CADTH

CADTH RAPID RESPONSE REPORT:
SUMMARY WITH CRITICAL APPRAISAL

Pharmacogenomic Testing in Depression: A Review of Clinical Effectiveness, Cost-Effectiveness, and Guidelines

Service Line:	Rapid Response Service
Version:	1.0
Publication Date:	January 31, 2020
Report Length:	38 Pages

CADTH

Authors: Ke Xin Li, Hannah Loshak

Cite As: Pharmacogenomic testing in depression: a review of clinical effectiveness, cost-effectiveness, and guidelines. Ottawa: CADTH; 2020 Jan. (CADTH rapid response report: summary with critical appraisal).

ISSN: 1922-8147 (online)

About CADTH: CADTH is an independent, not-for-profit organization responsible for providing Canada's health care decision-makers with objective evidence to help make informed decisions about the optimal use of drugs, medical devices, diagnostics, and procedures in our health care system.

Funding: CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

Questions or requests for information about this report can be directed to **Requests@CADTH.ca**

Abbreviations

ADHD	attention deficit hyperactivity disorder
AMSTAR II	A MeaSurement Tool to Assess systematic Reviews II
CRD	Centre for Reviews and Dissemination
DST	decision support tool
HAMA	Hamilton anxiety scale
HAMD-17, HAM-D17, or HDRS-17	17-item Hamilton Depression Rating Scale
HAM-D21	21-item Hamilton Depression Rating Scale
HTA	health technology assessment
HQO	Health Quality Ontario
ICER	incremental cost-effectiveness ratio
MA	meta-analysis
MDD	major depressive disorder
MeSH	medical subject headings
QALY	quality-adjusted life year
QALW	quality-adjusted life week
QIDS-C16	Clinician Rated Quick Inventory of Depressive Symptomatology
QoL	quality of life
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	randomized controlled trial
SR	systematic review
TDM	therapeutic drug monitoring
TESS	treatment emergent symptom scale
US	United States
WTP	Willingness-to-pay

Context and Policy Issues

In Canada, 11.3% of adults identified symptoms that met the criteria for depression in 2012.[1] It was estimated in 2012 that 11.2% of Canadians experience major depressive disorder (MDD) at least once in their lifetimes and 4.7% of Canadian experience MDD per year.[2] Depression severity is often measured by the 17-item Hamilton Depression Rating Scale (HAMD-17, also known as HDRS-17).[3] A score of zero to seven indicates no depression; a score of eight to 16 for mild depression; 17 to 23 for moderate depression; and 24 or greater for severe depression.[3]

Genetic variants in patients with depression could be the explanation for about 42–50% of individual differences in the antidepressant response rates.[4,5] With the decreasing cost of genotyping, genetic testing-guided medication therapy has been increasing in popularity around the world.[6] Pharmacogenetics is the study of genes that cause variability in drug response, while pharmacogenomics is broader in context, referring to the collective effect of variability across the genome to modulate drug response.[7] Pharmacogenetics and pharmacogenomics are often used interchangeably in the published literature.[7] There is a growing number of pharmacogenomics testing-guided decision support tools (DSTs) for patients with depression available worldwide.[6] DSTs work by utilizing algorithms to combine results of genetic variants into a report to guide healthcare practitioners in prescribing of antidepressants and choosing the dosing regimen based on whether the patient is a poor, normal, extensive or ultra-metabolizer.[8,9] The more recently developed second-generation tools differ from the first-generation tools in their ability to simultaneously assess and interpret multiple genetic markers.[6] Given that most psychiatric medications are processed by and interact with multiple biological pathways, the multiple genetic marker approach is

suggested to be important in evaluating pharmacotherapy and drug response.[6] However, the clinical effectiveness and cost effectiveness of these pharmacogenomics testing tools remain uncertain and has been the topic of much debate.[6]

The purpose of this report is to examine the clinical effectiveness, cost-effectiveness of pharmacogenomic testing versus treatment as usual for treating all severities of diagnosed depression. Additionally, evidence-based guidelines regarding the pharmacogenomic testing in patients with all severities of diagnosed depression will be reviewed.

Research Questions

1. What is the clinical effectiveness of pharmacogenomic testing for treating all severities of diagnosed depression?

2. What is the cost-effectiveness of pharmacogenomic testing for treating all severities of diagnosed depression?

3. What are the evidence-based guidelines for pharmacogenomic testing in patients with all severities of diagnosed depression?

Key Findings

This review was comprised of one health technology assessment report, two systematic reviews with meta-analyses, one randomized controlled trial, and three economic evaluations regarding pharmacogenomic testing versus treatment as usual for treating all severities of diagnosed depression.

One health technology assessment report suggested that the evidence for pharmacogenomic testing for depressive disorders was limited and of low to very low quality for different outcomes measured. The authors concluded that the evidence was insufficient for forming conclusions regarding clinical use. One systematic review with meta-analysis suggested that pharmacogenetic-guided prescribing had a positive effect on the likelihood of achieving symptom remission which may be confined to individuals with moderate to severe depression and a history of inadequate response or intolerability to previous psychotropic medications. One systematic review with meta-analysis suggested that the evidence was limited in quality and quantity and that primary studies suggestive of a positive effect of pharmacogenomic testing in major depressive disorders were mostly of low quality. One randomized controlled trial reported no significant difference in the improvement of depressive symptoms or safety outcomes between pharmacogenomic testing guided and unguided groups.

The health technology assessment report stated that results in one cost-effectiveness study suggested moderate cost-effectiveness of pharmacogenomics testing given the probability of having an incremental cost-effectiveness ratio below the international $1,926 cost-effectiveness threshold was 90%, while another study suggested that based on the commonly used threshold of $50,000 per quality-adjusted life year, pharmacogenomics testing would not be cost-effective. One included systematic review reported the probability of pharmacogenomics testing being cost-effective at the willingness to pay threshold of $50,000 was 94.5%. One[9] of the three included economic evaluations reported the lack of conclusion on cost-effectiveness of screening for CYP2D6 in primary care patients using antidepressants. Two economic evaluations reported that pharmacogenomics testing was dominant over treatment as usual.

One guideline within the health technology assessment report recommended that a combination of therapeutic drug monitoring and genotyping may be informative in potentially nonadherent patients.

Methods

Literature Search Methods

A limited literature search was conducted by an information specialist on key resources including Medline and PsycINFO via OVID, the Cochrane Library, the University of York Centre for Reviews and Dissemination (CRD) databases, the websites of Canadian and major international health technology agencies, as well as a focused Internet search. The search strategy was comprised of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concepts were pharmacogenomics and depression. No filters were applied to limit the retrieval by study type. Where possible, retrieval was limited to the human population. The search was also limited to English language documents published between January 1, 2014 and December 16, 2019.

Selection Criteria and Methods

One reviewer screened citations and selected studies. In the first level of screening, titles and abstracts were reviewed and potentially relevant articles were retrieved and assessed for inclusion. As pharmacogenetics and pharmacogenomics are often used interchangeably in the published literature,[7] articles that used either terminology were screened and considered for inclusion. The final selection of full-text articles was based on the inclusion criteria presented in Table 1.

Table 1: Selection Criteria

Population	Q1-3: Adults (18 to 60 years old)[a] with diagnosed depression of all severities (e.g., major depressive disorder, refractory depression)
Intervention	Q1-3: Guided care (e.g., guiding the drug selection or dose) with pharmacogenomic testing, either before or after treatment is initiated.
Comparator	Q1,2: Treatment as usual (e.g., no testing) Q3: Not applicable
Outcomes	Q1: Clinical effectiveness (e.g., response rate, remission rate, optimized dosing regimen, number of changes in treatment choice) and harm (e.g., adverse events, morbidity, mortality) Q2: Cost-effectiveness (e.g., cost per quality adjusted life years, cost per patient adverse event avoided, cost per clinical outcome) Q3: Recommendations regarding the use of pharmacogenomic testing for depression
Study Designs	Health technology assessments, systematic reviews, randomized controlled trials, non-randomized studies

[a] Studies that included adults without specifying the age range or systematic reviews that included adult population with a broader age range were included; as it was assumed that the majority of included adults would be between 18 and 60 years old.

Exclusion Criteria

Articles were excluded if they did not meet the selection criteria outlined in Table 1, they were duplicate publications, or were published prior to 2014. Guidelines with unclear

methodology were also excluded. Publications that included patients with dysthymia, Seasonal Affective Disorder, and post-partum depression were excluded.

Critical Appraisal of Individual Studies

The included systematic reviews (SRs) and health technology assessment (HTA) were critically appraised by one reviewer using A MeaSurement Tool to Assess systematic Reviews II (AMSTAR II),[10] randomized controlled trials (RCTs) were critically appraised using the Downs and Black Checklist,[11] and economic studies were assessed using the Drummond checklist.[12] Summary scores were not calculated for the included studies; rather, the strengths and limitations of each included study were described narratively.

Summary of Evidence

Quantity of Research Available

A total of 424 citations were identified in the literature search. Following screening of titles and abstracts, 389 citations were excluded and 35 potentially relevant reports from the electronic search were retrieved for full-text review. Five potentially relevant publications were retrieved from the grey literature search for full text review. Of these potentially relevant articles, 33 publications were excluded for various reasons, and seven publications met the inclusion criteria and were included in this report. These comprised one HTA report,[13] two SRs,[6,14] one RCT,[15] and three economic evaluations.[9,16,17] No non-randomized studies or evidence-based guidelines were identified. Appendix 1 presents the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[18] flowchart of the study selection. Additional references of potential interest are provided in Appendix 6: Additional References of Potential Interest.

Summary of Study Characteristics

Seven publications met the inclusion criteria and were included in this report. These comprised two SRs,[6,14] one HTA report,[13] one RCT,[15] and three economic evaluations.[9,16,17] No non-randomized studies or evidence-based guidelines not already included in the HTA report were identified. Additional details regarding the characteristics of included publications are provided in Appendix 2

Study Design

Of the seven included publications, one HTA report,[13] two SRs,[6,14] and one RCT[15] were identified regarding of pharmacogenomic testing versus treatment as usual for treating all severities of diagnosed depression. The HTA report,[13] published in 2016, searched for systematic reviews, meta-analyses, economic evaluations, primary clinical studies and practice guidelines regarding the clinical utility, clinical effectiveness, and cost-effectiveness of genetic testing.[13] The searches were conducted up to November 2016 and identified 14 primary studies and 12 guidelines, of which four clinical effectiveness studies and one guideline were relevant to the clinical effectiveness research question of this report.[13] One SR with meta-analysis (MA)[6], published in 2019, searched for RCTs up to May 2018. It included five RCTs, all of which were relevant to this report.[6] The second SR with MA,[14] published in 2017, searched for published reviews, MAs and primary studies up to October 2015. Ten studies were included, of which nine studies were relevant to this report, including four clinical effectiveness studies and five cost-effectiveness studies.[6] The

included RCT,[15] published in 2019, was a prospective single-blinded and single-center study, with data collected from September 2017 to July 2018.

