Common Drug Review
Clinical Review Report

November 2016

Drug	aripiprazole (Abilify)
Indication	As an adjunct to antidepressants for the treatment of major depressive disorder (MDD) in adult patients who had an inadequate response to prior antidepressant treatments during the current episode
Listing request	As per indication
Manufacturer	Bristol-Myers Squibb Canada[a]

[a] Trademark of Otsuka Pharmaceutical Co., Ltd. used under license by Bristol-Myers Squibb Canada.

This report was prepared by the Canadian Agency for Drugs and Technologies in Health (CADTH). Through the CADTH Common Drug Review (CDR) process, CADTH undertakes reviews of drug submissions, resubmissions, and requests for advice, and provides formulary listing recommendations to all Canadian publicly funded federal, provincial, and territorial drug plans, with the exception of Quebec.

The report contains an evidence-based clinical and/or pharmacoeconomic drug review, based on published and unpublished material, including manufacturer submissions; studies identified through independent, systematic literature searches; and patient-group submissions. In accordance with *CDR Update — Issue 87*, manufacturers may request that confidential information be redacted from the CDR Clinical and Pharmacoeconomic Review Reports.

The information in this report is intended to help Canadian health care decision-makers, health care professionals, health systems leaders, and policy-makers make well-informed decisions and thereby improve the quality of health care services. The information in this report should not be used as a substitute for the application of clinical judgment with respect to the care of a particular patient or other professional judgment in any decision-making process, nor is it intended to replace professional medical advice. While CADTH has taken care in the preparation of this document to ensure that its contents are accurate, complete, and up-to-date as of the date of publication, CADTH does not make any guarantee to that effect. CADTH is not responsible for the quality, currency, propriety, accuracy, or reasonableness of any statements, information, or conclusions contained in the source documentation. CADTH is not responsible for any errors or omissions or injury, loss, or damage arising from or relating to the use (or misuse) of any information, statements, or conclusions contained in or implied by the information in this document or in any of the source documentation.

This document is intended for use in the context of the Canadian health care system. Other health care systems are different; the issues and information related to the subject matter of this document may be different in other jurisdictions and, if used outside of Canada, it is at the user's risk. This disclaimer and any questions or matters of any nature arising from or relating to the content or use (or misuse) of this document will be governed by and interpreted in accordance with the laws of the Province of Ontario and the laws of Canada applicable therein, and all proceedings shall be subject to the exclusive jurisdiction of the courts of the Province of Ontario, Canada.

CADTH takes sole responsibility for the final form and content of this document, subject to the limitations noted above. The statements and conclusions in this document are those of CADTH and not of its advisory committees and reviewers. The statements, conclusions, and views expressed herein do not necessarily represent the views of Health Canada or any Canadian provincial or territorial government. Production of this document is made possible by financial contributions from Health Canada and the governments of Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland and Labrador, Northwest Territories, Nova Scotia, Nunavut, Ontario, Prince Edward Island, Saskatchewan, and Yukon.

You are permitted to make copies of this document for non-commercial purposes, provided it is not modified when reproduced and appropriate credit is given to CADTH. You may not otherwise copy, modify, translate, post on a website, store electronically, republish, or redistribute any material from this document in any form or by any means without the prior written permission of CADTH.

Please contact CADTH's Vice-President of Corporate Services at corporateservices@cadth.ca with any inquiries about this notice or other legal matters relating to CADTH's services.

