmacoeconomic evaluation and is a part of the QHES checklist (16,101,102). As a result of these more comprehensive sensitivity analyses, the quality of economic evaluations appears to be improved over the last decade. However, we found that differences in genotyping costs were not always included in the sensitivity analysis which leaves room for improvement.

Among the new studies included in this update (i.e. since 2010), considerable differences in the length of the time horizon were noticed, varying from two weeks to lifelong. A time horizon shorter than a year was primarily related to the expected relevant clinical outcomes. Interestingly, there were differences in the time horizon between studies investigating the same genes. For example, Kazi et al. and Panattoni et al. applied a lifelong time horizon (34,39), whereas others applied a time horizon of 15 months (35,38). All studies discussed at least some limitations and a lack of robust clinical evidence was most frequently mentioned as an important limitation contributing to uncertainty in the analysis. Although studies frequently made assumptions about analytical validity of PGx tests as well as about the rapid clinical availability, these were often not discussed as limitations or studied in sensitivity or scenario analyses. Besides the analytical validity, the clinical validity of PGx tests is important. In general this is the same as an effect estimate of a PGx test. As such, the clinical validity is correctly embedded in an economic evaluation. However, this does not apply for all PGx tests, because some PGx tests involve an analysis of multiple genetic variations. These variations can all contribute to a similar genotype prediction and therefore the clinical validity depends on the number and type of variant alleles analysed. For example, CYP2C19 genotyping which depends on general molecular genetic analysis of single nuclear polymorphisms to detect variant alleles. Several studies included in this review studied only the CYP2C19*2 allele (37-39). However, other allele like the *3 allele can also effect CYP2C19 activity and give a similar clinical effect as the *2 allele (103). When more variant alleles are analysed the clinical validity of the PGx test will increase, although some variants are rare and will contribute little. With the ongoing and rapid increase of knowledge about variant alleles, it is important economic evaluations report the variant alleles on which their assumptions about effects of genotyping (i.e. clinical validity) were based. For the example of CYP2C19, two out of the six studies involving the CYP2C19 gene, did not specify which alleles were included (34,36). This hampers the extrapolation of their findings towards populations of a different ethnicity and other laboratories.

Compared to studies published before 2008, studies were more commonly funded by pharmaceutical companies. Interestingly, all of these sponsored studies concluded that PGx tests were dominant, cost-saving, or cost-effective whereas the few studies which concluded otherwise were financed by academic, governmental, or unknown resources. It is known that studies funded by pharmaceutical companies publish more positive results when compared to studies funded by other resources which is in line with our
results (104). These positive biased results are not related towards the quality of the studies, but to the comparison which is made or publication bias (105). For many of the economic evaluations in this review, assumptions about the effect of the genetic test were made. Furthermore, analytical validity was often not included in the model. This leaves room to bias result in favour of a preferred treatment strategy. Therefore, attention should be given toward assumptions about these aspects, especially when studies are funded by pharmaceutical companies or if the funding is not reported.

Most studies included in this review concluded that the application of PGx tests was cost-effective. Yet, the conclusions were not unambiguous, often due to the uncertainties in the economic models. Another reason for this was that most of the newly included studies did not assess the intrinsic value of the PGx test itself, but a scenario involving one or more PGx tests. Although such estimates are important for the economic impact of the application of personalized medicine, they do not provide information about the cost-effectiveness of the PGx test itself and therefore outcomes can be different. To improve generalizability between studies, an additional scenario analysis in which equal costs and efficacy of the compared treatment strategies are assumed, to assess the intrinsic value of the PGx tests, could be used. Such an approach would especially be interesting when the comparator drug is under patent and drug costs are likely to decrease in the future. Another aspect creating different conclusions was the genetic variety between study populations. This was clearly described by some papers which mentioned the specific conditions like a specific geographic region or a disease for which the genetic testing was cost-effective. For example, PGx testing for HLA-B*5702 was cost-effective in Singaporean Chinese and Malays, but not in Singaporean Indians (23). In addition, Rattanavipapong et al. (2013) found that HLA-B*5702 testing was cost-effective in epileptic patients, but not in neuropathic pain patients (47).

Comparison with previous literature reviews

For the new studies included in this update (i.e. since 2010), an average QHES-quality score of 76 was found. This was is in line with the results from the review from Wong et al. about pharmacoeconomics of PGx tests. They found an average quality score of 77 (13). However, it was lower compared to a related review of Djalalov et al., who found an average quality score of 90 (10). This is likely due to the subjective nature of some items in the quality assessment. For example, ‘item 3’ asks if the used estimates for the analysis were from the best available source, and is therefore quite sensitive to the interpretation of the reviewer (16). Previous review studies pointed out that the methodology of economic evaluations of PGx tests is often heterogeneous and of insufficient quality (9,11). Assasi et al. found that economic PGx studies which were of low quality (i.e. QHES score of < 50), frequently failed to handle uncertainty, did not inform about the study’s limitations, and did not discuss direction and magnitude of potential uncertainty (11).
Among the studies included in this update (i.e. since 2010), most of the authors discussed uncertainties and except for the study by Tiamkao et al., all incorporated uncertainty in a sensitivity analysis (24). However, direction and magnitude of potential bias was still not sufficiently discussed and remains a major point for improvement.

**Limitations of our approach**

This review has some limitations. Firstly, we did not include studies from other databases than PubMed or grey literature and we only included English written studies. Therefore some studies might have been missed. In general, studies that are not indexed in MEDLINE or written in English do not have a large impact on reviews’ outcomes (106). Nevertheless, these studies are frequently of lower quality, and therefore the average quality of the studies included in our review might have been slightly overestimated. However, based on the comparison with other reviews we found a slightly lower average quality score. Lastly, publication bias can always influence the findings of a review. Therefore cost-effectiveness of PGx tests could be overestimated.

**Recommendations**

Based on the quality assessment, reporting of the reasons behind the chosen perspective and the type of economic model can improve the quality of economic evaluations of PGx tests. In addition, reporting of both short-and long-term outcomes and the influence of potential bias, in terms of direction and magnitude on the cost-effectiveness estimates could be improved. Although a substantial and persistent increase in the use of both univariate and probabilistic sensitivity analyses was observed since 2008, there is still room for improvement by using a combination of these techniques instead of one technique, among several studies published since 2008. Publication bias or biased comparators might favour cost-effectiveness of PGx tests. Among studies funded by companies with conflicting interests, the risk on this kind of bias should be critically assessed. In our previous reviews the main limitation that was identified was the unavailability of clinical evidence (1,9). Although this remains an important issue, based on our new findings we can add some recommendations which are more applicable to implementation of PGx tests in clinical practice. First, the clinical validity of a test, i.e. the capability of the test to predict phenotypes with a clinical effect and the analytical validity should be reported as is recommended by the US Academy of Managed Care Pharmacy (107). Both parameters should be included in a sensitivity analysis. In addition, the variant alleles on which these parameters were based should be reported. For some PGx tests, these estimates might be unknown. In this case, a better approach towards this problem would be inclusion of an univariate sensitivity analysis with different cut-off values for the analytical and clinical validity of the PGx test. In this way, a minimum for the analytical and clinical validity can be generated. Note that with an
increase in analysed alleles and as such the clinical validity, the price for the PGx test usually increases as well. Secondly, different turnaround times in which PGx test results would become available for healthcare professionals after requesting the test would be very informative. If direct availability of the genetic test is assumed, for example in the case of pre-emptive genotyping, this should be clearly stated in the method section. This way, the generalizability of results to other countries where PGx tests are available would improve. In addition, a range of costs for the genetic test should be evaluated in univariate and if applicable probabilistic sensitivity analyses. Furthermore, we recommend the addition of a scenario analysis in which drug costs between comparator groups are equalized to give information about the intrinsic value of the PGx test itself.