One HTA report,[13] one SR,[14] and three economic evaluations[9,16,17] were identified regarding the cost-effectiveness of pharmacogenomic testing versus treatment as usual for treating all severities of diagnosed depression. The 2016 HTA report,[13] included three economic evaluation studies and one guideline that were relevant to the cost-effectiveness research question of this report. The 2017 SR [14] included five cost-effectiveness studies relevant to the research question. Additionally, three economic evaluation publications[9,16,17] were included in this Rapid Response report. The first economic study,[9] published in 2019, was a model-based cost-utility analysis from a Netherlands societal perspective with a Markov model. The time horizon was 12 weeks.[9] The study had a hypothetical cohort of 1000 patients and assumed no delay effect of therapy and a switch to a different drug in the case of a failed therapy.[9] The second economic study,[16] published in 2018, was a cost-effectiveness analysis with a Markov model with a time horizon of three years. It was conducted with a United States (US) societal perspective.[16] The study included 260 patients with moderate to severe depression from an RCT as the source of the clinical and cost inputs for the analysis.[16] The model assumed a three year catch-up period in which the standard of care group response becomes equal to the pharmacogenetic testing group.[16] The probability of spontaneous transition between response and nonresponse was assumed to be equivalent in both groups and was thus left at zero.[16] The third economic evaluation,[17] published in 2017, was a cost-effectiveness analysis from a US societal perspective with a Markov model with three years as the time horizon and a willingness-to-pay threshold of $50,000 per quality-adjusted life year (QALY). The study used a hypothetical cohort of 10,000 individuals with a baseline 17-item Hamilton depression scale (HAMD-17 or HDRS-17) score with diagnosis of depression with or without anxiety.[17] It assumed: both groups to be identical to control for variability in outcome due to differences in populations; patients who responded to treatment or achieved remission could relapse, requiring new treatment; all patients received citalopram or an equivalent medication; patients were considered in remission in they if they stopped receiving treatment over 6 months; relapse rates declined the longer a patient stayed in remission; the relative risk of suicide attempts decreased by 0.49 if the patient responded to treatment (based on the results of an observational study[19]); the occurrence of an adverse drug event negatively affected a patient's quality of life (QoL) and therefore penalized the QoL over that treatment period; and the quality of life for a patient in the remission state was identical to that of the general population.[17]

Country of Origin

The included HTA report[13] was by authors in the US. The two systematic reviews were by authors in Canada.[6,14]

The included RCT[15] was conducted in China.

One of the economic evaluations[9] was by authors in in the Netherlands. Two of the economic evaluations[16,17] were by authors in the US.

Patient Population

The included HTA report[13] included clinical effectiveness studies, cost effectiveness studies and guidelines for patients of any age who were being prescribed medications for treatment of depression, mood disorder, psychosis, anxiety, attention deficit hyperactivity disorder (ADHD), or substance use disorder. The patient population with depression was relevant to

this review and accounted for the 4 relevant studies.[13] Of the two included SRs,[6,14] the first SR[6] included RCTs regarding the efficacy of pharmacogenetic-guided decision support tools for antidepressant treatment. The patient population included adults diagnosed with MDD.[6] The second SR included RCTs, non-randomized studies, and cost-effectiveness studies assessing the effects of utilizing pharmacogenomic testing on improving clinical outcomes of MDD.[14] The eligible patient population was adults 18 to 75 years old.[14]

The included RCT[15] recruited patients aged 18 to 51 years with HAMD-17 total scores higher than 17 at baseline and the first item of the HAMD-17 (depressive mood) with a score higher than 2; who had never received psychiatric treatment or had interrupted antidepressant medication for more than 2 weeks; and with no psychotic symptoms. While the severity of depression associated with a HAMD-17 total score higher than 17 was not specified in this study,[15] this likely represents at least moderate depression.

The first included economic evaluation[9] modelled a hypothetical cohort of adult patients age ≥ 18 years with major depression in primary care. The patients were divided into three groups (poor metabolizers, extensive metabolizers, and ultra-metabolizers) based on prevalence data from the literature.[9] The second economic evaluation[16] used a study population of treatment-naïve patients with MDD or patients with inadequately controlled MDD and a score of 20 or greater on the HAMD-17, indicating moderate or severe depression. The mean age was 48 years old.[16] The third included economic evaluation[17] modelled a hypothetical cohort of patients with baseline HAMD-17 score ≥ 20 with or without a HAMA score ≥ 18, diagnosed with depression with or without anxiety. The mean age was 48 years old in both groups.[17]

Interventions and Comparators

The HTA report[13] included clinical effective and cost effectiveness studies that compared genetic testing (GeneSight,[20-22] CNSDose,[23] ABCB1,[24] test for at least one of CYP2D6, CYP2C19, CYP2C9, and/or serotonin transporter genotype 5-HTTLPR,[25] HTR2A,[26] 5-HTTLPR,[27] CYP2D6[28]) to usual care or no genetic testing. The interventions in the two SRs[6,14] were pharmacogenetic-guided treatment compared to unguided treatment.

The included RCT[15] compared pharmacogenetic testing for guided medication therapy to unguided therapy.

The three economic evaluations[9,16,17] compared pharmacogenomic testing to standard of care.

Outcomes

The HTA report[13] included clinical effectiveness studies, cost effectiveness studies and guidelines regarding genetic testing. The relevant outcomes measured were patient adherence to treatment regimen measured by percentage of drug therapy adherence; change in patient response to informed treatment (measured by treatment response with 50% reduction in HAMD-17, and remission with HAMD-17 <7, reduction in symptoms, Clinician Rated Quick Inventory of Depressive Symptomatology (QIDS-C16)) and change in adverse event rates as a result of informed treatment; cost-effectiveness measured by incremental QALY, or quality-adjusted life week (QALW) and incremental cost-effectiveness ratio (ICER), and recommendations regarding genetic testing. The outcome in one included systematic review[6] was remission measured by HDRS-17 (also known as HAMD-17). The outcomes in the other included systematic review[14] were depressive symptom severity measured by treatment response with 50% reduction in HAMD-17, and remission with

reduction in symptoms, remission rate (remission defined as HAMD-17 <7), and cost-effectiveness measured by change in QALYs and probability of test being cost effective at the willingness to pay (WTP) threshold of $50,000.

The included RCT[15] reported outcomes including depression outcomes measured by HAMD-17 total score, response rate (reduction ratio in HAMD-17 by ≥ 50%), remission rate (remission defined as HAMD-17 <7) Hamilton anxiety scale (HAMA) total score, routine blood test, liver function, renal function, electrocardiogram examination, and adverse reactions measured by incidence rate.

Three economic evaluation[9,16,17] estimated incremental cost, QALY, incremental QALY, ICER for comparing the pharmacogenetic testing guided group to the unguided group.

Summary of Critical Appraisal

Additional details regarding the strengths and limitations of included publications are provided in Appendix 3: Critical Appraisal of Included Publications.

Health Technology Assessment and Systematic Reviews

The included HTA report[13] had clearly defined research questions, objectives, and eligibility criteria. Key search terms and the dates of the searches were provided, increasing the reproducibility of the literature search, and literature searches were performed in multiple databases (PubMed and OVID-Embase).[13] In addition, the literature search included a grey literature search, decreasing the risk for missing relevant, non-database literature.[13] The review included a flow chart illustrating guideline selection and provided reasons for article exclusion.[13] The review included a list of the included and excluded studies.[13] Finally, the review authors stated that they had no related conflicts of interest.[13]

As for the limitations of the HTA report, it was not reported that review methods were prospectively registered in a published protocol, increasing the risk for selective reporting.[13] It was unclear if guideline selection, data extraction, and quality assessment were conducted in duplicate, increasing the risk for inconsistencies in these processes.[13] Lastly, the included primary studies were published in Spain, Australia, Germany, Italy and the United States and their relevance to the Canadian healthcare setting was unclear.[13]

The two included SRs[6,14] both had clearly defined research questions, objectives, and eligibility criteria. Protocols were established prior to the conduct of the review.[6,14] The authors used a comprehensive literature search strategy, providing the key search terms and searching multiple databases.[6,14] The included studies were described in adequate detail.[6,14] There were no concerns with the reported sources of funding and the potential conflicts of interest.[6,14], In one SR,[6] study selection and data extraction were performed in duplicate to improve consistency in the process.

There were also limitations in the included SRs.[6,14] The review authors for both SRs did not report the list of excluded studies.[6,14] In one SR,[6] the review authors did not justify their decision to limit to studies with an RCT design for inclusion in the review. In the second SR,[14] a protocol was mentioned but not reported to be registered prior to the conduct of the review,[14] which decreased transparency about the rigor of the SR. The study selection and data extraction were not reported to be performed in duplicate by two reviewers,[14] increasing the risk for inconsistencies in these processes. An investigation of publication bias and its impact on the results of the review were not reported.[14]

Randomized Controlled Trials

The strengths of the identified RCT[15] include clearly described objectives, main outcomes, characteristics, interventions, randomization, potential confounders and main findings. Patients from different treatment groups were recruited from the same population over the same time period.[15] The statistical tests used to assess the main outcomes were appropriate.[15] Patient adherence to the interventions was likely reliable with in-hospital observed administration of the medications.[15] The main outcome measure used was reliable as it was depression severity measured by a validated tool (HAMD-17).[15] Probability values were reported as p values for the main outcomes. Potential conflicts of interest were reported in the article.[15]

There were also several limitations identified in the included RCT.[15] The study was single-blinded, with the blinded group being the clinical psychiatrist, which may lead to response bias in patient reported outcomes.[15] It was unclear whether the patients who participated, staff, places, and facilities in the study in China were representative of the Canadian population.[15] The patients who were asked to participate and prepared to participate in the study were recruited via convenience sample (patients presenting to the hospital).[15] It was unclear whether they were representative of the entire population of adult patients with depression, which may lead to issues with the external validity of the studies.[15] A power calculation was not conducted to determine if the sample was of an adequate size for statistical significance.[15]

Economic Evaluations

In the three included economic evaluations,[9,16,17] the research question, economic importance of the research question, and rationale for choosing alternative interventions compared were clearly stated. The viewpoint/perspective of the analysis, treatment strategies being compared, form of economic evaluation were clearly stated.[9,16,17] The choice of form of economic evaluation was justified in relation to the questions addressed.[9,16,17] Additionally, the sources of effectiveness estimates and screening/treatment costs were provided.[9,16,17] The primary outcome measures for the economic evaluation, methods to value benefits, and time horizon of costs and benefits were clearly stated.[9,16,17] Details of the subjects from whom valuations were obtained and methods for the estimation of unit costs were reported.[9,16,17] With the use of figures, the structure of the model was clearly described.[9,16,17] Discount rates, choices of discount rates, and explanation of costs/benefits not being discounted were provided.[9,16,17] The approach to sensitivity analysis was given and the choice of variables for the sensitivity analysis were justified.[9,16,17] Major outcomes were presented in a disaggregated form and the answer to the study question was given. Incremental analysis was reported.[9,16,17] Conclusions follow from the data were reported and accompanied by appropriate caveats.[9,16,17] Lastly, the authors declared potential conflicts of interests and disclosed sources of funding.[9,16,17]

The included economic valuations[9,16,17] had various limitations. The viewpoint and perspective were not justified in all three studies.[9,16,17] The relevance of productivity changes and details of currency adjustments or conversions were not discussed.[9,16,17] Being economic evaluations with societal perspectives based in the Netherlands[9] or the United States[16,17], the results of these reports may not be and may not be generalizable to the Canadian health system. There was uncertainty with key parameters (e.g., treatment length, parameters, cycle length) in the model which may affect the ICER.[9,16,17] In one economic evaluation study,[9] a single study was the source of model parameters and studied one antidepressant treatment (venlafaxine; assumptions were made for other

antidepressants), which may overestimate the effect and results of this study. Multiple assumptions made in the study may affect the results and interpretation of the study.[9] In two of the included economic evaluation studies,[16,17] data inputs were taken from single references, rather than a synthesis or meta-analysis of estimates from multiple sources. In one economic evaluation study,[16] the data and studies used as model parameters were based on 12-week follow-up periods which may inaccurately depict the patient's response to treatment.[16] There may be a slight over-estimation of mortality rates for depression, as suicide rates were derived from a study focusing on treatment-resistant MDD.[16] Utility scores may be based on data from samples that were different from the populations that they were applied to affecting the overall results of the model, as utility scores for patients with MDD were taken from a study of patients at primary care centers being treated for depression.[16] There was uncertainty regarding the length of the follow-up of patients and treatment.[16] In another economic evaluation, the analysis used the same cost estimates for each severity level of depression.[17]

Summary of Findings

Appendix 4 presents a table of the main study findings and authors' conclusions.