TABLE OF CONTENTS

Tables

Figures

ABBREVIATIONS

5-HT	5-hydroxytryptamine (serotonin)
AAP	atypical antipsychotic
ADT	antidepressant therapy
AE	adverse event
ANCOVA	analysis of covariance
CDR	CADTH Common Drug Review
CGI-I	Clinical Global Impression–Improvement
CGI-S	Clinical Global Impression–Severity of Illness
CR	controlled release
DSM-IV-TR	*Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition, Text Revision
EPS	extrapyramidal symptoms
HAM-D	Hamilton Depression Rating Scale
IDS-SR	Inventory of Depressive Symptomatology–Self-Report
ITT	intention-to-treat
LOCF	last observation carried forward
MADRS	Montgomery–Åsberg Depression Rating Scale
MCID	minimal clinically important difference
MDD	major depressive disorder
MDE	major depressive episode
MDSC	Mood Disorders Society of Canada
NICE	National Institute for Health and Care Excellence
NMA	network meta-analysis
QALY	quality-adjusted life-year
QIDS-SR	Quick Inventory of Depressive Symptomatology–Self-Report
Q-LES-Q	Quality of Life Enjoyment and Satisfaction Questionnaire
RCT	randomized controlled trial
SAE	serious adverse event
SDS	Sheehan Disability Scale
SNRI	serotonin and norepinephrine reuptake inhibitor
SSRI	selective serotonin reuptake inhibitor
TEAE	treatment-emergent adverse event
XR	extended release

EXECUTIVE SUMMARY

Introduction

Major depressive disorder (MDD) is characterized by the occurrence of one or more major depressive episodes (MDEs), which persist for at least two weeks and are characterized by a depressed mood and/or markedly diminished interest or pleasure in all, or almost all, activities.[1] The clinical manifestation of MDD is heterogeneous. The duration of MDEs can also vary significantly in duration, ranging from weeks to years.[1] MDD is one of the most prevalent chronic conditions in Canada, with an annual prevalence reaching 4.8%[2] and a lifetime prevalence of 10.8% of the population.[1,3,4] The prevalence of MDD is twice as high for women as for men, but this difference declines with age.[5] According to the Global Burden of Disease Study and other studies, MDD is a major cause of disability.[6-9] Because of its early age of onset and frequent recurrences, MDD is also among the leading causes of disability, as measured by disability-adjusted life-years, worldwide.[6]

The goal of treatment in patients with MDD is the resolution of symptoms (remission) in order to restore psychosocial and occupational functioning. Traditional antidepressant therapy (ADT) is the mainstay of treatment. Despite the availability of various ADTs, as many as 50% to 60% of patients do not respond to ADT.[10]

Aripiprazole (Abilify) is an atypical antipsychotic (AAP) and is the only drug approved by Health Canada for the adjunctive treatment of MDD in adults who had an inadequate response to prior ADT. The objective of this review is to evaluate the beneficial and harmful effects of aripiprazole (oral tablets 2 mg, 5 mg, 10 mg, and 15 mg) as an adjunct to ADT for the treatment of MDD in adult patients who had an inadequate response to prior ADT during the current episode.

Results and Interpretation

Included Studies

Three double-blind randomized controlled trials were included in the review. (In this review, study CN138-139, study CN138-163, and study CN165 are simplified as study 139,[11] study 163,[12] and study 165,[13] respectively.) All three trials were identically designed and consisted of three phases: screening phase (phase A), run-in phase (phase B), and double-blind randomized phase (phase C). The objectives of phase A were to select patients with a diagnosis of MDD, as defined by *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition, Text Revision (DSM-IV-TR) criteria, and to ensure discontinuation of previous psychotropic drugs. The objective of phase B was to identify patients with MDD with an inadequate response to ADT; patients in this phase were assigned to open-label ADT plus placebo, and their response was measured at eight weeks. Patients who were non-responders were eligible to enter phase C. The objective of phase C was to establish the comparative efficacy and safety of aripiprazole over placebo as an adjunct to ADT in patients with MDD with an inadequate response to ADT. Patients (N = 1,092 in total) who had an incomplete response to prior ADT were randomized to aripiprazole plus ADT or placebo plus ADT; patients remained on the ADT regimen assigned in phase B. All three studies were six weeks in duration. The primary outcome of the included studies was the between-treatment group difference of change during phase C in Montgomery–Åsberg Depression Rating Scale (MADRS) total score. Other outcomes included the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) subscale scores, Sheehan Disability Scale (SDS), Hamilton Depression Rating Scale, 17-item version (HAM-D17) total score, Clinical Global Impression–Severity of Illness (CGI-S), Inventory of Depressive Symptomatology–Self-Report (IDS-SR), and Quick Inventory of Depressive Symptomatology–Self-Report (QIDS-SR), and the proportion of patients achieving MADRS response (defined as a ≥ 50% reduction in MADRS total score) and remission (defined as MADRS total

score of ≤ 10 and reduction of ≥ 50% from the end of phase B). Safety outcomes included mortality, serious adverse events (SAEs), and adverse events (AEs).