4.1.5. CONCLUSION

Application of PGx tests was mostly found to be a cost-effective or cost-saving strategy, although some studies concluded otherwise which underlines the importance of future studies assessing the cost-effectiveness of PGx tests. We found that only the minority of recent pharmacoeconomic evaluations assessed the intrinsic value of the PGx tests. New compounds that are not affected by genetics, are emerging as cost-effective alternatives for pharmacogenetic testing strategies. Over the last decade, there was an increase in the number of studies and in the reporting of quality associated characteristics. Due to rapid development in analytical techniques, reporting of analytical and clinical validity of the assessed PGx test is recommended for future evaluations. Furthermore robust clinical evidence regarding PGx tests' efficacy is warranted.
4.1.6. REFERENCES


Economic evaluations of pharmacogenetic tests.


Economic evaluations of pharmacogenetic tests.


Chapter 4.1


Economic evaluations of pharmacogenetic tests.


**S1 Table.** Overview of pharmaco-economic PGx studies published between August 2010 and September 2014 analysing the intrinsic value of a PGx test (A), or comparing different treatment strategies involving PGx testing (B).

<table>
<thead>
<tr>
<th>1st Author (reference)</th>
<th>Disease area</th>
<th>PGx test</th>
<th>PGx test costs</th>
<th>Drug</th>
<th>Analysis</th>
<th>Sensitivity Analyses</th>
<th>Time horizon</th>
<th>Discounting</th>
<th>Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Klang 2010 (17)</strong></td>
<td>ER⁺ LN⁻ ESBC</td>
<td>21 gene assay</td>
<td>$3,460</td>
<td>PGx test + chemotherapy/no chemotherapy vs. chemotherapy/no chemotherapy</td>
<td>CUA</td>
<td>UV and PB</td>
<td>30 yrs and lifetime</td>
<td>3%</td>
<td>Payer’s</td>
</tr>
<tr>
<td><strong>Bacchi 2010 (18)</strong></td>
<td>ER⁺; ALN⁻ ESBC</td>
<td>21 gene assay</td>
<td>$2,294</td>
<td>PGx test + chemotherapy/no chemotherapy vs. chemotherapy</td>
<td>CMA</td>
<td>UV</td>
<td>Not mentioned</td>
<td>No discounting</td>
<td>Third party payer’s</td>
</tr>
<tr>
<td><strong>Hall 2012 (19)</strong></td>
<td>ER⁺ LN⁺ ESBC</td>
<td>21 gene assay</td>
<td>£2,000-£7,000</td>
<td>3,257</td>
<td>PGx test + chemotherapy/no chemotherapy vs. chemotherapy</td>
<td>CUA</td>
<td>UV and PB</td>
<td>30 yrs and lifetime</td>
<td>3.5%</td>
</tr>
<tr>
<td><strong>Vanderlaan 2011 (20)</strong></td>
<td>ER⁺ N⁺-HER2⁻ ESBC</td>
<td>21 gene assay</td>
<td>$3,975</td>
<td>PGx test + chemotherapy/no chemotherapy vs. chemotherapy</td>
<td>CUA</td>
<td>UV</td>
<td>30 yrs</td>
<td>3%</td>
<td>Payer’s</td>
</tr>
<tr>
<td><strong>Verhoef 2013 (21)</strong></td>
<td>AF</td>
<td>CYP2C9 and VKORC1</td>
<td>€20-€160</td>
<td>Base case: €40</td>
<td>Phenprocoumon vs. PGx test + phenprocoumon</td>
<td>CUA</td>
<td>UV, SA and PB</td>
<td>Lifetime</td>
<td>4% (cost) 1.5% (effects)</td>
</tr>
<tr>
<td><strong>Dong 2012 (22)</strong></td>
<td>Epilepsy</td>
<td>HLA-B*1502</td>
<td>$80-$380</td>
<td>Base case: $270</td>
<td>Carbamazepine and phenytoin*</td>
<td>CUA</td>
<td>UV and PB</td>
<td>30 yrs</td>
<td>3%</td>
</tr>
<tr>
<td><strong>Tiamkao 2013 (23)</strong></td>
<td>Neurological diseases</td>
<td>HLA-B*1502</td>
<td>3,00 Baht</td>
<td>Carbamazepine</td>
<td>CMA</td>
<td>None</td>
<td>Unknown</td>
<td>No discounting</td>
<td>Payer’s</td>
</tr>
<tr>
<td><strong>Shiroiwa 2010 (24)</strong></td>
<td>mCRC</td>
<td>KRAS</td>
<td>$220-$1,100</td>
<td>Base case: $220</td>
<td>Cetuximab</td>
<td>CEA + CUA</td>
<td>UV and PB</td>
<td>2.5 yrs</td>
<td>3%</td>
</tr>
<tr>
<td><strong>Blank 2011 (25)</strong></td>
<td>mCRC</td>
<td>KRAS and BRAF</td>
<td>€394</td>
<td>Cetuximab</td>
<td>CUA</td>
<td>UV and PB</td>
<td>Lifetime</td>
<td>3%</td>
<td>Payer’s</td>
</tr>
<tr>
<td>1st Author</td>
<td>Disease area</td>
<td>PGx test</td>
<td>PGx test costs</td>
<td>Drug</td>
<td>Analysis</td>
<td>Sensitivity Analyses</td>
<td>Time horizon</td>
<td>Discounting</td>
<td>Perspective</td>
</tr>
<tr>
<td>------------</td>
<td>--------------</td>
<td>----------</td>
<td>----------------</td>
<td>------</td>
<td>----------</td>
<td>---------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Behl 2012 (26)</td>
<td>mCRC</td>
<td>KRAS and/or BRAF</td>
<td>KRAS: $112-$336 Base case: $224 KRAS and BRAF: $152-$455 Base case: $303</td>
<td>Cetuximab</td>
<td>CEA</td>
<td>PB</td>
<td>1, 2, 5 and 10yrs</td>
<td>3%</td>
<td>Not reported</td>
</tr>
<tr>
<td>Shiffman 2012 (27)</td>
<td>CVD</td>
<td>LPA</td>
<td>$100-$200 Base case: $150</td>
<td>Aspirin</td>
<td>CUA</td>
<td>UV and PB</td>
<td>10yrs</td>
<td>3.5%</td>
<td>Payer’s</td>
</tr>
<tr>
<td>Donnan 2011 (28)</td>
<td>acute lymphoblastic leukaemia</td>
<td>TPMT</td>
<td>$83-$414 Base case: $380</td>
<td>6-Mercaptopurine</td>
<td>CEA</td>
<td>UV and PB</td>
<td>3 months</td>
<td>No discounting</td>
<td>Societal</td>
</tr>
<tr>
<td>Schackman 2013 (29)</td>
<td>HIV</td>
<td>UGT1A1</td>
<td>$10 or $107 Base case: $107</td>
<td>Atazanavir vs. PGx test + atazanavir/darunavir*</td>
<td>CUA</td>
<td>UV and PB</td>
<td>lifetime</td>
<td>3%</td>
<td>Payer’s</td>
</tr>
</tbody>
</table>