Clinical Effectiveness of Pharmacogenomic Testing

Health Technology Assessment and Systematic Reviews

There was considerable overlap in the primary studies providing clinical effectiveness results that were included in the HTA report and the two SRs; results from individual primary studies are reported only once; therefore, the results from one included SR[14] are not described separately in the clinical effectiveness section of the report. A description of the overlap is presented in Appendix 5: Overlap between Included Systematic Reviews.

In the HTA report,[13] the remission outcome was reported in four relevant primary studies, as measured by various depression rating scales.[21-24] One study[22] reported a non-significant difference between the remission rate of pharmacogenomic testing guided group when compared to the unguided group. Three primary studies[21,23,24] reported that statistically significantly more patients in the guided group achieved remission when compared to the unguided group (at 12, 8 and 4 weeks, respectively), with one study[24] reporting uncertainty of the required change in 21-item Hamilton depression rating scale (HAM-D21) score's clinical relevance. The SR authors concluded that the quality of the evidence was low and the confidence that the results represent a true effect was also low, despite consistency of results favoring improved remission rates as a result of genotyping.[13] The response to treatment outcome was reported in four primary studies.[20-22,25] In one RCT[22] and poor-quality retrospective comparative study,[25] there was no significant differences between groups in treatment response. One non-randomized study[21] reported that statistically significantly more patients in the guided group than unguided group achieved a response. Another non-randomized study,[20] reported improved response for a statistically significantly more guided group patients than unguided group using HAM-D17 depression severity and QIDS-C16 scores. Two primary studies[23,24] reported on the adherence, tolerance, adverse events, and hospital stay outcomes for patients with depressive disorders. While the results of the two studies[23,24] were statistically significant favouring pharmacogenomic genotyping, the HTA report authors concluded that the evidence related to adverse events and to duration of hospital stay was of very low quality and insufficient for forming conclusions.[13]

In one included SR,[6] five relevant primary studies,[22,23,29-31] three of which [29-31] were not included in the included HTA report. However, due to the pooled risk ratio analysis of the results, all five primary studies[22,23,29-31] were taken into consideration for this report. The random-effects pooled risk ratio (RR) suggested a significant association between pharmacogenetic-guided prescribing and remission.[6]

Randomized Controlled Study

The included RCT[15] reported no significant differences in depressive symptoms measured by HAMD-17 total scores, response rate, remission rate, anxiety symptoms measured by HAMA total scores, adverse reactions and medication tolerability between guided and unguided groups at the end of the treatment. The authors concluded that pharmacogenomic testing may not significantly improve the efficiency and safety of the treatment for the guided group compared with those in the unguided group.[15]

Cost-Effectiveness of Pharmacogenomic Testing

Health Technology Assessment and Systematic Review

The HTA report[13] stated that results in one cost-effectiveness study[26] suggested moderate cost-effectiveness of pharmacogenomics testing given the probability of having an incremental cost-effectiveness ratio (ICER) below the international $1,926 cost-effectiveness threshold suggested by the World Health Organization was 90%. Cost-effectiveness analyses were performed using state-transition probability models.[13] The incremental benefit of the pharmacogenomic approach was “0.062 quality-adjusted life-weeks (QALW) for clinical response plus 0.016 QALWs for side effect burden” (p.60).[13] Assuming that patients will have 2 recurrent episodes, the overall incremental benefit of pharmacogenomic testing was 0.156 QALWs.[13] The incremental cost of pharmacogenomic testing was 179 international dollars (Intl $) and the ICER was Intl $1,147.[13] Multivariate sensitivity analyses were performed using estimated ICER values ranging from Intl $638 to Intl $1,738 (10th to 90th percentiles).[13] Another study[27] in the HTA report suggested that based on the commonly used threshold of $50,000 per quality-adjusted life year (QALY), pharmacogenomics testing would not be cost-effective. The discrepancies were reported to be due to the economic evidence base including studies of different designs and study populations, and differences in pharmacogenomic tests that were compared with no-test treatment regimens.[13]

One included systematic review[14] reported the probability of pharmacogenomics testing being cost-effective at the willingness to pay threshold of $50,000 was 94.5%, which suggested that pharmacogenomic testing could be a cost-effective intervention.

Economic Evaluations

One[9] of the three included economic evaluations[9,16,17] reported an inability to conclude that screening for CYP2D6 in primary care patients using antidepressants would be cost-effective. The QALYs were reported to be 0.146 for tested group and 0.145 for the non-tested group.[9] The ICER was reported to be €77,406 per QALY.[9] Forty-eight percent of the simulations were below the WTP threshold of €80,000 per QALY.[9]

Two economic evaluations[16,17] reported potentially improved QALYs in the guided group when compared to treatment as usual. In one economic evaluation study,[16] the incremental QALYs with guided treatment were 0.01 and 0.17 in moderately to severely depressed patients and severely depressed patients, respectively. The ICER for the pharmacogenetics testing-guided group was reported to dominate the standard of care group (i.e., more

QALYs gained and cost savings with pharmacogenetic testing than with standard of care).[16] Another economic evaluation study[17] included in this report stated the incremental QALY to be 0.15, with the ICER of pharmacogenetics testing-guided group dominating the treatment as usual group.[17]

Evidence-based Guidelines

One guideline[32] within the HTA report[13] recommended that a combination of therapeutic drug monitoring (TDM) and genotyping may be informative in potentially nonadherent patients (for example, patients with low drug plasma levels despite high doses of the antidepressant). The guideline authors suggested that TDM and genotyping could help identify slow or rapid metabolizers of certain antidepressants.[13] The guideline was reported to be of fair quality, with the limitation that the search terms, date ranges of the literature search, criteria for selecting evidence, and how the body of evidence was evaluated for bias were not reported.[13]

No other relevant evidence-based guidelines regarding the pharmacogenomic testing for depression was identified.

Limitations

A number of limitations were identified in the critical appraisal tables in Appendix 3: Critical Appraisal of Included Publications, however, additional limitations exist.

Even though a HTA report,[13] and two SRs[6,14] were included in this report, the primary studies in these included publications were of low to moderate quality, as reported by the HTA and SR authors.[6,13,14]

As most included studies were conducted in countries outside of Canada[9,13,15-17] (with two exceptions of systematic reviews conducted in Canada but which included studies from other countries[6,14]), the applicability of the evidence to Canadian settings was unclear. With the different demographic components and health care systems, determining whether evidence is relevant and able to be generalized to the Canadian context requires an assessment of the differences in the health care systems. The availability of the testing tools and antidepressants used in the studies are unknown in Canada.

There was a gap in the recent evidence regarding the cost-effectiveness of patients with mild depression with or without anxiety and the evidence-based guideline recommendations regarding the use of pharmacogenomic testing in guiding depression treatment.

Conclusions and Implications for Decision or Policy Making

This report provides a summary of recent evidence regarding the clinical effectiveness, cost-effectiveness and guidelines regarding pharmacogenomic testing for treating all severities of diagnosed depression. This review was comprised of one HTA report,[13] two SRs with meta-analyses,[6,14] one RCT,[15] and three economic evaluations[9,16,17] regarding pharmacogenomic testing versus treatment as usual for treating all severities of diagnosed depression.

Regarding the clinical effectiveness of pharmacogenomic testing for treating all severities of diagnosed depression, the evidence was found to be inconclusive with respect to depression severity outcomes and safety outcomes. Evidence of limited quality from one HTA report[13] suggested that the evidence for pharmacogenomic testing for depressive

disorders was limited and of low to very low quality for different outcomes measured. The authors concluded that the evidence was insufficient for forming conclusions regarding clinical use.[13] One SR with meta-analysis[6] suggested that pharmacogenetic-guided prescribing had a positive effect on the likelihood of achieving symptom remission which may be confined to individuals with moderate to severe depression and a history of inadequate response or intolerability to previous psychotropic medications. One SR with meta-analysis[14] suggested that the evidenced was limited and that primary studies suggestive of a positive effect of pharmacogenomic testing in major depressive disorders were mostly of low quality. One RCT[15] reported no significant difference in the improvement of depressive symptoms or safety outcomes between pharmacogenomic testing guided and unguided groups. It may be premature to draw conclusions about the comparative effectiveness of pharmacogenomic testing versus treatment as usual given the mixed results of clinical evidence[6,13-15] for this comparison.

Regarding the cost-effectiveness of pharmacogenomic testing, the results were also variable.[9,13,14,16,17] The HTA report[13] stated that the results in one cost-effectiveness study[26] suggested moderate cost-effectiveness of pharmacogenomics testing given the probability of having an incremental cost-effectiveness ratio below the international $1,926 cost-effectiveness threshold was 90%, while another study[27] suggested that based on the commonly used threshold of $50,000 per QALY, pharmacogenomics testing would not be cost-effective. One included SR[14] reported the probability of pharmacogenomics testing being cost-effective at the willingness to pay threshold of $50,000 was 94.5%, based on a cost-effectiveness study[33] included in the SR.[13] One[9] of the three[9,16,17] included economic evaluations reported the lack of conclusion on cost-effectiveness of screening for CYP2D6 in primary care patients using antidepressants.[9] Two economic evaluations[16,17] reported that pharmacogenomics testing was dominant over unguided treatment.

Regarding evidence-based guideline recommendations on the topic, one guideline[32] within the HTA report[13] recommended that a combination of therapeutic drug monitoring and genotyping may be informative in potentially nonadherent patients. There was a general paucity of guideline recommendations on this topic.

In 2017, Health Quality Ontario (HQO) published an HTA report[34] on the Assurex GeneSight psychotropic test, a pharmacogenomic testing tool for psychotropic medication selection. The HQO report[34] was relevant but not included in this report due to the complete overlap of included primary studies with the included HTA report[13] and SRs.[6,14] The overall conclusions by the HQO authors were that there is uncertainty about the use of GeneSight Psychotropic test to guide medication selection.[34] When compared to treatment as usual, the GeneSight test was associated with improvements in response to depression therapy, measures of depression, and patient and clinician satisfaction.[34] No observed differences in rates of complete remission were found between groups.[34] The HQO authors reported low confidence in these findings due to the limitations in the body of evidence.[34]

The limitations of the included studies and of this report should be considered when interpreting the results. Additional studies of high methodological quality may further aid in making definitive conclusions about pharmacogenomic testing for treating all severities of diagnosed depression.