The main limitations of the body of evidence are the following: an active combination comparator (such as other AAPs) is lacking; physician-rated MADRS improvement with aripiprazole did not translate into similar improvements in the patient-rated IDS-SR or QIDS-SR; and data demonstrating superiority of aripiprazole over placebo on validated quality of life instruments or on function/disability outcomes are lacking.

Efficacy

A statistically significantly greater improvement with aripiprazole than with placebo during phase C in the Q-LES-Q overall life satisfaction item was demonstrated in all three studies: mean (95% confidence interval) between-group differences in change during phase C were 0.25 points (0.07 to 0.44), 0.19 points (0.01 to 0.36), and 0.33 points (0.13 to 0.52) in study 139, study 163, and study 165, respectively. A statistically significantly greater improvement with aripiprazole than placebo during phase C in the Q-LES-Q overall general subscore was observed in two of the three studies (study 163 and study 165). No statistically significant differences were observed between the groups in Q-LES-Q satisfaction with medication in any of the studies. A minimal clinically important difference (MCID) has not been defined for this questionnaire; hence, it remains unclear whether the observed statistically significant benefits with aripiprazole are clinically meaningful.

A statistically significant difference on the overall SDS mean score change between treatment groups, in favour of aripiprazole, was demonstrated in only one of the studies (study 163). However, a statistically significant difference between the treatment groups in favour of aripiprazole was demonstrated on the family life item score in all three studies. As well, a statistically significant improvement with aripiprazole compared with placebo on the social life item was observed in study 139 and study 163. None of the studies showed that aripiprazole was superior to placebo in the work/school item. Again, the clinical relevance of these findings is uncertain.

Compared with placebo, the aripiprazole group showed statistically significantly higher MADRS response (aripiprazole versus placebo: 34% versus 24%, 32% versus 17%, and 47% versus 27% in studies 139, 163, and 165, respectively) and remission rates (aripiprazole versus placebo: 26% versus 16%, 25% versus 15%, and 37% versus 19% in studies 139, 163, and 165, respectively). The numbers needed to treat ranged from 5 to 11 for response and from 6 to 10 for remission. The findings were consistent with those reported in a previous meta-analysis.[14] The manufacturer-conducted network meta-analysis[15,16] [redacted].

Aripiprazole, as compared with placebo, demonstrated a statistically significantly greater improvement in symptom scores, as measured by the MADRS: mean difference in MADRS total score for aripiprazole versus placebo was –3.01 (–4.66 to –1.37), –2.84 (–4.53 to –1.15), and –3.73 (–5.44 to –2.02) in studies 139, 163, and 165, respectively. The magnitude of the symptom improvement measured by the MADRS across three studies was clinically meaningful. The superiority of aripiprazole over placebo was also demonstrated on the HAM-D17 total score, in which the mean difference versus placebo ranged from [redacted]. The between-group differences fall within the 2-point clinically significant difference suggested by Montgomery and Möller[17] and the 3-point difference suggested by the National Institute for Health and Care Excellence.[18] This range of clinical significance

appears to be opinion-based, and no MCID for the HAM-D17 has been derived. Hence, the clinical significance of the findings using HAM-D17 observed in the included studies remains unclear.

No statistically significant treatment-group difference was found at the end of phase C on the IDS-SR total score or QIDS-SR score.