*A Authors assumed equal costs and efficacy for different pharmaceuticals
AF, atrial fibrillation; BRAF, v-Raf murine sarcoma viral oncogene homolog B1; CEA, cost-effectiveness analysis; CMA, cost-minimization analysis; CUA, cost-utility analysis; CVD, cardiovascular disease; CYP, cytochrome P-450; ER, oestrogen receptor; ESBC, early stage breast cancer; HER2, Human Epidermal growth factor Receptor 2; HLA, human leukocyte antigen; KRAS, Kirsten rat sarcoma viral oncogene homolog; LPA, lipoprotein-a; (A)axillary LN, lymph node; mCRC, metastatic colorectal cancer; MV, multivariate; PB, probabilistic; PGx, pharmacogenetic; TPMT, thiopurine S-methyltransferase; UGT, UDP-glucuronosyltransferase; UV, univariate; VKORC, Vitamin K epoxide reductase complex.
<table>
<thead>
<tr>
<th>1st Author (reference)</th>
<th>Disease area</th>
<th>PGx test</th>
<th>PGx test costs</th>
<th>Drug</th>
<th>Analysis</th>
<th>Sensitivity Analyses</th>
<th>Time horizon</th>
<th>Discounting</th>
<th>Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olgiati 2012 (30)</td>
<td>Depression</td>
<td>5-HTTLPR</td>
<td>$100-$300 Base care: $200</td>
<td>Citalopram and/or bupropion vs. PGx test + citalopram and/or bupropion.</td>
<td>CUA</td>
<td>UV, MV, and PB</td>
<td>Short term</td>
<td>No discounting</td>
<td>Payer's</td>
</tr>
<tr>
<td>Serretti 2011 (31)</td>
<td>Depression</td>
<td>5-HTTLPR</td>
<td>$142-$326 Base case: $234</td>
<td>Citalopram or bupropion vs. PGx test + citalopram/ bupropion</td>
<td>CUA</td>
<td>UV, MV, and PB</td>
<td>12 weeks</td>
<td>No discounting</td>
<td>Payer's</td>
</tr>
<tr>
<td>Reed 2011 (32)</td>
<td>Prostate cancer</td>
<td>8-14 risk alleles</td>
<td>$200-$600 Base case: $400</td>
<td>No chemotherapy vs. PGx test + finasteride</td>
<td>CUA</td>
<td>UV and PB</td>
<td>Lifetime</td>
<td>3%</td>
<td>Payer's</td>
</tr>
<tr>
<td>Djalalov 2012 (33)</td>
<td>Amnestic cognitive impairment</td>
<td>APOE ε4</td>
<td>CAN$150-CAN$1,625 Base case: CAN$325</td>
<td>Standard of care vs. donepezil + PGx test</td>
<td>CUA</td>
<td>UV and PB</td>
<td>30 yrs</td>
<td>5%</td>
<td>Societal</td>
</tr>
<tr>
<td>Kazi 2014 (37)</td>
<td>ACS</td>
<td>CYP2C19</td>
<td>$100-$700 Base case: $235</td>
<td>Clopidogrel - vs. prasugrel; - vs ticagrelor; - vs PGx test + clopidogrel/ticagrelor - vs PGx test + clopidogrel/prasugrel</td>
<td>CUA</td>
<td>UV, SA, and PB</td>
<td>Lifetime</td>
<td>3%</td>
<td>Societal</td>
</tr>
<tr>
<td>Reese 2012 (34)</td>
<td>ACS</td>
<td>CYP2C19</td>
<td>$310</td>
<td>- Clopidogrel vs. PGx test + clopidogrel/prasugrel; - Prasugrel vs. PGx test + clopidogrel/prasugrel - Generic clopidogrel vs. PGx test + generic clopidogrel/prasugrel; - Prasugrel vs. generic clopidogrel/prasugrel</td>
<td>CEA</td>
<td>PB</td>
<td>15 months</td>
<td>5%</td>
<td>Payer's</td>
</tr>
</tbody>
</table>
### B. (continued)