References

1. Public Health Agency of Canada. What is depression? 2016; https://www.canada.ca/en/public-health/services/chronic-diseases/mental-illness/what-depression.html. Accessed 2020 Jan 31.
2. Knoll AD, MacLennan RN. Prevalence and correlates of depression in Canada: findings from the Canadian Community Health Survey. *Can Psychol.* 2017;58(2):116-123.
3. Zimmerman M, Martinez JH, Young D, Chelminski I, Dalrymple K. Severity classification on the Hamilton Depression Rating Scale. *J Affect Disord.* 2013;150(2):384-388.
4. Tansey KE, Guipponi M, Hu X, et al. Contribution of common genetic variants to antidepressant response. *Biol Psychiatry.* 2013;73(7):679-682.
5. Concetta C, Fabbri C, Porcelli S, et al. Pharmacogenetics of antidepressants. *Front Pharmacol.* 2011;2:6.
6. Bousman CA, Arandjelovic K, Mancuso SG, Eyre HA, Dunlop BW. Pharmacogenetic tests and depressive symptom remission: a meta-analysis of randomized controlled trials. *Pharmacogenomics.* 2019;20(1):37-47.
7. Bishop JR. Neurogenetics, part I. In: Steegers E, Cipolla M, Miller E, eds. *Handbook of clinical neurology*: Elsevier; 2018.
8. Zeier Z, Carpenter LL, Kalin NH, et al. Clinical implementation of pharmacogenetic decision support tools for antidepressant drug prescribing. *Am J Psychiatry.* 2018;175(9):873-886.
9. Sluiter RL, Janzing JGE, van der Wilt GJ, Kievit W, Teichert M. An economic model of the cost-utility of pre-emptive genetic testing to support pharmacotherapy in patients with major depression in primary care. *Pharmacogenomics J.* 2019;19(5):480-489.
10. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017;358:j4008. http://www.bmj.com/content/bmj/358/bmj.j4008.full.pdf. Accessed 2020 Jan 31.
11. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health.* 1998;52(6):377-384. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1756728/pdf/v052p00377.pdf. Accessed 2020 Jan 31.
12. Higgins JPT, Green S, editors. Figure 15.5.a: Drummond checklist (Drummond 1996). *Cochrane handbook for systematic reviews of interventions*. London (GB): The Cochrane Collaboration; 2011: http://handbook-5-1.cochrane.org/chapter_15/figure_15_5_a_drummond_checklist_drummond_1996.htm. Accessed 2020 Jan 31.
13. Pharmacogenomic testing for selected conditions: final evidence report. Olympia (WA): Washington State Health Care Authority; 2016: https://www.hca.wa.gov/assets/program/pharmacogenomics-final-rpt-20161209.pdf. Accessed 2020 Jan 31.
14. Rosenblat JD, Lee Y, McIntyre RS. Does pharmacogenomic testing improve clinical outcomes for major depressive disorder? A systematic review of clinical trials and cost-effectiveness studies. *J Clin Psychiatry.* 2017;78(6):720-729.
15. Shan X, Zhao W, Qiu Y, et al. Preliminary clinical investigation of combinatorial pharmacogenomic testing for the optimized treatment of depression: a randomized single-blind study. *Front Neurosci.* 2019;13:960.
16. Groessl EJ, Tally SR, Hillery N, Maciel A, Garces JA. Cost-effectiveness of a pharmacogenetic test to guide treatment for major depressive disorder. *J Manag Care Spec Pharm.* 2018;24(8):726-734.
17. Najafzadeh M, Garces JA, Maciel A. Economic evaluation of implementing a novel pharmacogenomic test (IDgenetix) to guide treatment of patients with depression and/or anxiety. *Pharmacoeconomics.* 2017;35(12):1297-1310.
18. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol.* 2009;62(10):e1-e34.
19. Olin B. Mortality and suicide risk in treatment-resistant depression: an observational study of the long-term impact of intervention. *PLoS One.* 2012;7(10):e48002.
20. Hall-Flavin DK, Winner JG, Allen JD, et al. Using a pharmacogenomic algorithm to guide the treatment of depression. *Transl Psychiatry.* 2012;2(10):e172.
21. Hall-Flavin DK, Winner JG, Allen JD, et al. Utility of integrated pharmacogenomic testing to support the treatment of major depressive disorder in a psychiatric outpatient setting. *Pharmacogenet Genomics.* 2013;23(10):535-548.
22. Winner JG, Carhart JM, Altar CA, Allen JD, Dechairo BM. A prospective, randomized, double-blind study assessing the clinical impact of integrated pharmacogenomic testing for major depressive disorder. *Discov Med.* 2013;16(89):219-227.
23. Singh AB. Improved antidepressant remission in major depression via a pharmacokinetic pathway polygene pharmacogenetic report. *Clin Psychopharmacol Neurosci.* 2015;13(2):150-156.
24. Breitenstein B, Scheuer S, Holsboer F. Are there meaningful biomarkers of treatment response for depression? *Drug Discov Today.* 2014;19(5):539-561.
25. Rundell JR, Harmandayan M, Staab JP. Pharmacogenomic testing and outcome among depressed patients in a tertiary care outpatient psychiatric consultation practice. *Transl Psychiatry.* 2011;1:e6.

26. Perlis RH, Patrick A, Smoller JW, Wang PS. When is pharmacogenetic testing for antidepressant response ready for the clinic? A cost-effectiveness analysis based on data from the STAR*D study. *Neuropsychopharmacology.* 2009;34(10):2227-2236.

27. Olgiati P, Bajo E, Bigelli M, Montgomery S, Serretti A, C.E.A.P (Cost-Effectiveness Analysis on Psychiatric Disorders)-group. Challenging sequential approach to treatment resistant depression: cost-utility analysis based on the sequenced treatment alternatives to relieve depression (STAR* D) trial. *Eur Neuropsychopharmacol.* 2012;23(12):1739-1746.

28. Herbild L, Bech M, Gyrd-Hansen D. Estimating the Danish populations' preferences for pharmacogenetic testing using a discrete choice experiment. The case of treating depression. *Value Health.* 2009;12(4):560-567.

29. Perez V, Salavert A, Espadaler J, et al. Efficacy of prospective pharmacogenetic testing in the treatment of major depressive disorder: results of a randomized, double-blind clinical trial. *BMC Psychiatry.* 2017;17(1):250.

30. Bradley P, Shiekh M, Mehra V, et al. Improved efficacy with targeted pharmacogenetic-guided treatment of patients with depression and anxiety: a randomized clinical trial demonstrating clinical utility. *J Psychiatric Res.* 2018;96:100-107.

31. Greden JF, Rosthschild AJ, Thase M, et al. Combinatorial pharmacogenetics significantly improves response and remission for major depressive disorder: a large, blinded, randomized controlled trial. *CNS Spectr.* 2018;24(1):202-203.

32. Bauer M, Pfennig A, Severus E, Whybrow PC, Angst J, Moller H-J. World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders, part 1: update 2013 on the acute and continuation treatment of unipolar depressive disorders. *World J Biol Psychia.* 2013;14:334-358.

33. Hornberger J, Li Q, Quinn B. Cost-effectiveness of combinatorial pharmacogenomic testing for treatment-resistant major depressive disorder patients. *Am J Manag Care.* 2015;21(6):e357-365.

34. Health Quality O. Pharmacogenomic testing for psychotropic medication selection: a systematic review of the Assurex GeneSight psychotropic test. *Ont Health Technol Assess Ser.* 2017;17(4):1-39.

35. Winner JG, Carhart JM, Altar CA, et al. Combinatorial pharmacogenomic guidance for psychiatric medications reduces overall pharmacy costs in a 1 year prospective evaluation. *Curr Med Res Opin.* 2015;31(9):1633-1643.

36. Trangle M, Gursky J, Haight R, et al. Adult depression in primary care. Bloomington (MN): Institute for Clinical Systems Improvement; 2016: https://roar.nevadaprc.org/system/documents/4068/original/NPRC.3054.Depr-Interactive0512b.pdf?1471560355. Accessed 2020 Jan 31.

37. Espadaler J, Tuson M, Lopez-Ibor JM, Lopez-Ibor F. Pharmacogenetic testing for the guidance of psychiatric treatment: a multicenter retrospective analysis. *CNS Spectr.* 2016;22(4):315-324.

38. VA/DoD clinical practice guideline for the management of major depressive disorder. Washington (DC): Department of Veterans Affairs, Department of Defense; 2016: https://www.healthquality.va.gov/guidelines/MH/mdd/. Accessed 2020 Jan 31.

39. Brennan FX, Gardner KR, Lombard J, et al. A naturalistic study of the effectiveness of pharmacogenetic testing to guide treatment in psychiatric patients with mood and anxiety disorders. *Prim Care Companion CNS Disord.* 2015;17(2).

40. Oslin DW, Leong SH, Lynch KG, et al. Naltrexone vs placebo for the treatment of alcohol dependence: a randomized clinical trial. *JAMA Psychiatry.* 2015;72(5):430-437.

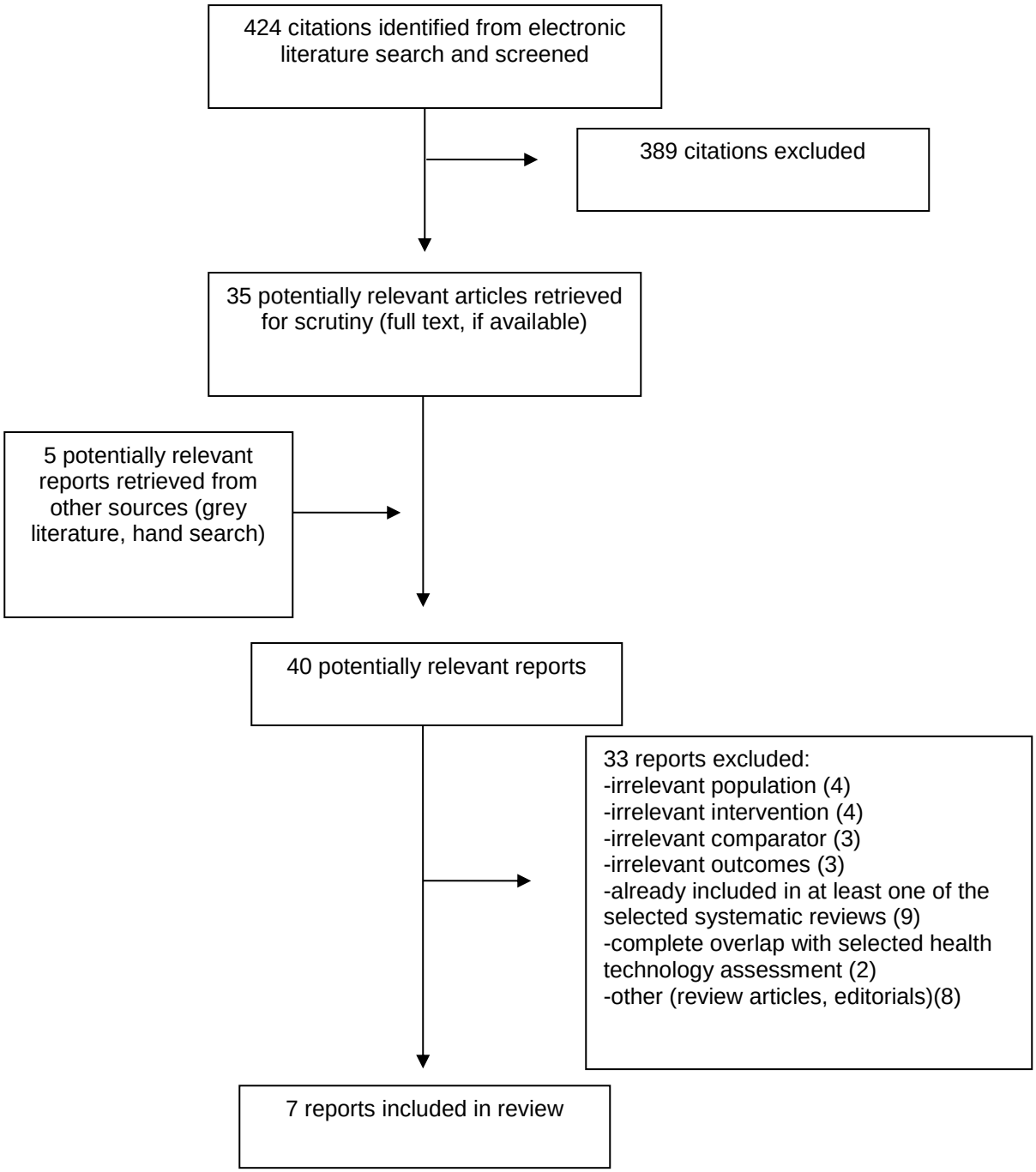
41. Fagerness J, Fonseca F, Hess GP, et al. Pharmacogenetic-guided psychiatric intervention associated with increased adherence and cost savings. *Am J Manag Care.* 2014;20(5):e146-156.

42. Herbild L, Andersen SE, Werge T, Rasmussen HB, Jurgens G. Does pharmacogenetic testing for CYP450 2D6 and 2C19 among patients with diagnoses within the schizophrenic spectrum reduce treatment costs? *Basic Clin Pharmacol Toxicol.* 2013;113(4):266-272.

43. Möller HJ, Bitter I, Bobes J, et al. Position statement of the European Psychiatric Association (EPA) on the value of antidepressants in the treatment of unipolar depression. *Eur Psychiatry.* 2011;27(2):114-128.

44. McDermott B, Baigent M, Chanen A, et al. Clinical practice guidelines: depression in adolescents and young adults. Melbourne (AU): Australian Government; 2010: http://www.smartteen.net/main/eng/_admin/download/4-662-1459406819.pdf. Accessed 2020 Jan 31.