Harms

There were no deaths reported in the three studies. Overall, the incidence of SAEs was infrequent and comparable in the two treatment groups. The SAE rate was 0% to 1.7% for placebo and 0.5% to 1.1% for aripiprazole. The most frequent SAEs included infections, injuries, exostosis, cellulitis, arterial occlusive disease, and suicidal ideation. Overall, AEs were more common with aripiprazole than with placebo across the three studies (80.7% to 82% versus 62.5% to 63.2%, respectively). The most common AEs included akathisia, restlessness, fatigue, insomnia, blurred vision, somnolence, and constipation. More patients in the aripiprazole group (25.6% to 33.9%) reported extrapyramidal symptoms than the placebo group (7.6% to 9.7%). Weight gain was infrequent and comparable in both treatment groups. Sexual dysfunction was reported in only one patient who received aripiprazole. More patients in the aripiprazole group discontinued the study treatment due to AEs than in the placebo group (5.9% versus 1.7%).

Pharmacoeconomic Summary

The manufacturer submitted a cost-utility analysis based on a patient-level simulation model. The primary analysis compares aripiprazole with quetiapine (Seroquel extended release [XR]); while a secondary analysis compares aripiprazole with quetiapine, risperidone, and olanzapine. The manufacturer considers a patient lifetime time horizon with four possible health states: MDE in the acute phase, remission, symptom-free, and death. The analysis is conducted from the perspective of the health care payer. Costs were applied per week, according to the health state. Utility values were sourced from published literature. Clinical efficacy was based on a manufacturer-conducted indirect comparison. The submitted price varies depending on dose: 2 mg = $3.0013, 5 mg = $3.3783, 10 mg, 15 mg, 20 mg, and 30 mg = $3.8933. The submitted prices are in some cases substantially less than the list prices of several public drug formularies.

Results of Manufacturer's Analysis

In the base case, the manufacturer reported that aripiprazole was associated with an additional 0.020 quality-adjusted life-years (QALYs) per patient and an addition cost of $97 compared with quetiapine, leading to an incremental cost per QALY gained of $4,829. The manufacturer's secondary analysis reported that risperidone dominated, as it was both cheaper and more effective than aripiprazole, quetiapine, and olanzapine.

Interpretations and Key Limitations

CADTH Common Drug Review (CDR) identified several limitations of the model:

- **Uncertain comparative efficacy:** The manufacturer submitted an indirect treatment comparison to evaluate the relative efficacy of quetiapine and aripiprazole in terms of remission. The comparison contained several limitations that hindered the interpretability of results.
- **Assumptions regarding quetiapine use may not be appropriate:** The manufacturer assumed that quetiapine would be used at 300 mg per day; however, quetiapine can be used from 150 mg to 300 mg per day for the treatment of MDE. Furthermore, the model assumed that quetiapine would be used as adjunctive therapy only; however, quetiapine could be used as monotherapy or as an adjunct, with higher-level evidence supporting its use as monotherapy.

- **Model time horizon:** A time horizon of 999 years was used in the base case, which is extremely high and underestimates the incremental cost-utility ratio of aripiprazole compared with quetiapine.

Results of CADTH Common Drug Review Analysis

Given the issues identified with manufacturer's model, CDR conducted a reanalysis using the following assumptions:

- A lifetime horizon of 30 to 55 years, based on life expectancy of 75 to 100 years.
- Equal distribution of the three quetiapine doses for 150 mg, 200 mg, and 300 mg per day.
- Half of patients treated with quetiapine as monotherapy, instead of as an adjunct to another ADT.

The reanalysis showed that aripiprazole was associated with an additional 0.02 QALYs per patient and an addition cost of $165 compared with quetiapine, leading to an incremental cost per QALY gained of $8,231.

Based on the results of the manufacturer's indirect treatment comparison, the efficacy of aripiprazole and quetiapine appear similar, which would render aripiprazole more costly than quetiapine. There is, however, considerable uncertainty around the magnitude and direction of the numerical differences. CDR identified several limitations in the manufacturer's analysis, which, when adjusted for in the model, resulted in a higher incremental cost per QALY gained of $8,231 for aripiprazole versus quetiapine.