<table>
<thead>
<tr>
<th>1st Author (reference)</th>
<th>Disease area</th>
<th>PGx test</th>
<th>PGx test costs</th>
<th>Drug</th>
<th>Analysis</th>
<th>Sensitivity Analyses</th>
<th>Time horizon</th>
<th>Discounting</th>
<th>Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorich 2013 (35)</td>
<td>ACS</td>
<td>CYP2C19</td>
<td>AUSS47</td>
<td>- Clopidogrel vs. PGx test + clopidogrel/ticagrelor - PGx test + clopidogrel/ticagrelor vs. ticagrelor</td>
<td>CUA</td>
<td>UV and PB</td>
<td>Lifetime</td>
<td>5%</td>
<td>Healthcare / (Payers?)</td>
</tr>
<tr>
<td>Crespin, 2011 (36)</td>
<td>ACS</td>
<td>CYP2C19 *2</td>
<td>$100-$300 Base case: $200</td>
<td>Ticagrelor vs. PGx test + clopidogrel</td>
<td>CUA</td>
<td>UV and PB</td>
<td>5yrs</td>
<td>3%</td>
<td>Payer's</td>
</tr>
<tr>
<td>Lala 2013 (38)</td>
<td>ACS</td>
<td>CYP2C19 *2</td>
<td>$60-$750 Base case: $500</td>
<td>PGx test + clopidogrel/ prasugrel - vs. clopidogrel - vs. prasugrel</td>
<td>CUA</td>
<td>UV and PB</td>
<td>15 months and 10yrs</td>
<td>3%</td>
<td>Payer's</td>
</tr>
<tr>
<td>Panattoni 2012 (39)</td>
<td>ACS</td>
<td>CYP2C19 *2</td>
<td>NZS 175</td>
<td>- Clopidogrel vs. PGx test + clopidogrel/prasugrel; - Prasugrel vs. PGx test+ clopidogrel/prasugrel.</td>
<td>CUA</td>
<td>PB</td>
<td>Lifetime</td>
<td>3%</td>
<td>Payer's</td>
</tr>
<tr>
<td>Pink 2013 (40)</td>
<td>Nonvalvular AF</td>
<td>CYP2C9 and VKORC1</td>
<td>£20</td>
<td>Warfarin vs. PGx test+ warfarin/novel oral anticoagulants.</td>
<td>CUA</td>
<td>UV and PB</td>
<td>Lifetime</td>
<td>3.5%</td>
<td>Payer's</td>
</tr>
<tr>
<td>You 2014 (41)</td>
<td>AF</td>
<td>CYP2C9 and VKORC1</td>
<td>$50-$200 Base case: $75</td>
<td>Warfarin vs. PGx test+ warfarin/novel oral anticoagulants.</td>
<td>CUA</td>
<td>UV, MV and PB</td>
<td>25yrs</td>
<td>3%</td>
<td>Payer's</td>
</tr>
<tr>
<td>You 2012 (42)</td>
<td>AF</td>
<td>CYP2C9 and VKORC1</td>
<td>$50-$200 Base case: $72</td>
<td>PGx test +warfarin - vs. warfarin; - vs. dabigatran 150mg; - vs. dabigatran mg.</td>
<td>CUA</td>
<td>UV, MV and PB</td>
<td>Lifetime</td>
<td>3%</td>
<td>Payer's</td>
</tr>
<tr>
<td>de Lima Lopes 2012 (43)</td>
<td>NSCLC</td>
<td>EGFR</td>
<td>SG$190-SG$760 Base case: SG$380</td>
<td>Standard chemotherapy vs. PGx test + gefitinib/standard chemotherapy</td>
<td>CUA</td>
<td>UV and SA</td>
<td>Lifetime</td>
<td>No discounting</td>
<td>Payer's</td>
</tr>
<tr>
<td>1st Author (reference)</td>
<td>Disease area</td>
<td>PGx test</td>
<td>PGx test costs</td>
<td>Drug</td>
<td>Analysis</td>
<td>Sensitivity Analyses</td>
<td>Time horizon</td>
<td>Discounting</td>
<td>Perspective</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------</td>
<td>----------</td>
<td>----------------</td>
<td>------</td>
<td>----------</td>
<td>----------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Handorf 2012 (44)</td>
<td>NSCLC</td>
<td>EGFR</td>
<td>$122-$365</td>
<td>Platinum combination vs. PGx test + erlotinib/platinum combination</td>
<td>CUA</td>
<td>UV and MV</td>
<td>12 months</td>
<td>No discounting</td>
<td>Payer's</td>
</tr>
<tr>
<td>Zhu 2013 (45)</td>
<td>NSCLC</td>
<td>EGFR</td>
<td>$381-$635</td>
<td>Routine care vs. PGx test + gefitinib/routine care</td>
<td>CEA and CUA</td>
<td>UV and PB</td>
<td>10yrs</td>
<td>3%</td>
<td>Payer's</td>
</tr>
<tr>
<td>Kauf 2010 (46)</td>
<td>HIV</td>
<td>HLA-B*5701</td>
<td>$50-$150</td>
<td>Abacavir+lamivudine+ efavirenz vs. PGx test + abacavir+lamivudine+ efavirenz/ tenofovir+emtricitabine</td>
<td>CEA</td>
<td>UV</td>
<td>60 days and</td>
<td>No discounting</td>
<td>Payer's</td>
</tr>
<tr>
<td>Rattanavipapong 2013 (47)</td>
<td>Epilepsy and neuropatic pain</td>
<td>HLA-B *1502</td>
<td>THB 1,000</td>
<td>Carbamazepine vs. PGx test + carbamazepine; - vs. alternative (sodium valproate and gabapentin.)</td>
<td>CUA</td>
<td>PB</td>
<td>Lifetime</td>
<td>3.5%</td>
<td>Payer's</td>
</tr>
<tr>
<td>Liu 2012 (48)</td>
<td>Chronic hepatitis C</td>
<td>IL-28B</td>
<td>$186-$557</td>
<td>Standard therapy (interferon + ribavirin) vs. PGx + standard therapy/triple therapy (interferon + ribavirin + protease inhibitor)</td>
<td>CUA</td>
<td>UV and PB</td>
<td>Lifetime</td>
<td>3%</td>
<td>Societal</td>
</tr>
<tr>
<td>Greeley 2011 (49)</td>
<td>Neonatal diabetes mellitus</td>
<td>KCNJ11 and ABCC8</td>
<td>$500-$5,000</td>
<td>Insulin vs. PGx test + sulfonylurea</td>
<td>CUA</td>
<td>UV and PB</td>
<td>10, 20 and 30yrs</td>
<td>3%</td>
<td>Societal</td>
</tr>
<tr>
<td>Parthanan 2013 (50)</td>
<td>ACS</td>
<td>KIF6</td>
<td>$100</td>
<td>Pravastatin vs. PGx test + pravastatin/ atorvastatin</td>
<td>CUA</td>
<td>UV and PB</td>
<td>Lifetime</td>
<td>3%</td>
<td>Payer's</td>
</tr>
</tbody>
</table>
### Economic evaluations of pharmacogenetic tests.

<table>
<thead>
<tr>
<th>1st Author (reference)</th>
<th>Disease area</th>
<th>PGx test</th>
<th>PGx test costs</th>
<th>Drug</th>
<th>Analysis</th>
<th>Sensitivity Analyses</th>
<th>Time horizon</th>
<th>Discounting</th>
<th>Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vijaya-raghavan 2012 (51)</td>
<td>mCRC</td>
<td>KRAS</td>
<td>$243</td>
<td>- Panitumumab vs. PGx test + panitumumab/other chemo; - Cetuximab vs. PGx test + cetuximab/other chemotherapy; - Combination therapy (cetuximab + irinotecan) vs. PGx test + combination therapy /irinotecan.</td>
<td>CEA</td>
<td>UV</td>
<td>Lifetime</td>
<td>No discounting</td>
<td>Payer’s</td>
</tr>
<tr>
<td>Hagaman 2010 (53)</td>
<td>idiopathic pulmonary fibrosis</td>
<td>TPMT</td>
<td>$300</td>
<td>Azathioprine vs. PGx test + azathioprine/ alternative</td>
<td>CUA</td>
<td>UV and MV</td>
<td>Lifetime</td>
<td>No discounting</td>
<td>Payer’s</td>
</tr>
<tr>
<td>Thompson 2014 (52)</td>
<td>Autoimmune disease</td>
<td>TPMT</td>
<td>£20</td>
<td>Azathioprine vs. PGx test + azathioprine/ alternative</td>
<td>CUA</td>
<td>UV</td>
<td>4 months</td>
<td>No discounting</td>
<td>Health service</td>
</tr>
<tr>
<td>Pichereau 2010 (54)</td>
<td>mCRC</td>
<td>UGT1A1 *28</td>
<td>€71</td>
<td>FOLFIRI vs. PGx test + FOLFIRI/ FOLFIRI + CSF</td>
<td>CEA</td>
<td>UV and PB</td>
<td>2 weeks</td>
<td>No discounting</td>
<td>Hospital’s</td>
</tr>
</tbody>
</table>

ABCC, ATP-binding cassette transporter syb family C; ACS, acute coronary syndrome; AF, atrial fibrillation; APOE, apolipoprotein-E; CEA, cost-effectiveness analysis; CSF, Colony Stimulating Factor; CUA, cost-utility analysis; CYP, cytochrome P-450; EGFR, epidermal growth factor receptor; FOLFIRI, Folinic acid, fluorouracil, irinotecan; HLA, human leukocyte antigen; HTTLPR, serotonin-transporter-linked polymorphic region; IL, interleukine; KCNJ, Potassium inwardly-rectifying channel, subfamily J; KIF, kinase family member; KRAS, Kirsten rat sarcoma viral oncogene homolog; LPA, lipoprotein-a; mCRC, metastatic colorectal cancer; MV, multivariate; NSCLC, non-small cell lung cancer; PB, probabilistic; PGx, pharmacogenetic; SA, scenarios analyses; THB, Thai Baht; TPMT, thiopurine S-methyltransferase; UGT, UDP-glucuronosyltransferase; UV, univariate; VKORC, Vitamin K epoxide reductase complex.
### Supplementary file 2. Main limitations which were discussed in the papers published between August 2010 and September 2014.