Appendix 1: Selection of Included Studies



Appendix 2: Characteristics of Included Publications

Table 2: Characteristics of Included Systematic Reviews and Meta-Analyses

First Author, Publication Year, Country	Study Designs and Numbers of Primary Studies Included	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
Health Technology Assessment				
Washington State, 2016[13] **United States**	**Study design:** HTA **Literature search strategy:** The authors searched PubMed (January 1, 2000 to August 15, 2016), OVID-Embase (1996 to 2016, week 33) and PsycINFO (1987 to July, week 4, 2016) databases for systematic reviews, meta-analyses, economic evaluations, primary clinical studies and practice guidelines; searches were updated to November 28, 2016 **Included study characteristics:** 14 studies, 12 guidelines included Included studies published 2009-2016 Relevant studies: 4 clinical effectiveness studies • Singh, 2015[23] • Winner, 2013[22] • Breitenstein, 2014[24] • Winner, 2015[35] 3 economic evaluation studies: • Perlis, 2009[26] • Olgiati, 2012[27] • Herbild, 2009[28] 1 guideline • WFSBP (Bauer, 2013)[32] **Quality assessment tool:** Hayes Evidence-Grading Guides and Quality Checklists **Objective**: to determine the clinical utility, clinical effectiveness, and cost-effectiveness of genetic testing	N = NR **Included:** Patient population was people of any age who were being prescribed medications for treatment of depression, mood disorder, psychosis, anxiety, ADHD, or substance use disorder. The interventions were clinical laboratory tests for genetic variants in targeted genes or in panels of genes. Test results were available to the medication prescriber in the experimental arm of the study. The settings were inpatient and outpatient settings. **Excluded:** NR	**Intervention:** GeneSight,[20-22] CNSDose,[23] ABCB1,[24] test for at least one of CYP2D6, CYP2C19, CYP2C9, and/or serotonin transporter genotype 5-HTTLPR,[25] HTR2A,[26] 5-HTTLPR,[27] CYP2D6,[28]) **Comparator:** usual care or no genetic testing	**Outcomes:** Patient Management: physician and patient decision-making regarding drug choice and/or dose; patient adherence to treatment regimen measured by percentage of drug therapy adherence; change in patient response to informed treatment (measured by treatment response with 50% reduction in HAMD-17, and remission with HAMD-17 <7, reduction in symptoms, Clinician Rated Quick Inventory of Depressive Symptomatology (QIDS-C16)) and change in adverse event rates as a result of informed treatment; cost-effectiveness measured by incremental QALY, or quality-adjusted life week (QALW) and incremental cost-

First Author, Publication Year, Country	Study Designs and Numbers of Primary Studies Included	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
				effectiveness ratio (ICER), and recommendations regarding genetic testing **Length of follow-up**: 5-12 weeks
Systematic Reviews with Meta-Analyses				
Bousman, 2019[6] **Canada**	**Study design:** SR with MA **Literature search strategy:** The authors searched Embase, PsycINFO, Ovid Medline, ScienceDirect, Google Scholar and the Cochrane Central Register of Controlled Trials database up to May 2018. Studies published in English were included. **Included study characteristics:** 5 RCTs included Included studies published 2013-2018 Relevant studies: 5 RCTs **Quality assessment tool:** Cochrane Risk of Bias Assessment Tool **Objective:** To conduct a SR and MA of prospective, RCTs to examine the remission rates of pharmacogenetic-guided decision support tools relevant to depressive symptom remission in MDD	N = 1,737 patients **Included:** adult participants (aged ≥18) diagnosed with MDD Mean age: NR	**Included interventions:** pharmacogenetic-guided treatment **Relevant Interventions:** unguided treatment	**Outcomes:** Remission measured by HDRS-17 (score ≤ 7) **Length of follow-up**: 8-12 weeks
Rosenblat, 2017[14] **Canada**	**Study design:** SR with MA **Literature search strategy:** The authors searched MEDLINE/PubMed and Google Scholar databases were searched from inception to October 2015 for published reviews, meta-analyses and primary studies.	**Included:** adults (18-75 years old) Mean age: NR **Excluded:** NR	**Included interventions:** pharmacogenomic testing-guided treatment **Relevant Interventions:** unguided treatment	**Outcomes:** Depressive symptom severity, remission rate, cost-effectiveness

First Author, Publication Year, Country	Study Designs and Numbers of Primary Studies Included	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
	Included study characteristics: 10 studies included Included studies published 2009-2015 Relevant studies: 4 clinical effectiveness studies and 5 cost-effectiveness studies **Quality assessment tool:** Newcastle-Ottawa Scale **Objective**: to determine the cost effectiveness and the effect on clinical outcomes of pharmacogenomic testing-guided treatment in the treatment of MDD as compared to unguided treatment			**Length of follow-up**: 8 weeks-3 months

AHRQ = Agency for Healthcare Research and Quality; DST = decision support tool; HDRS-17 = Hamilton Depression Rating Scale-17; HTA = health technology assessment; MA = meta-analysis; MDD = major depressive disorder; NR = not reported; RCT = randomized controlled trial; SR = systematic review; WFSBP = World Federation of Societies for Biological Psychiatry.

Table 3: Characteristics of Included Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
Shan, 2019[15] **China**	**Study design:** single-blinded RCT **Setting:** inpatients and outpatients of Department of Psychiatry of the Second Xiangya Hospital, Central South University from September 2017 to July 2018. **Objective:** to assess the effectiveness pharmacogenetic testing for medication therapy selection in patients with depression.	**Inclusion criteria:** Patients aged 18 to 51 years of age with a HAMD-17 total scores of > 17 at baseline and the first item of the HAMD-17 (depressive mood) > 2; never received psychiatric treatment or have interrupted antidepressant medication for more than 2 weeks (fluoxetine for at least 4 weeks); and with no psychotic symptoms **Excluded:** having any other psychiatric	**Intervention of interest**: pharmacogenetic testing for guided medication therapy selection **Comparator**: Unguided therapy selection	**Relevant Outcome:** depression outcomes measured by HAMD-17 total score, response rate, remission rate, HAMA total score, blood routine, liver function, renal function, and electrocardiogram examination, adverse reactions measured by TESS **Length of follow-up**: 8 weeks

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
		diagnoses according to DSM-5; any physical illness such as liver and kidney diseases, cardiovascular diseases; any combination with other antipsychotic medications (both low and high doses), including typical and atypical antipsychotic and mood stabilizer; and pregnancy **Number of patients:** 71 patients n = 31 in guided group n= 40 in unguided group **Mean age:** 26.52 ± 7.92 years in guided group; 28.85 ± 8.93 years in unguided group		

DSM-5 = Structural Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, fifth edition; HAMA = Hamilton anxiety scale; HAMD-17 = 17-item Hamilton depression scale; RCT = randomized controlled trial; TESS = treatment emergent symptom scale.

Table 4: Characteristics of Included Economic Evaluations

First Author, Publication Year, Country	Type of Analysis, Time Horizon, Perspective	Decision Problem	Population Characteristics	Intervention and Comparator(s)	Clinical and Cost Data Used in Analysis	Main Assumptions
Sluitter, 2019 [9] **Netherlands** Funding source: Grant by KNMP	**Analysis**: CUA **Approach**: Model-based analysis; Markov model	*"The aim of this study was to assess the likely cost-utility of PGx for CYP2D6 genotypes before the start of antidepressant drug treatment,*	Adult patients with major depression (age ≥ 18 years) in primary care. EM - one or two active alleles of CYP2D6 and a normal metaboliser activity;	**Intervention:** PGx **Comparator:** no PGx group = patients receiving standard care	**Clinical Inputs/Probabilities:** - Probabilities of being EM, PM or UM - Probabilities of SE for EM, PM or UM - Probabilities of having no SE for EM, PM, UM - After PGx, probability of PM having SEs - After PGx, probability of PMs having no SEs	- Base case assumed no delay effect of therapy - Case of a failed attempt, a switch to a different drug was assumed, with outcomes similar to those of patients who did not show clinical improvement and who did not attempt suicide. - Concerning treatment effects, a general effect among

First Author, Publication Year, Country	Type of Analysis, Time Horizon, Perspective	Decision Problem	Population Characteristics	Intervention and Comparator(s)	Clinical and Cost Data Used in Analysis	Main Assumptions
	Time horizon: 12 weeks **Cycle length:** 12 weeks **Perspective**: Netherlands societal **Discount rate**: None **Analytic Approach:** Univariate sensitivity analysis; scenario analysis	*compared to no PGx." (pg. 2)*	PM - no active allele of CYP2D6, and a decreased metabolic activity UM - more than two active alleles of CYP2D6 and increased enzymatic activity **Number of participants:** Hypothetical cohort of 1000 patients was used in the model **Mean age:** Not specified		- Probability having a treatment effect on PM - Probabilities of having no treatment effect being PM, EM, or UM - Probabilities of having a treatment effect being PM, EM or UM - Probability of titration - Probability of switching to another class - Probability of waiting for treatment decision due to SE - Probability of suicide attempt - Probability of fatal suicide attempt **Cost Inputs:** - Associated with suicide attempts - Fatal suicide - PGx - GP visit - Psychiatrist visit - Due to labor loss - Prescription cost **Utility Inputs:** - Being depressed - No SEs - After titration - Switching - Waiting - Effect (remission) - No effect - Suicide attempt	depressive patients was assumed for the EM class. - Due to the lack of evidence for PM and UM metabolisers, assumptions on treatment effects had to be made. - For the PM group with an increased antidepressant blood level, the same treatment effect as for the EM group was assumed. - For the UM group with a decrease antidepressant blood level, a smaller treatment effect was assumed. - Probabilities of dosage titration, switching or waiting due to SE or no treatment effect were assumed to be equal for those without treatment effects, regardless of the metaboliser classes.
Groessl, 2018[16] **United States** **Funding source**: AltheaDx, manufact	**Analysis**: CEA **Approach**: Model-based analysis; Markov model	*"To evaluate the cost-effectiveness of a IDGx that has demonstrated effectiveness compared with SOC medication management*	Treatment naïve patients with MDD or patients with inadequately controlled MDD and a score of 20 or greater on the HAM-D17	**Intervention**: IDGx **Comparator**: SOC	**Clinical Inputs/Probabilities:** - Clinical response measured by HAMD-17 - Suicide rates for non-responders and responders - RR of all-cause mortality (responders and non-responders)	- Model assumed a 3 year catch up period in which the SOC group response becomes equal to the IDGx group. - The probability of spontaneous transition between response and nonresponse was assumed to be equivalent in both arms and was thus left at zero.