Conclusions

In the three included double-blind, randomized, placebo-controlled trials, remission and response rates according to MADRS, and changes in MADRS and HAM-D17 scores demonstrated statistically significant greater improvements with adjunctive aripiprazole compared with placebo. A statistically significant improvement in health-related quality of life (by Q-LES-Q overall life satisfaction score and CGI-S) was observed in all three studies. As well, improvement in function/disability (by SDS) was statistically significant in some of the subscores in favour of aripiprazole in some studies. However, the clinical significance of the between-group differences in health-related quality of life and functional capacity is uncertain because of the lack of an MCID for these. In terms of patient-reported symptoms, measured with IDS-SR or QIDS-SR, no statistically significant benefit was reported with aripiprazole compared with placebo. SAEs were infrequent and are too few to draw conclusions. Common AEs associated with aripiprazole were extrapyramidal symptoms, such as akathisia. Other treatment-emergent AEs such as upper respiratory infection, insomnia, and blurred vision also occurred more frequently with aripiprazole than placebo. Akathisia, somnolence, [redacted], and insomnia were the common reasons for discontinuing treatment with aripiprazole. In summary, compared with placebo, adding aripiprazole to ADT statistically significantly improved symptoms on investigator-rated scales and increased response and remission rates in patients with MDD with inadequate response to ADT. However, the findings from the three studies do not clearly indicate whether aripiprazole is superior to placebo in improving health-related quality of life or patients' functioning. A manufacturer–submitted NMA suggested [redacted] The heterogeneity between the studies and patients included in the NMA are key limitations of the analysis.

The main limitations of the body of evidence for aripiprazole as adjunctive therapy to ADT in MDD are the following: an active comparator is lacking; clinician-rated MADRS did not translate into similar improvements with aripiprazole with the patient-rated IDS-SR or QIDS-SR; and data demonstrating clear superiority over placebo for health-related quality of life or for function/disability outcomes are lacking.

TABLE 1: SUMMARY OF RESULTS

Outcome	Study 139		Study 163		Study 165	
	Placebo	Aripiprazole	Placebo	Aripiprazole	Placebo	Aripiprazole
Patients with response (MADRS)						
n/N (%)	41/172 (23.8)	61/181 (33.7)	32 /184 (17.4)	60/185 (32.4)	45/169 (26.6)	81/174 (46.6)
RR (95% CI)	[redacted] *P* = 0.027		[redacted] *P* < 0.001		[redacted] *P* < 0.001	
NNT	11		7		5	
Patients in remission (MADRS)						
n/N (%)	27/172 (15.7)	47/181 (26.0)	28/184 (15.2)	47/185 (25.4)	32/169 (18.9)	64 /174 (36.8)
RR (95% CI)	[redacted] *P* = 0.011		[redacted] *P* = 0.016		[redacted] *P* < 0.001	
NNT	10		10		6	
MADRS total score						
Change during phase C (mean ± SE)	–5.77 ± 0.67	–8.78 ± 0.63	–5.65 ± 0.64	–8.49 ± 0.66	–6.39 ± 0.74	–10.12 ± 0.74
Between-group difference of changes during phase C: mean (95% CI) (aripiprazole – placebo)[b]	–3.01 (–4.66 to –1.37) *P* < 0.001		–2.84 (–4.53 to –1.15) *P* = 0.001		–3.73 (–5.44 to –2.02) *P* < 0.001	
Withdrawals						
Total, n (%)	18 (10.1)	22 (12.1)	28 (14.7)	29 (15.2)	23 (13.4)	30 (16.9)
SAEs						
n/N (%)	3/176 (1.7)	2/182 (1.1)	0/190	1/189 (0.5)	1/172 (0.6)	1/176 (0.6)
RR	[redacted] *P* = 0.63		Not estimable		Not estimable	
NNH	100		Not estimable		Not estimable	
WDAEs						
n/N (%)	4/178 (2.2)	6/182 (3.3)	2/190 (1.1)	7/189 (3.7)	3/172 (1.7)	11/176 (6.2)
RR (95% CI)	[redacted]		[redacted]		[redacted]	
NNH	100		33		20	
Notable harms						
EPS-related AEs						
n/N (%)	17/178 (9.7)	50/182 (27.5)	[redacted]		[redacted]	
RR (95% CI)	[redacted]		[redacted]		[redacted]	
NNH	5		4		5	