<table>
<thead>
<tr>
<th>Study</th>
<th>Discussion of limitations</th>
</tr>
</thead>
</table>
| Bacchi et al, 2010            | - The hypothetical nature inherent to economic models  
|                               | - Not all direct and indirect costs were taken into account  
|                               | - The effectiveness of therapy based on the risk prediction of the assay was not considered  
|                               | - There was variability in the definition and incidence of febrile neutropenia  
|                               | - It was not possible to include Brazilian estimates for stage distribution upon diagnosis of breast cancer                                                                                                           |
| Behl et al, 2012              | - No progression-free survival outcomes for the different strategies was provided  
|                               | - Limitations in data for differences among the treatment options in quality of life, and was therefore not incorporated  
|                               | - Gain in overall survival may not have been seen if the health impact of treatment were expressed in quality-adjusted survival  
|                               | - Differences in quality of life among patients receiving alternate treatments have not been quantified in a way that allows quality adjusted survival to be modelled  
|                               | - It was not taken in account that patients could discontinue the treatment due to adverse events  
|                               | - The analysis assumes that all differences in survival are due to a lack of response to cetuximab. Though mutations may independently predict prognosis, not the kind of treatment  
|                               | - The estimated savings might be overestimated if any part of the survival difference is due to mutations, and independent of the therapy                                                                                   |
| Blank et al, 2011             | - Lack of clinical data of a clearly define patient population from Switzerland. Hence, clinical and utility data originated from few studies conducted outside Switzerland  
|                               | - Uncertainty in trial data and potentially limited transferability to routine clinical practice populations  
|                               | - The quality of life and utility data allowed to differentiate on the basis of treatment, but not of mutation status  
|                               | - *BRAF* mutation seemed to have no impact on response to the antibody, suggesting that *BRAF* mutation may not have the same predictive value in first-line and chemorefractory tumours. |
| Crespin et al, 2011           | - The ultimate effectiveness of antiplatelet medication likely cannot be determined, because the short duration of the clinical trials  
|                               | - The estimates might be biased if the effectiveness of ticagrelor differs after 12 months of therapy  
|                               | - Results may not apply to subpopulations extracted from clinical trials  
|                               | - The costs for adverse event might vary between subpopulations due to events that significantly differ  
|                               | - Assumptions that MI and death were independent, which is likely unrealistic  
<p>|                               | - Results cannot be used to determine universal cost-effectiveness relative to other viable treatment options for secondary prevention after an acute coronary syndrome  |</p>
<table>
<thead>
<tr>
<th>Study</th>
<th>Discussion of limitations</th>
</tr>
</thead>
</table>
| de Lima Lopes et al. 2012   | - Assumptions were made about length and time spent in treatment by extrapolating from clinical data, and assuming that the time spent in first- or second-line treatment with chemotherapy of gefitinib was equivalent to account for similar overall survival  
- Cost data is obtained from 3 Singapore oncology centres, and might not be representative for other countries and/or hospitals  
- Several assumptions about quality of life values were made, hence the estimated ICERs might be biased  
- Lack of data to capture robust individual level variation in treatment responses  
- The standard of care for first line treatment has shift to include pemetrexed, bevacizumab and cetuximab in first-line treatment and to include pemetrexed as potential maintenance option after initial chemotherapy  
- The model specifically compares a standard practice, which includes no EGFR mutation testing, first-line treatment with chemotherapy and second-line treatment with gefitinib, to EGFR testing and guided therapy based on the results of the test.  
- In the base case it was assumed that gefitinib did not benefit patients without activating mutations, even beyond first line treatment |
| Djeralov et al. 2012        | - There is limited evidence on the effectiveness of donepezil treatment in delaying progression from AMCI to AD among APOE e4 carriers.  
- Surveillance costs were based on results from a study that used older patients, and was conducted in a different healthcare system  
- They did not used some data from a study, because that data might have had selection biases that limited the generalisation of the results  
- Adequate and widely accepted criteria for diagnosing AMCI are unavailable, but incorrect diagnoses not incorporated in the study. Incorrect diagnosing might increase the ICER  
- They did not included patients that were not seen and diagnosed with AMCI before the development of AD |
| Dong et al. 2012            | - The study assumes that health-related QoL is restored to perfect health. However, the QALY gains might be overestimated if the health related QoL is in reality lower due to an imperfect response to drugs, epilepsy recurrence or other health problems  
- Treatment decisions are in reality influenced by many factors and may substantially deviate from the initial assumptions  
- The model also includes several additional assumptions and simplifications:  
  - Some patient subpopulations excluded  
  - Treatment rules simplified for who receives the treatment  
  - The model assumes that VPA, CBZ, and PHT have similar efficacy.  
- Effectiveness data from a study using Caucasians, ethnic differences cannot be ruled out  
- They assumed that genotyping results are immediately accessible, though this may not hold universally |
| Donnan et al. 2011          | - The rarity of the disease limits amount and type of evidence available, therefore some values based on expert opinion  
- No QALY analysis possible, due to lack of literature on utility scores for children with ALL  
- There was some uncertainty in the values used for the unit prices of TPMT genotype and enzymatic tests |
### Supplementary file 2. (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Discussion of limitations</th>
</tr>
</thead>
</table>
| Greeley et al. 2011 | - In this study it was assumed that the test was 100% sensitive and specific, although there was not accounted for factors that might diminish the effective clinical sensitivity and specificity  
- Lack of data, therefore some of the model’s assumptions for selected complications were derived from studies performed on patients with diabetes type 2  
- Due to a lack of quality-of-life data in neonatal diabetes, it was assumed that patients would experience a utility gain of 0.1 based on survey data from diabetes type 2 subjects  
- Due to insufficient longitudinal data, it was assumed that patients with treatable genetic defects would remain responsive to the treatment over 30 years  
- The model did not account for the therapy’s potential to improve neurodevelopmental outcomes |
| Hagaman et al. 2010  | - There were no published trials that specifically explored their issue; therefore they had to draw data from numerous sources.  
- The study assumed that the low-dose treatment with reduced TPMT activity is the same as for full-dose therapy in patients with normal TPMT activity  
- The study lacked data describing the incidence of BMT in patients with intermediate TPMT activity  
- The breakdown of TPMT activity in patients with leukopenia has not been clearly elucidated  
- The authors assumed that conservative therapy was equal in efficacy to the placebo arm in another study, in which the patients received azathioprine and steroids with a NAC placebo. This assumption likely overestimates the marginal cost-effectiveness of therapy with azathioprine, NAC and steroids. |
| Hall et al. 2012  | - The authors conducted an analysis of the immediate budget impact to illustrate the financial implications during the chemotherapy period only  
- The budget impact results do not characterise the strength evidence, and do not consider long term costs (for cancer recurrence and treatment toxicity, relative life expectancy and quality of life)  
- Only the Oncotype DX was considered, and not the relative value and alternative tests  
- It is important to recognize that any model is a simplification of reality, and the model presented here is no exception.  
- It is credible that taking into account the transferability of data from a US trial into a UK setting would generate additional uncertainty and even introduce bias into the results.  
- The use of decision aids such as Adjuvant Online! was not specifically incorporated into the model  
- The possibility that the price of the test might change with the introduction of alternatives was not considered  
- Estimates of long-term costs and quality of life (e.g., associated with cancer recurrence or cardiac toxicity) are derived from higher quality evidence than the short-term data sources (e.g., costs of toxicity) in this analysis because they rely on previously published dedicated cost and quality of life studies. They are, however, subject to assumptions of data transferability to our patient population and require confirmation in a formal study with long-term follow-up.  
- This analysis presents results only for the average patient aged 60 years |
### Supplementary file 2. (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Discussion of limitations</th>
</tr>
</thead>
</table>
| Handorf et al. 2012 | - The authors were not able to model the full course of lifetime therapy, which often included second- and third-line treatment. The inputs for such a model were not available  
  - A model-based approach was applied, rather than a randomised trial. Therefore, the conclusions depend on the validity of the assumptions that were used to develop the model  
  - The authors were not able to gather costs and effects for claims and registry data  
  - The current analyses were subject to modifications in costs (reduction of costs in the future makes the strategies more favourable)                                                                                                                                                                         |
| Kauf et al. 2010  | - This analysis relied on several assumptions specific to the development of the ADVANCE model as described in the text and the supplemental data.  
  - Care should be taken in the interpretation of the results in which screening dominated the alternative therapy, because differences in effectiveness very small in the dominant screening compared with alternative strategy  
  - The results only apply to those patients for whom abacavir and tenofovir are considered appropriate treatment alternatives  
  - This analysis does not consider all the possible benefits of screening                                                                                                                                                                                                                                                                 |
| Kazi et al. 2014  | - “Estimated differences in outcomes between various CYP2C19 genotypes were largely based on post hoc analyses of randomized trials”  
  - The efficacy and safety of prasugrel and ticagrelor were based on only one large, randomized clinical trial  
  - The indirect comparison of ticagrelor with prasugrel was limited due to structural differences in the design and execution of the used clinical trials  
  - It was assumed that the clinical outcomes from the PLATO trial can be translated to U.S. patients on low dose aspirin therapy, however this has yet to be investigated  
  - In order to define the actual relative cost-effectiveness, the long-term effects of newer antiplatelet agents on mortality rate has to be determined                                                                                                                                                              |
| Klang et al. 2010 | - The clinical data on which the analyses were based were derived from a non-randomly selected sample of patients  
  - Utilities were derived from English speaking literature, thus may not fully characterise the preferences of patients in Israel  
  - Validation of the essay was based on clinical trials conducted in the US  
  - Due to limited data, the authors omitted some potential long-term implications of breast cancer and its treatments such as the risk of local recurrence and risk of second primary tumours associated with chemotherapy  
  - The authors examined the effect of the test from the payer’s perspective, hence indirect costs were not included                                                                                                                                                                                                 |
<table>
<thead>
<tr>
<th>Study</th>
<th>Discussion of limitations</th>
</tr>
</thead>
</table>
| Lala et al. 2013 | - Healthcare delivering costs were not available at the individual patient level  
- The authors only accounted for the impact of testing for CYP2C19*2 allele was taken in account, and the cost-effectiveness of genetic testing for other mutations was not taken in account  
- This analysis did not address the economics of platelet reactivity testing  
- For patients with a prior history of TIA or stroke, body weight <60 or age >75 prasugrel may not be considered optimal therapy  
- It was assumed that there was a constant magnitude of benefit and a constant rate of adverse events  
- The authors assumed that the relative risk of death in our population, compared with the general US population, remained fixed over time.  
- There studied that found differences in bleeding events between CYP2C19*2 carriers and wild type  
- Although genetic testing was shown to be cost-saving compared with treating all ACS patients undergoing PCI empirically with prasugrel or clopidogrel, the absolute health and cost differences were small |
| Liu et al. 2012 | - Because of a lack of evidence, strategies involving retreatment with triple therapy after initial failure was not taken into account  
- They did not included the effect of reductions in HCV transmission due to successful treatment  
- The results are limited to mono-infected individuals |
| Olgiati et al. 2012 | - It assumes that 5-HTTLPR variants have the same distribution and effect size in all European countries.  
- The influence of 5-HTTLPR on SSRI response is documented in randomized trials, but biased by various limitations  
- The issue of the transferability of the results was implicitly addressed by considering several regions at different income levels  
- According the authors it is arguable that reduction in health expenditures for approximately 4% of new responders under pharmacogenetic treatment cannot offset incremental costs for genetic test |
| Panattoni et al. 2012 | - The findings are based on models of outcomes rather than a randomised trial  
- Whether the shorter duration of clopidogrel therapy contributed to the higher events rates was not clear  
- The definitions and classification of adverse events differ between New Zealand hospital DRG data and the TRITON-TIMI 38 clinical trial  
- The ethnicity data has limitations, because many patients had heritage from more than 1 ethnical group |
### Supplementary file 2. (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Discussion of limitations</th>
</tr>
</thead>
</table>
| Parthan et al. 2013          | - The analysis was based on a post hoc substudy, which was not designed to investigate the cost-effectiveness of KIF6 testing  
- It was assumed that the patients with a major cardiovascular event remain in this state until death, although they have a higher risk of a second event  
- Cardiovascular event rates were extrapolated because data was only available for 2 years  
- The estimates of differential statin benefit may have been overestimated, because the association of KIF6 with the differential reduction of CHD event rates from statin therapy may be limited to high-dose atorvastatin and standard dose pravastatin  
- There was no data on event rates for non-adherent patients by KIF6 carrier status, therefore the 4-month event rate for patients in the placebo arm was used as a proxy  
- The authors did not accounted for a possible relationship between multiple events and risk of death, therefore the estimation of life expectancy may not be accurate  
- The cost of secondary events was estimate to be the same as the costs of a primary event  
- “The acute costs of UA requiring hospitalization is represented in the model as a 36-month cost in light of the absence of data on UA as an event secondary to other cardiovascular events.”                                                                                                                                                                                                                   |
| Pichereau et al. 2010        | - The authors did not opt for a full economic model in the analysis as the primary intention was to evaluate whether or not UGT1A1 genotype testing would be an efficient use of additional resources from the hospital.  
- Only few studies were used to build our model as most of the available data were not associated with complete information of polymorphisms prevalence or FN incidence  
- The authors assumed 100% efficacy of CSF therapy, thus inflating the estimates of neutropenic events avoided, if efficacy is in fact lower.  
- The costs of CSF were not included because this drug was not provided by the hospital  
- Indirect and intangible costs were excluded  
- Irinotecan toxicity might also be affected by other polymorphisms, though this was not taken in account for this study                                                                                                                                                                                                                                                                                                                                                        |
| Pink et al. 2013             | - The adjusted indirect comparison is necessary to include all possible treatment options. However, this may introduce bias through differences in trial design, a lack of access to individual patient data and the need to extrapolate the available data from trial to lifetime horizons.  
- The results were obtained from a number of data sources are difficult to validate externally  
- Both the PKPD and the economic models are parameter extensive, increasing the probability that some of the values used are inaccurate  
- Each of the three stages of the methodology introduces uncertainties                                                                                                                                                                                                                                                                                                                                                                         |
| Rattanavipapong et al. 2013  | - Due to the rarity of SJS/TEN cases, only a small number of patients was included, representing both cost and utility limitations  
- There is no surveillance system to quantify the prevalence of CBZ-induced SJS/TEN in the Thai population  
- This study employed data from only one study, which was conducted in a medical school in Thailand                                                                                                                                                                                                                                                                                                                                 |
Chapter 4.1