First Author, Publication Year, Country	Type of Analysis, Time Horizon, Perspective	Decision Problem	Population Characteristics	Intervention and Comparator(s)	Clinical and Cost Data Used in Analysis	Main Assumptions
urer of the IDgenetix	**Time horizon**: 3 years **Perspective**: US Societal **Discount rate**: 2.5-3.5% **Analytic Approach:** Univariate, one-way sensitivity analysis	*among patients with varied MDD severity."* (pg.1)	**Number of participants**: 260 participants with moderate to severe depression from an RCT went through the Markov model. **Mean age**: 48 years **Sex**: 100% female		**Cost Inputs:** - IDGx test - Annual direct medical costs (responder and non-responders - Annual indirect medical costs (responders and non-responders) **Utility Inputs:** - Responders - Non-responders	

Najafzadeh 2017 [17]
 	
United States
 	
Funding source: No funding received
	Analysis: CEA
	
	Approach: Model-based analysis; Markov model
	
	Time horizon: 3 years (1 year, 2 years, 1o years and lifetime conducted)
	
	Cycle length: 3 months
	
	Perspective: US Societal
	
	Discount rate: 3%
		"In this study, we aimed to assess the cost effectiveness of using IDgenetix-guided treatment compared with SOC in patients with moderate or severe depression and/or anxiety, based on the results of the published RCT and other published observational studies. The secondary aim of this study was to evaluate the cost-effectiveness results in different	Hypothetical cohort of individuals with baseline HAMD score ≥ 20 and/or a HAMA score ≥ 18[30], diagnosed with depression with/without anxiety
			
			Number of participants: 10,000 individuals
			
			Mean Age: 48 years in both groups
			
			Sex: 27% male and 73% females in both groups
			
			Anxiety and Depression: 65% of patients had both anxiety and depression;
				Intervention: IDgenetix (pharmacogenetic therapy)
				
				Comparator: SOC
					Clinical inputs/probabilities:
					- Decline in remission rates (Level 1, 2, 3 or 4)
					  ○ Level 1 treatmemt = patients receiving citalopram or an equivalent medication
					  ○ level 2-4 = non-responders and includedoptions such as bupropion, cognitive therapy, sertraline, extended-release venlafaxine, tranylcypromine, or extended-release venlafaxine plus mirtazapine
					- Decline in response rates (Level 1, 2, 3 or 4)
					- Increase in ADEs rate by treatment level (Level 1, 2, 3 or 4)
					- Suicide attempts among patients with depression
						- Assumed both groups to be identical to control for variability in outcome due to differences in populations
						- assumed that patients who responded to treatment or achieved remission could relapse, requiring new treatment
						- Assumed all patients received Treatment 1 (receiving citalopram or an equivalent medication.)
						- Assumed a 3-year time horizon for base case analysis
						- Assuming at WTP threshold of $50,000 per QALY
						- Patients were considered in remission in they if they stopped receiving treatment over two courses (6 months)
						- Assumed that relapse rates declined the longer a patient stayed in remission
						- Assumed that the relative risk of suicide attempts decreased by 0.49 if the patient responded to treatment [and assigned a

First Author, Publication Year, Country	Type of Analysis, Time Horizon, Perspective	Decision Problem	Population Characteristics	Intervention and Comparator(s)	Clinical and Cost Data Used in Analysis	Main Assumptions
	Analytic Approach: Univariate, one-way sensitivity analysis	*patient subgroups and under different assumptions about model parameters." (pg. 2)*	35% had anxiety only		- RR risk of suicide attempt among those who respond to treatment - Proportion of suicides considered successful - Relapse rates for remission patients (Level 1, 2, 3 or 4) - Relapse rates for responding patients (Level 1, 2, 3 or 4) **Cost inputs:** - Direct costs to depression/anxiety per year - Indirect costs to depression/anxiety per year - IDgenetix costs **Utility inputs:** - QoL for remission - QoL for response - QoL for no response - ADE impact on QoL	rate of zero for those who achieved remission - Assumed that the occurrence of an ADE negatively affected a patient's QoL and therefore penalized the QoL over that treatment period. - Assumed that the QoL for a patient in the remission state was identical to that of the general population

ADE = adverse drug event; CEA = cost-effectiveness analysis; CUA = cost-utility analysis; EM = extensive metabolizers; HAM-D17 =17 item Hamilton Rating Scale for Depression; HAMA = Hamilton Rating Scale for Anxiety; IDGx = pharmacogenetic testing; KNMP = Royal Dutch Pharmacists Association; MDD = major depressive disorder; PGx = pharmacogenetic screening; PM = poor metabolizers; QALY= quality-adjusted life year; QoL = quality of life; RCT = randomized controlled trial; SE= side effects; SOC = standard of care; UM = ultra-metabolizers; WTP = willingness-to-pay.

Appendix 3: Critical Appraisal of Included Publications

Table 5: Strengths and Limitations of Systematic Reviews and Meta-Analyses using AMSTAR II[10]

Strengths	Limitations
Washington State, 2016[13]	
• The research questions and inclusion criteria for the review included the components of population, intervention, comparison, and outcomes. • The review authors used a comprehensive literature search strategy. • The review included a flow chart illustrating guideline selection and provided reasons for article exclusion • The review authors explained their selection of all study designs for inclusion in the review. • The included studies were described in adequate detail. • A list of excluded studies was published. • The sources of funding for the included studies were reported. The potential sources of conflict of interest and funding were disclosed.	• It was unclear whether a published protocol was established prior to the conduct of the review. • It was unclear whether the study selection and data extraction were performed in duplicate by two reviewers. • The included primary studies were published in Spain, Australia, Germany, Italy and the United States and their relevance to the Canadian healthcare setting was unclear.
Bousman, 2019[6]	
• The research questions and inclusion criteria for the review included the components of population, intervention, comparison, and outcomes. • A protocol was established prior to the conduct of the review. The PROSPERO registration number was provided in the report. • The review authors used a comprehensive literature search strategy. • The literature search, study selection and data extraction were performed in duplicate by two reviewers. • The included studies were described in adequate detail. • The authors used a satisfactory technique, the Cochrane Risk of Bias Assessment Tool, for assessing the risk of bias of included individual studies • The authors justified combining the data in a meta-analysis. The discussion and explanation of any heterogeneity observed in the results of the review was reported. • The sources of funding for the included studies were reported. • The potential sources of conflict of interest and funding were disclosed.	• A list of excluded studies was not published. • The review authors did not explain their selection of RCT study design for inclusion in the review.
Rosenblat, 2017[14]	
• The research questions and inclusion criteria for the review included the components of population, intervention, comparison, and outcomes. • The review authors used a comprehensive literature search strategy.	• A protocol was established but not registered prior to the conduct of the review. • The study selection and data extraction were not reported to be performed in duplicate by two reviewers. • A list of excluded studies was not published.

Strengths	Limitations
• The review authors explained their selection of all study designs for inclusion in the review. • The included studies were described in adequate detail. • The sources of funding for the included studies were reported. • The discussion and explanation of any heterogeneity observed in the results of the review was reported. • The potential sources of conflict of interest and funding were disclosed.	• An investigation of publication bias and its impact on the results of the review were not reported.

PROSPERO = the International Prospective Register of Systematic Reviews; RCT = randomized controlled trial.

Table 6: Strengths and Limitations of Clinical Study using Downs and Black Checklist[11]

Strengths	Limitations
Shan, 2019[15]	
• The objective, main outcomes, characteristics, interventions, confounders and main findings of the study were clearly described. • The patient adherence to with the interventions were likely reliable due to observed dosing of medication in hospital. • The patients were randomized to the treatment groups using a random number list. • The patients in different intervention groups were recruited from the same population and over the same time period. • The statistical tests used to assess the main outcomes were appropriate. • The main outcome measures used were reliable as it was depression severity measured as a validated HAMD-17 score. • Probability values were reported as p values for the main outcomes. • Potential conflicts of interest were reported in the article.	• The study was single-blinded, with the blinded group being the clinical psychiatrist. • It was unclear whether the patients who participated, staff, places, and facilities in the study in China were representative of the Canadian population. • The patients who were asked to participate and prepared to participate in the study were recruited via convenience sample (patients presenting to the hospital). It was unclear whether they were representative of the entire population of adult patients with depression. • A power calculation was not conducted to determine if the sample was of an adequate size for clinical significance.

HAMD-17 = 17-item Hamilton depression scale.

Table 7: Strengths and Limitations of Economic Studies using the Drummond Checklist[12]

Strengths	Limitations
Sluitter, 2019[9]	
Study design • The research question, economic importance of the research question, and rationale for choosing alternative interventions compared were clearly stated • The viewpoint/perspective of the analysis were clearly stated. • The treatment strategies being compared were clearly described • The form of economic evaluation used was stated	• The viewpoint and perspective were not justified. • The relevance of productivity changes was not discussed • Details of currency adjustments or conversions were not provided • Evaluation is a Netherlands-based study and may not be generalizable to other health systems • A societal perspective may not be appropriate in the Canadian context

Strengths	Limitations
• The choice of form of economic evaluation was justified in relation to the questions addressed **Data collection** • The sources of effectiveness estimates and screening/treatment costs were provided • The design and results of effectiveness studies from which assumptions were drawn were provided although not clear as some points. • The primary outcome measures for the economic evaluation were clearly stated • Methods to value benefits were stated • Details of the subjects from whom valuations were obtained were given • Methods for the estimation of unit costs were described • The structure of the model was clearly described using figures **Analysis and interpretation of results** • Time horizon of costs and benefits was stated (12 weeks) • Discount rates, choices of discount rates, and explanation of costs/benefits not being discounted were provided • The approach to sensitivity analysis was given • The choice of variables for the sensitivity analysis were justified • Major outcomes were presented in a disaggregated form • The answer to the study question was given • Incremental analysis was reported • Conclusions follow from the data reported • Conclusions were accompanied by appropriate caveats **Miscellaneous** • The authors declared that they had no potential conflicts of interest • Sources of funding were disclosed and were unlikely to have had an effect on the findings of the study	• The time horizon may be considered short although the authors did describe their reasoning for this. • There is some uncertainty with key parameters in the model which may affect the ICER. • One study was the source of model parameters and studied one antidepressant treatment (venlafaxine; assumptions were made for other antidepressants), which may overestimate the effect and results of this study • Multiple assumptions made in the study may affect the results and interpretation of the study.
Groessl, 2018[16]	
Study design • The viewpoint/perspective of the analysis was stated • The research question, economic importance of the research question, and rationale for choosing alternative interventions compared were clearly stated • The treatment strategies being compared were clearly described • The design and results of effectiveness studies from which assumptions were drawn were not provided • The form of economic evaluation used was stated • The choice of form of economic evaluation was justified in relation to the questions addressed **Data collection**	• The author disclosed a potential conflict of interest as being consultant for the manufacturer of IDGx • Sources of funding were disclosed and may have had an effect on the findings of the study • The viewpoint/perspective of the analysis was not justified although the authors noted that there may be some uncertainty with this. • Data input were taken from single references, rather than a synthesis or meta-analysis of estimates from multiple sources • The relevance of productivity changes was not discussed • Details of currency adjustments or conversions were not provided (assumed to be US dollars) • The data and studies used as model parameters were based on different populations and studies

Strengths	Limitations
• The sources of effectiveness estimates and screening/treatment costs were provided • The primary outcome measures for the economic evaluation were clearly stated • Methods to value benefits were stated • Details of the subjects from whom valuations were obtained were given • Methods for the estimation of unit costs were described • The structure of the model was clearly described using figures **Analysis and interpretation of results** • Time horizon of costs and benefits was stated (12 weeks) • Discount rates, choices of discount rates, and explanation of costs/benefits not being discounted were provided • The approach to sensitivity analysis was given • The choice of variables for the sensitivity analysis were justified • Major outcomes were presented in a disaggregated form • The answer to the study question was given • Incremental analysis was reported • Conclusions follow from the data reported • Conclusions were accompanied by appropriate caveat	• Data and studies are based on a 12-week follow-up periods which may inaccurately depict the patient's response to treatment • There may be slights over-estimation of mortality rates for depressions, since suicide rates were derived from a study focusing on treatment-resistant MDD • The findings of this US based evaluation may not be generalizable to other health systems • A societal perspective may not be appropriate in the Canadian context • There was some uncertainty with key parameters in the model which may affect the ICER. • Utility scores may be based on data from samples that were different from the populations that they were applied affecting the overall results of the model. • There was uncertainty regarding the length of the follow-up of patients and treatment.
Najafzadeh 2017[17]	
Study design • The viewpoint/perspective of the analysis was stated while supplemental time horizons were used for another analysis (12 weeks, 1, 2, 10 years and lifetime) • The research question, economic importance of the research question, and rationale for choosing alternative interventions compared were clearly stated • The treatment strategies being compared were clearly described • The design and results of effectiveness studies from which assumptions were drawn were not provided • The form of economic evaluation used was stated • The choice of form of economic evaluation was justified in relation to the questions addressed **Data collection** • The sources of effectiveness estimates and screening/treatment costs were provided • The primary outcome measures for the economic evaluation were clearly stated • Methods to value benefits were stated • Details of the subjects from whom valuations were obtained were given • Methods for the estimation of unit costs were described • The structure of the model was clearly described using figures	• The author disclosed a potential conflict of interest as a consultant for the manufacturer of IDGenetix • The viewpoint/perspective of the analysis was not justified • Data input were taken from single references, rather than a synthesis or meta-analysis of estimates from multiple sources • The relevance of productivity changes was not discussed • Details of currency adjustments or conversions were not provided (assumed to be US dollars) • The findings of this US based evaluation may not be generalizable to other health systems • A societal perspective may not be appropriate in the Canadian context • There was some uncertainty with key parameters in the model which may affect the ICER. • Results of the study did not apply to patients with mild depression with/without anxiety. • The analysis used the same cost estimates for each severity level of depression.