Supplementary file 2. (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Discussion of limitations</th>
</tr>
</thead>
</table>
| Reed et al. 2011 | - The impact of finasteride outside clinical trial settings is unclear. In the PCPT active surveillance for prostate cancer continued throughout the trial, so it is conceivable that the effect of finasteride could be attenuated among patients followed in routine practice  
- It is also unclear whether the treatment effects measured in the PCPT apply equally to patients at higher risk for prostate cancer, whether treatment effects decrease over time and whether reductions in the prevalence of prostate cancer lead to decreases in prostate cancer specific mortality.  
- There is no established threshold in the US for determining whether an intervention is cost-effective  
- Efficiency gains were limited from a cost-effectiveness perspective |
| Reese et al. 2012 | - The authors included both IMs and PMs for prasugrel in the genotype-guided treatment arm, although a boxed warning in the clopidogrel label only includes PMs. However, the majority of the published data indicates that both IMs and PMs are at increased risk.  
- The probabilities used in the base-case models were obtained from one randomised trial and substudies of that trial. This concerns age and ethnicity  
- Substudy analyses could have introduced biases  
- In this study the treatment lasted 15 months, whereas the treatment guidelines recommend antiplatelet therapy for at least one year  
- The cost of genetic tests will probably influence the use. |
| Schackman et al. 2012 | - The results from a retrospective analysis were used, though they may not be generalizable to the US or other populations  
- The authors assumed in the base case that atazanavir and darunavir had equivalent efficacy and costs  
- The assumption that clinicians or patients might prefer to initiate atazanavir was not captured in the base case QALYs or costs  
- The range of QoL effects of hyperbilirubinaemia that was considered may not capture the full spectrum of clinical situations  
- The authors did not consider the potential future benefit of UGT1A1 testing to inform prescribing of other non-HIV drugs |
### Study 1: Serretti et al. 2011
- There was no experimental data about effective gain in antidepressant response due to pharmacogenetic approach.
- The assumptions that patients with s/s genotype treated with citalopram would have the greatest benefit from genetic testing would, due to the remission rates, imply that virtually all the antidepressant response in the 5-HTTLPR s/s genotype group was due to a placebo response and none to the SSRI antidepressant.
- At best estimate 50-60% of change in response outcome might be due to a true pharmacologic effect.
- This study focused on antidepressant response and side effects. However the 5-HTTLPR polymorphism has also been found to moderate depressive response to environmental stress.
- A long follow-up would have been methodologically correct, but constituted a serious hurdle to build a realistic model. Firstly, because only a minority of patients do not drop out from treatment after a few months. Secondly, even in those who remain in treatment, adherence is seldom at optimal level, and this may have negative consequences for effectiveness and cost. For a long-term assessment of major depressive disorder it is necessary to estimate recurrence rate. This is not an easy task, because follow-up studies of depression are characterized by marked differences in terms of designs, outcome definitions and crude measures of pharmacotherapy.
- The influence of 5-HTTLPR on SSRI response is documented in randomised trials, but biased by various limitations.
- The real effect is not yet established.
- The model did not account for recent discoveries that changed the structure and function of the 5-HTTLPR polymorphism.
- In order to simplify the association between 5-HTTLPR variants and antidepressant response, second-order interactions with gender and life-events were not featured.
- Information was missing regarding costs to caregiver or family members and psychotherapy.
- The used typical starting doses for SSRI treatment might be suboptimal and interfere with the assessment of clinical response.
- The impact of antidepressant treatment on suicidal risk was not featured.
- The results are only provisional because key assumptions regarding gain in antidepressant response and reduction in side-effect burden are speculative and not supported by empirical findings.

### Study 2: Shiffman et al. 2012
- The authors were unable to measure or find a published estimate of the risk of stroke events associated with the 2 variants of the LPA gene in women. Therefore, they assumed that the increased stroke risk in women was the same as the increased MI risk estimated in men.
- It was assumed that frequency of the LPA risk alleles is unchanged at different levels of the FRS.
- There was no source documenting the timing of GI bleeding events that occur after initiation of aspirin therapy, therefore it was assumed that GI bleeding events would occur in the first year after initiation of therapy.
- There were no reports of an association between these LPA variants and the risk of CVD in populations of non-European ancestry.
### Supplementary file 2. (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Discussion of limitations</th>
</tr>
</thead>
</table>
| Shiroiwa et al. 2010   | - The authors estimated the BSC costs, which were based on the fee-for-service costs of daily opioid use although some patients do not use them, or less frequently  
                        - There was not data available for Japanese utility scores, so Canadian scores were extrapolated                                                                                 |
| Sorich et al. 2013     | - Cost between health systems are different and result may not apply to other healthcare systems  
                        - No comparison with prasugrel or genotype guided dosing including alternative genotypes like Ultra Rapid metabolizers.  
                        - It was assumed that clopidogrel or ticagrelor were used for 12 months after ACS and then stopped.                                                                 |
| Thompson et al. 2014   | - The study’s sample may not be generalizable to all potential patients who will be offered TPMT genotyping in clinical practice.  
                        - The study time horizon should have been longer to capture any potential long-term costs and benefits  
                        - The perspective of the study might be a limitation, because it does not take into account the costs beyond health care resource use |
| Tiamkao et al. 2013    | - This study was based on reviews of literature, so the details could be slightly different and could result in discrepancies  
                        - Costs of treatment of other complications after discharging the patient, re-admission or out-patient follow up were not included                                         |
| Vanderlaan et al. 2011 | - The estimate of the net reduction in chemotherapy associated with assay testing was based on a single published study for the general N+ (1-3)/ER+ population and on a separate analysis for the >65-year age group  
                        - The model used an age distribution representative for the US population estimating the incidence of breast cancer, though the model was applied to a potentially younger managed care population  
                        - A payer’s perspective was applied, and indirect costs were excluded  
                        - Clinical trial data was used to estimate costs of treatment, hospitalisation and medication associated with adverse events, yet the costs are often higher in a non-controlled setting  
                        - The authors modelled change in chemotherapy use only among those with low recurrent score results in order to provide a conservative estimate of the benefits gained from the assay’s use in this population  
                        - The plan costs of chemotherapy drugs and supportive care may have been overestimated  
                        - Unpublished survey data was used to estimate practice patterns of chemotherapy-related supportive care  
                        - It was estimated that 90% of the cancers would be non-HER2 overexpressing, although this might be uncertain  
                        - The estimated 2.7% excess risk among patients receiving chemotherapy may be an overestimate, because it was based on data that included radiotherapy and hormonal treatment  
                        - The model did not address the use of the assay in the population with the current greatest use  
                        - The model assumed that cancer and all-cause mortality were estimated based on national statistics but did not incorporate one additional benefit of the oncotype DX assay (if the recurrent score is known, patients’ risk of cancer-related death is stratified into 3 categories)  
                        - There is great variability in the value that patients attribute to chemotherapy treatment |
### Supplementary file 2. (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Discussion of limitations</th>
</tr>
</thead>
</table>
| Verhoef et al. 2013        | - Lack of data concerning the effectiveness of genotyping in phenprocoumon patients from clinical trials.  
|                            |   - Cost of genetic test                                                                  |
|                            |   - Surrogate end point (INR)                                                              |
|                            |   - Possible correlation between parameters which were independently varied in the probabilistic sensitivity analysis |
| Vijayaraghavan et al. 2012 | - The analysis was restricted to the use of EGFR inhibitors in second and subsequent lines of therapies  
|                            |   - It was not taken in account that recent findings indicated that use of cetuximab in chemotherapy-refractory colorectal cancer patients is associated with longer overall survival compared to patients with other KRAS-mutated tumours  
|                            |   - Due to a lack of data on utilities for patients with mCRC receiving therapy, no quality of life adjustments were incorporated in the model  
|                            |   - There is limited data on the precise sensitivity and specificity of the KRAS test, therefore the authors estimated a 95% sensitivity and 100% specificity |
| You et al. 2012            | - The model projected life-long events based on key factors derived from a 2 years clinical trial  
|                            |   - The cost items were limited to the resources of anticoagulation therapy and related complications |
| You JH, 2014               | - Long-term events were projected by using short-term clinical trial data                    |
|                            |   - Monitoring of new data regarding NOACs is required to update the decision model         |
|                            |   - The cost items were limited to resources of anticoagulation therapy and related complications |
|                            |   - A TTR of 60% was assumed, which might not be applicable to clinics with higher TTR      |
|                            |   - In the CoumaGen-II trial patients with various warfarin indications were mixed, which could result in uncertainty in the estimation of genotype-guided dosing effectiveness in AF patients in this study. |
| Zhu et al. 2013            | - The current analysis did not evaluate the cost-effectiveness of gefitinib maintenance treatment for the whole cohort without EGFR genotyping because this data was not available |
|                            |   - The present model did not include other EGFR-targeted agents used as maintenance treatments, such as erlotinib, to assess the incremental cost-effectiveness in comparison with gefitinib because no head-to-head trial data are currently available. |
|                            |   - A budget impact analysis for the addition of gefitinib maintenance treatment on society was not conducted |
|                            |   - The current analysis incorporated PSF and OS data after cancer progression from different trials. |
|                            |   - Some model inputs were obtained from literature published abroad due to a lack of Chinese clinical data |
|                            |   - The sensitivity and specificity of different genotyping facilities was not accounted |
|                            |   - In order to simplify the evaluation, other adjuvant therapies were excluded |
### S3. PRISMA checklist