Strengths	Limitations
Analysis and interpretation of results • Time horizon of costs and benefits was stated (12 weeks) • Discount rates, choices of discount rates, and explanation of costs/benefits not being discounted were provided • The approach to sensitivity analysis was given • The choice of variables for the sensitivity analysis were justified • Major outcomes were presented in a disaggregated form • The answer to the study question was given • Incremental analysis was reported • Conclusions follow from the data reported • Conclusions were accompanied by appropriate caveats **Miscellaneous** • Sources of funding were disclosed	

ICER = incremental cost-effectiveness ratio; US = united states.

Appendix 4: Main Study Findings and Authors' Conclusions

Table 8: Summary of Findings Included Systematic Reviews and Meta-Analyses

Main Study Findings	Authors' Conclusion
Washington State, 2016[13]	
Key Question #1b Impact of pharmacogenomics testing on patient outcomes for depressive disorders **Remission:** **Winner, 2013[22]** (guided - GeneSight PGx test panel n = 26 vs unguided n = 25) • At 10 weeks, patients achieved remission (Ham-D17 <7): • 20% of guided patients vs, 8.3% of unguided (OR=2.75; 95% CI, 0.48-15.8; *P* = NS). **Singh, 2015[23]** (guided - CNSDose assay n = 74 vs unguided n = 74) · At 12 weeks, guided patients more often obtained remission (HAM-D17 <7) • OR=2.52; 95% CI, 1.71-3.73; *P* ＜0.0001 • Number needed to test for remission = 3 (95% CI, 1.7-3.5). **Hall-Flavin, 2013[21]** (guided - GeneSight n = 114 vs unguided n = 113) • At 8 weeks, more guided patients obtained remission (QIDS-C16<6) compared with unguided patients (OR=2.42; 95% CI, 1.09-5.39; *P* = 0.03). • HAM-D17 and PHQ-9 results not significantly different except for results using data imputation to account for 27% lost to follow-up. **Breitenstein, 2014[24]** (guided - ABCB1 test n = 58 vs unguided n = 58) • guided patients more often in remission (HAM-D21 <10) at treatment week 4 compared with unguided patients (83.6% vs 62.1%; *P* = 0.005). HAM-D21 at admission >14. • Required change in score may not be clinically relevant. **Response to treatment:** **Winner, 2013[22]** (guided - GeneSight n = 26 vs unguided n = 25, all genotyped) • At 10 weeks, 36% of guided patients responded (>50% reduction in HAM-D17) vs 20.8% of unguided patients (OR=2.14; 95% CI, 0.59-7.69; *P* = NS). **Hall-Flavin, 2013[21]** (guided - GeneSight n = 114 vs unguided n = 113, all genotyped) • At 8 weeks more guided patients responded (>50% reduction in score from baseline) vs unguided patients as measured by: • QIDS-C16 (OR=2.58; 95% CI, 1.33-5.03; *P* = 0.005) • HAM-D17 (OR=2.06; 95% CI, 1.07-3.95; *P* = 0.03) • PHQ-9 (OR=2.27; 95% CI 1.20-4.30; *P* = 0.01) • Results using data imputation to account for 27% loss to follow-up were statistically significant except for QIDS-C16. **Hall-Flavin, 2012[20]** (guided - GeneSight n = 25 vs unguided n = 26; all genotyped) • 8-wk score reductions: • QIDS-C16: 31.2% for guided patients vs 7.2% for controls (*P* = 0.002). • HAM-D17: 30.8% for guided patients vs 18.2% for controls (*P* = 0.04).	**Clinical effectiveness:** **Remission** "In all studies, the direction of results suggests that genotyped pts are more likely to obtain remission. But results are not consistently statistically significant and in 1 study may not be clinically relevant." (p.15)[13] **Response to treatment:** "Results are in the direction of improved response for genotyped patients. Only 1 study used defined measures of response and obtained statistically significant results." (p. 15-16)[13] **Adherence, tolerance, adverse events:** "In 2 of 3 studies, results indicate increased tolerance of medications when prescribed with knowledge of PGx results." (p. 16)[13] **Hospital stay/Healthcare utilization:** "Results indicate PGx for ABCB1 variants may result in better anti-depressant dosing and shorter hospital stays; not generalizable" (p. 17)[13] **Cost effectiveness:** "One study found PGx testing not to be cost-effective; 1 modeling study of a hypothetical pt cohort estimated an increased overall cost of healthcare with PGx vs Ctl for an incremental benefit in QALW." (p.20) [13] **Cost utility:** "Utility increases with decreases in the number of changes in meds or ↓ times for dosage adjustments." (p.20) [13] **Guideline recommendation:** "In possibly nonadherent patients (e.g., low drug plasma levels despite high doses of the antidepressant), a combination of TDM and genotyping may be informative. Such analyses can aid in identifying those individuals who are slow or rapid metabolizers of certain antidepressants." (p.22)[13] "In summary, the evidence base for pharmacogenomic testing for the psychiatric disorders of interest for this report is extremely

Main Study Findings	Authors' Conclusion
Rundell, 2011[25] (guided n = 29 vs unguided n = 17) – Intervention is At least one of CYP2D6, CYP2C19, CYP2C9, and/or serotonin transporter genotype 5-HTTLPR vs. no test ordered • CYP450 categories: No significant differences in serial PHQ-9 scores over time. • 5-HTTLPR categories: L/L genotype patients had greater PHQ-9 score improvement than other genotypes at times 4 and 5 (*P* = 0.02 to *P* = 0.05). • Adjusted post-day 14 PHQ-9 scale slopes and differences in pre- to post-baseline scale slopes were not significantly different among genotype categories. **Adherence, tolerance, adverse events:** **Singh, 2015[23]** (guided - CNSDose n = 74 vs unguided n = 74) • Unguided patients were less able to tolerate medications, requiring dose reduction or cessation (OR=1.13; 95% CI, 1.01-1.25; *P* = 0.0272). • guided patients took sick leave less often (4% vs 15%; *P* = 0.0272) and of less duration when needed (4.3 vs 7.7 days; *P* = 0.014). **Hospital stay/Healthcare utilization:** **Breitenstein, 2014[24]** (guided - ABCB1 test n = 58 vs unguided n = 58) • Dose increases in genotype-appropriate antidepressants were associated with shorter hospital stays (*P* = 0.009). Hospital stay for patients with unfavorable ABCB1 genotype was reduced by 4.7 weeks if dose was increased more than 1.5-fold. **Key Question #4: What are the costs and cost-effectiveness of genetic testing to guide the selection or dose of medications?** **Cost-effectiveness studies:** **Perlis, 2009**[26] HTR2A PGx testing either before first-line tx (Test 1st) or after first-line tx failure (Test 2nd) vs no testing (Ctl): • Direct medical costs including outpatient and inpatient treatment, meds • Test 1st + bupropion tx for test-negative patients ↑ cost by $505/pt but provided 0.0054 QALY for ICER of $93,520/QALY; therefore, not cost-effective. **Olgiati, 2012[27]**, 5-HTTLPR PGx testing vs none in high income Western European countries: • Estimated costs of meds, outpatient and inpatient care, and genetic testing in Western European healthcare systems • Incremental benefit of PGx 0.062 QALWs for response + 0.016 QALWs for side effects • Overall incremental PGx benefit 0.156 QALWs • Estimated overall cost of healthcare Intl $2,242 (PGx) vs Intl $2,063 (Unguided) • Incremental cost of PGx testing was Intl $179 and the ICER was Intl.$1,147 **Cost-utility studies:** **Herbild, 2009[28]** CYP2D6 PGx testing vs none, willingness-to-pay for PGx: • Willingness to pay for a 10% probability of 1 antidepressant change or for the reduction of 1 month of time for dosage adjustments exceeded test cost in Denmark **Relevant practice guidelines** • **World Federation of Societies for Biological Psychiatry guideline, 2013[32]** recommended: "In possibly nonadherent patients (e.g., low	limited and compromised and is considered to be of low to very low quality, depending on the outcome measured. As such, the evidence is insufficient for forming conclusions regarding clinical use." (p.56)[13]

Main Study Findings	Authors' Conclusion
drug plasma levels despite high doses of the antidepressant), a combination of TDM and genotyping may be informative. Such analyses can aid in identifying those individuals who are slow or rapid metabolizers of certain antidepressants." (p.22)[13] ○ Quality: 5.0 – Fair ○ Limitations: Search terms and dates literature covered NR; criteria for selecting evidence and how the body of evidence was evaluated for bias NR	
Bousman, 2019[6]	
Relevant primary studies: **Bradley, 2018[30]** ○ Intervention vs. comparator: pharmacogenetic (NeuroIDgenetix)-guided (n = 352) vs unguided (n = 333) ○ Remission: ▪ Guided group, n/N: 14/40, ▪ Unguided group, n/N: 7/53, ▪ RR, 95%CI: 2.65, (1.18–5.95) **Greden, 2018[31]** ○ Intervention vs. comparator: pharmacogenetic (GeneSight)-guided (n = 681) vs unguided (n = 717): ○ Remission: ▪ Guided group, n/N: 93/607, ▪ Unguided group, n/N: 57/560, ▪ RR, 95%CI: 1.51, (1.11-2.05) **Perez, 2017[29]** ○ Intervention vs. comparator: pharmacogenetic (Neuropharmagen)-guided (n = 155) vs unguided (n = 161) ○ Remission: ▪ Guided group, n/N: 48/141, ▪ Unguided group, n/N: 46/139, ▪ RR, 95%CI: 1.03, (0.74-1.43) **Singh, 2015[23]** – reported in Washington State, 2016[13], analyzed in pooled analysis ○ Intervention vs. comparator: pharmacogenetic (CNSDose)-guided (n = 74) vs unguided (n = 74) ○ Remission: ▪ Guided group, n/N: 52/74, ▪ Unguided group, n/N: 21/74, ▪ RR, 95%CI: 2.52, (1.71-3.73) **Winner, 2013[22]** – reported in Washington State, 2016[13], analyzed in pooled analysis ○ Intervention vs. comparator: pharmacogenetic-guided (n = 25) vs unguided (n = 24) ○ Remission: ▪ Guided group, n/N: 5/25, ▪ Unguided group, n/N: 2/24, ▪ RR, 95%CI: 2.40, (0.51-11.21)	"Our meta-analysis showed pharmacogenetic-guided prescribing has a positive effect on the likelihood of achieving symptom remission. However, inclusion criteria of the included studies suggest this positive effect on remission may be confined to individuals with moderate to severe depression and a history of inadequate response or intolerability to previous psychotropic medications." (p. 43)[6] "Our updated systematic review and meta-analysis suggests pharmacogenetic-guided DST treatment is superior to treatment as usual in relation to remission likelihood, specifically among those with inadequate response or intolerability to previous psychotropic medications and perhaps more noticeably among individuals with more severe depressive symptoms. Thus, the results to date, suggest pharmacogenetic-guided DSTs merit consideration by clinicians treating patients who have not responded or have not been able to tolerate one or more psychotropic medications." (p. 43-44)[6] "We systematically identified and assessed five RCTs that examined the effect of pharmacogenetic-guided antidepressant prescribing on remission in the management of MDD. Our meta-analysis showed pharmacogenetic-guided prescribing has a positive effect on the likelihood of achieving symptom remission. However, inclusion criteria of the included studies suggest this positive effect on remission may be confined to individuals with moderate to severe depression and a history of inadequate response or intolerability to previous psychotropic medications." (p. 43)[6]
Rosenblat, 2017[14]	
Clinical effectiveness: **Hall-Flavin, 2012[20]** – reported in Washington State, 2016[13], data not extracted here **Hall-flavin, 2013[21]** – reported in Washington State, 2016[13], data not extracted here	Conclusion on Hornberger, 2015 "Therefore, their results suggested that combinational pharmacogenomic testing could be a cost-effective intervention. Notably, however, their projections were based mostly on studies of