<table>
<thead>
<tr>
<th>Section/topic</th>
<th>#</th>
<th>Checklist item</th>
<th>Reported on page #</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TITLE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>1</td>
<td>Identify the report as a systematic review, meta-analysis, or both.</td>
<td>1</td>
</tr>
<tr>
<td><strong>ABSTRACT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structured summary</td>
<td>2</td>
<td>Provide a structured summary including, as applicable: background; objectives;</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>data sources; study eligibility criteria, participants, and interventions;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>study appraisal and synthesis methods; results; limitations; conclusions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>and implications of key findings; systematic review registration number.</td>
<td></td>
</tr>
<tr>
<td><strong>INTRODUCTION</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>3</td>
<td>Describe the rationale for the review in the context of what is already known.</td>
<td>3-4</td>
</tr>
<tr>
<td>Objectives</td>
<td>4</td>
<td>Provide an explicit statement of questions being addressed with reference</td>
<td>3-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to participants, interventions, comparisons, outcomes, and study design</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(PICOS).</td>
<td></td>
</tr>
<tr>
<td><strong>METHODS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol and registration</td>
<td>5</td>
<td>Indicate if a review protocol exists, if and where it can be accessed (e.g.,</td>
<td>4-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Web address), and, if available, provide registration information including</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>registration number.</td>
<td></td>
</tr>
<tr>
<td>Eligibility criteria</td>
<td>6</td>
<td>Specify study characteristics (e.g., PICOS, length of follow-up) and report</td>
<td>4-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>characteristics (e.g., years considered, language, publication status) used</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>as criteria for eligibility, giving rationale.</td>
<td></td>
</tr>
<tr>
<td>Information sources</td>
<td>7</td>
<td>Describe all information sources (e.g., databases with dates of coverage,</td>
<td>4-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contact with study authors to identify additional studies) in the search and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>date last searched.</td>
<td></td>
</tr>
<tr>
<td>Search</td>
<td>8</td>
<td>Present full electronic search strategy for at least one database, including</td>
<td>4-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>any limits used, such that it could be repeated.</td>
<td></td>
</tr>
<tr>
<td>Study selection</td>
<td>9</td>
<td>State the process for selecting studies (i.e., screening, eligibility, included</td>
<td>4-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in systematic review, and, if applicable, included in the meta-analysis).</td>
<td></td>
</tr>
<tr>
<td>Data collection process</td>
<td>10</td>
<td>Describe method of data extraction from reports (e.g., piloted forms,</td>
<td>4-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>independently, in duplicate) and any processes for obtaining and confirming</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>data from investigators.</td>
<td></td>
</tr>
<tr>
<td>Data items</td>
<td>11</td>
<td>List and define all variables for which data were sought (e.g., PICOS, funding</td>
<td>4-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sources) and any assumptions and simplifications made.</td>
<td></td>
</tr>
<tr>
<td>Risk of bias in individual studies</td>
<td>12</td>
<td>Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level, and how this information is to be used in any data synthesis).</td>
<td>No meta-analysis was performed</td>
</tr>
<tr>
<td>Summary measures</td>
<td>13</td>
<td>State the principal summary measures (e.g., risk ratio, difference in means).</td>
<td>No meta-analysis was performed</td>
</tr>
<tr>
<td>Synthesis of results</td>
<td>14</td>
<td>Describe the methods of handling data and combining results of studies, if</td>
<td>No meta-analysis was performed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>done, including measures of consistency (e.g., $I^2$) for each meta-analysis.</td>
<td></td>
</tr>
</tbody>
</table>
### Economic evaluations of pharmacogenetic tests.

<table>
<thead>
<tr>
<th>Section/topic</th>
<th>#</th>
<th>Checklist item</th>
<th>Reported on page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of bias across studies</td>
<td>15</td>
<td>Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).</td>
<td>No meta-analysis was performed</td>
</tr>
<tr>
<td>Additional analyses</td>
<td>16</td>
<td>Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.</td>
<td>No meta-analysis was performed</td>
</tr>
</tbody>
</table>

#### RESULTS

**Study selection**

17 Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.

5

**Study characteristics**

18 For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.

5-13

**Risk of bias within studies**

19 Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).

No meta-analysis was performed

**Results of individual studies**

20 For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.

No meta-analysis was performed

**Synthesis of results**

21 Present results of each meta-analysis done, including confidence intervals and measures of consistency.

No meta-analysis was performed

**Risk of bias across studies**

22 Present results of any assessment of risk of bias across studies (see Item 15).

No meta-analysis was performed

**Additional analysis**

23 Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).

No meta-analysis was performed

#### DISCUSSION

**Summary of evidence**

24 Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).

14

**Limitations**

25 Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).

16

**Conclusions**

26 Provide a general interpretation of the results in the context of other evidence, and implications for future research.

16

#### FUNDING

27 Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.

Online submission info

---