Main Study Findings	Authors' Conclusion
Winner, 2013[22] – reported in Bousman, 2019[6] and Washington State, 2016[13], data not extracted here **Singh, 2015**[23] – reported in Bousman, 2019[6] and Washington State, 2016[13], data not extracted here **Cost-effectiveness:** **Hornberger, 2015**[33] Use model from Perlis 2009 for cost-effectiveness analysis Study data from Hall-Flavin 2012, Hall-flavin 2013, and Winner 2013 **Estimated change in QALYs:** increase by 0.316 years for PGx guided therapy **Projected savings:** • Saving in direct medical costs: $3,711 • Saving in work productivity costs per patient over the lifetime: $2,553 **Probability of GeneSight testing being cost-effective at the WTP threshold of $50,000:** 94.5%. **Winner, 2015**[35] – reported in Washington State, 2016[13], data not extracted here **Winner, 2013**[22] – reported in Bousman, 2019[6] and Washington State, 2016[13], data not extracted here **Olgiati, 2012**[27] – reported in Washington State, 2016[13], data not extracted here **Perlis, 2009**[26] – reported in Washington State, 2016[13], data not extracted here	poor quality; lacking appropriate randomization and blinding; to determine efficacy of GeneSight testing, 93.3% of the pooled results was based on 2 open-label, nonrandomized studies, while only 6.7% of their pooled results was based on a randomized, controlled, and double-blinded study. Therefore, the pooled efficacy (pooled effect of testing on response rate calculated to be 1.71 [95% CI 1.17 2.49]) was based mostly on low-quality studies. Since the model of cost-effectiveness is heavily weighted on intervention efficacy effect size, the validity of the results of this analysis is questionable as the reliability of the calculated efficacy may be poor."(p.727) [14]s "In conclusion, currently available evidence for improved clinical outcomes from pharmacogenomic testing is limited. Clinical trials suggestive of a positive effect of pharmacogenomic testing on clinical outcomes in MDD were mostly of low quality, often lacking randomization and blinding, and were vulnerable to bias from industry funding. Further, results from a randomized, double-blind clinical trial of GeneSight did not reach statistical significance; however, notably, they may have been underpowered. One randomized, double-blind clinical trial of CNSDose found a statistically significant increase in remission rates, but this has yet to be independently replicated.29 Taken together, results from these studies suggest that. further studies are required and merited to determine The impact of these tests on clinical outcomes, namely in the rate (ie, time to improvement) and amount (ie, response and remission rates) of therapeutic improvement. Well-designed clinical trials with adequate sample sizes) randomization, and blinding are required prior to the routine implementation of pharmacogenomic testing into clinical practice. If testing is found to improve clinical outcomes, the cost-effectiveness of testing should also be further evaluated based on the results from high-quality studies." (p.728) [14]

CI = confidence interval; Ctl = control NR = not reported; DST = decision support tool; HAM-D = Hamilton Depression Rating Scale; PGx = pharmacogenetic/pharmacogenomics testing; OR = odds ratio; pts = patients; PHQ-9 = Patient Health Questionnaire; QALW = quality-adjusted life week; QALY = quality-adjusted life year; QIDS-C16 = Quick Inventory of Depressive Symptomatology-Clinician Rated; RR = relative risk; TDM = therapeutic drug monitoring; WTP = willingness to pay.

Table 9: Summary of Findings of Included Primary Clinical Study

Main Study Findings	Authors' Conclusion
Shan, 2019[15]	

HAMD-17 total scores:
• Significantly decreased from baseline to 8 weeks within the guided and unguided groups (P < 0.01)
• No significant difference was found in the HAMD-17 total scores at each time point between the unguided and guided groups
  ○ At 2 weeks, *P* = 0.696
  ○ At 4 weeks, *P* = 0.901
  ○ At 8 weeks, *P* = 0.205
• The reduction ratio of HAMD-17 scores at 8 weeks**:** 60.86% in the guided group vs**.** 52.38% in the unguided group with no significant difference (*P* = 0.210)
  ○ Reduction ratio = (HAMD-17$_{total_1}$-HAMD-17$_{total_2}$)/HAMD-17$_{total_1}$.
  ○ HAMD-17$_{total_1}$ refers to the HAMD-17 total scores at baseline
  ○ HAMD-17$_{total_2}$ is the HAMD-17 total scores after 8 weeks of treatment
 	
Response rate after 8 weeks of treatment:
• Guided group 74.19% (23/31) vs unguided group 57.5% (23/40), not statistically significantly different (*P* = 0.144)
 	
Remission rate:
Guided group 61.29% (19/31) vs. unguided groups 45.0% (18/40) (*P* = 0.173)
Not statistically significantly different
 	
Anxiety symptoms measured by HAMA total scores:
No significant difference observed at each time point between two groups
  ○ At 2 weeks, *P* = 0.985
  ○ At 4 weeks, *P* = 0.889
  ○ At 8 weeks, *P* = 0.961
 	
Routine blood test, liver function, renal function, and electrocardiogram examinations:
After 8 weeks of treatment, no abnormalities were found
 	
Adverse reactions and medication tolerability measured by TESS:
Incidence rate of adverse reactions:
Guided group 55.56% vs. Unguided group 57.89%
No statistical difference between the two groups.
 	
Frequency of adverse reactions:
Guided groups 14 cases vs. Unguided groups 19
Cases, respectively.
 	
Main adverse reactions:
Headache, dizziness, drowsiness, nausea, vomiting, dry mouth, constipation, diarrhea, decreased appetite, and tachycardia
• Frequency NR
	"This study reports no significant difference in the improvement of depressive symptoms between guided and unguided groups at the end of the treatment. Pharmacogenomic testing might not significantly improve the clinical efficiency and safety for the guided group compared with those for the unguided group."(p. 9-10)[15]

DSM-5 = Structural Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, fifth edition; HAMA = Hamilton anxiety scale; HAMD-17 = 17-item Hamilton depression scale; RCT = randomized controlled trial; TESS = treatment emergent symptom scale.

Table 10: Summary of Findings of Included Economic Evaluations

Main Study Findings	Authors' Conclusion
Sluiter, 2019 [9]	
Costs: PGx = €970 (CI: €799 - €1244) No PGx = €881 (CI: €723 - €1130) **Incremental cost**: €89 (CI: €39 - €152) **QALYs:** PGx = 0.146 (CI: 0.133 – 0.159) No PGx = 0.145 (CI :0.132 – 0.157) **Incremental QALY :** 0.001 (CI : 0.001 – 0.002) **ICER :** €77,406 per QALY 48% of the simulations were below the WTP threshold of €80,000 per QALY **Scenario Analysis:** - Most sensitive to a delayed effect of a week for PGx; not cost-effective - Most sensitive to productivity losses were not considered, the ICER is not cost-effective	"*According to our model, we cannot unequivocally conclude that screening for CYP2D6 in primary care patients using antidepressants is be cost-effective, as the results are surrounded by large uncertainty. Therefore, information from ongoing studies should be used to reduce these uncertainties*." (pg.1) "*Although patients' genetic properties could be taken into account, other diagnostics, such as therapeutic drug monitoring, psychosocial factors, etc., remain necessary to optimise treatment as 'PGx' still misses a part of the patients with a poor metabolizer status or that are still not responding to treatment."* (pg.9)
Groessl, 2018[16]	
Total Costs: Moderate to Severe Depression SOC = $47,295 IDGx = $44, 697 Severe Depression SOC = $47,025 IDGx = $41,215 **Incremental cost**: Moderate to Severe Depression $-2,598 Severe Depression-$5,810 **QALYs:** Moderate to Severe DepressionSOC = 1.97 IDGx = 2.07 Severe DepressionSOC = 0.356 IDGx = 0.311 **Incremental QALY:** Moderate to Severe Depression= 0.01	*"The IDGx-guided treatment producing both QALYs gained and cost savings, the treatment "dominates" the SOC treatment." (pg.5)* *"Given the increased need for a variety of health care providers to prescribe and manage antidepressants, pharmacogenetic tests are a valuable tool that demonstrate improved patient outcomes in real-world settings and are strongly positioned to help reduce the economic burden of depression." (pg.7)*

Main Study Findings	Authors' Conclusion
Severe Depression= 0.17 **ICER:** IDGx dominates over SOC **Sensitivity Analysis:** - Model most sensitive to annual health care costs for responders - Model most sensitive to annual health care costs for non-responders	
Najafzadeh 2017 [17]	
Total Costs: TAU = $14,659 (CI : S10,384 – 19,275) IDgenetix = $14,124 (CI: $10,703 – 17,630) **Incremental cost**: -$535 **QALYs:** IDgenetix = 2.09 (CI: 1.88 – 2.88) TAU = 1.94 (CI: 1.66 – 2.21) **Incremental QALY:** 0.15 (0.04 – 0.28) **ICER:** IDGenetix dominates over TAU The probability of IDGenetix being cost-effective over TAU is 90% **Sub-group Analysis:** Patients with/without anxiety = $35/QALY Patients with severe depression and/or severe anxiety = IDGenetix dominates of TAU **Sensitivity Analysis:** • Model most sensitive most to assumption about remission and response rates of alternative treatment strategies	*"In summary, we found that implementing IDgenetix guided treatment of patients with moderate-to-severe depression and/or anxiety is likely to result in cost savings and improved QOL, compared with TAU….Overall, several features of the IDgenetix test, such as the efficacy of a guided treatment strategy, one-time test cost, and the prospect of using test results for guiding future episodes of depression or anxiety, make it a potentially dominant strategy compared with usual care." (pg.12)*

CI = confidence interval; ICER = incremental cost-effectiveness ratio; PGx = pharmacogenetic testing; QALY = quality-adjusted life year; TAU = treatment as usual; WTP = willingness to pay threshold.

Appendix 5: Overlap between Included Systematic Reviews

Table 11: Primary Study Overlap between Included Systematic Reviews

Primary Study Citation	Systematic Review Citation		
	Bousman, 2019[6]	Rosenblat, 2017[14]	Washington State, 2016[13]
Bradley, 2018[30]	X		
Greden, 2018[31]	X		
Perez, 2017[29]	X		
Trangle, 2016[36]			X
Espadaler, 2016[37]			X
VA/DoD, 2016[38]			X
Brennan, 2015[39]		X	
Singh, 2015[23]	X	X	X
Hornberger, 2015[33]		X	
Oslin, 2015[40]			X
Winner, 2015[35]		X	X
Fagerness, 2014[41]			X
Bauer, 2013[32]			X
Breitenstein, 2014[24]			X
Hall-Flavin, 2013[21]		X	X
Herbild, 2013[42]			X
Winner, 2013[22]	X	X	X
Hall-Flavin, 2012[20]		X	X
Olgiati, 2012[27]		X	X
Möller, 2011[43]			X
Rundell, 2011[25]			X
Beyondblue, 2010[44]			X
Perlis, 2009[26]		X	X
Herbild, 2009[28]			X

Appendix 6: Additional References of Potential Interest

Health Technology Assessment with Complete Overlap with Included Health Technology Assessment

Health Quality Ontario. Pharmacogenomic testing for psychotropic medication selection: a systematic review of the Assurex GeneSight psychotropic test. *Ont Health Technol Assess Ser.* 2017;17(4): **https://www.hqontario.ca/Portals/0/documents/evidence/reports/hta-genesight-13-03-2017-en.pdf**. Accessed 2020 Jan 31.