AE = adverse event; CI = confidence interval; EPS = extrapyramidal symptoms; MADRS = Montgomery–Åsberg Depression Rating Scale; NNH = number needed to harm; NNT = number needed to treat; RR = relative risk; SAE = serious adverse event; WDAE = withdrawal due to adverse event.

1. INTRODUCTION

1.1 Disease Prevalence and Incidence

Major depressive disorder (MDD) is characterized by the occurrence of one or more major depressive episodes (MDEs), which persist for at least two weeks and are characterized by a depressed mood (most of the day, nearly every day) and/or markedly diminished interest or pleasure in all, or almost all, activities (most of the day, nearly every day).[1] During the same two-week period, the presence of five or more additional symptoms associated with depressive disorder is an important criterion for the diagnosis of MDD. The clinical manifestation of MDD is heterogeneous, and may include dysphoria (any or a combination of feeling sad, helpless, hopeless, irritable or angry, and agitated or anxious), anhedonia (displeasure in previously enjoyed activities), a sense of worthlessness or guilt, inability to concentrate, loss of appetite, insomnia or sleep disturbances, suicidal thoughts or ideation, as well as somatic (physical) symptoms. The duration of MDEs can vary significantly in length, ranging from weeks to even years.[1]

MDD is one of the most prevalent chronic conditions in Canada, with an annual prevalence reaching 4.8%[2] and a lifetime prevalence of 10.8% of the population.[1,3,4] The prevalence of MDD is twice as high for women as for men, but this difference declines with age.[5] According to the Global Burden of Disease Study and other studies, MDD is a major cause of disability.[6-9] Because of its early age of onset and frequent recurrences, MDD is also among the leading causes of disability, as measured by disability-adjusted life-years, worldwide. Both impaired function in the workplace (presenteeism) and high levels of absenteeism caused by MDD have been shown to contribute to economic loss.[19]

1.2 Standards of Therapy

The goal of treatment in patients with MDD is the resolution of symptoms (remission) in order to restore psychosocial and occupational functioning.[20] Combined antidepressant therapy (ADT) and cognitive behavioural therapy or interpersonal psychotherapy are recommended as first-line treatments for acute MDD.[20] In terms of pharmacotherapy, traditional ADT is the mainstay of treatment, with selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, norepinephrine-dopamine reuptake inhibitors (e.g., bupropion), norepinephrine reuptake inhibitors (i.e., reboxetine), serotonin reuptake enhancers (i.e., tianeptine), alpha 2-adrenergic agonists or 5-hydroxytryptamine-2 (5-HT_2) antagonists (e.g., mirtazapine, mianserin), melatonin receptor 1 and 2 agonists (e.g., agomelatine), and reversible inhibitors of monoamine oxidase-A (e.g., moclobemide) regarded as “first-line” options.[21] Despite the availability of various ADTs with established efficacy, as many as 50% to 60% of patients do not respond to ADT,[10] and up to 30% fail to benefit from a series of treatments. In the Sequenced Treatment Alternatives to Relieve Depression study, approximately two-thirds of patients with MDD had insufficient relief of depressive symptoms with an antidepressant. This type of depression is often referred to as treatment-resistant depression or refractory depression.[20]

1.3 Drug

Aripiprazole (Abilify) is the only pharmacotherapy approved by Health Canada for the adjunctive treatment of MDD in adults who had an inadequate response to prior ADT.[22] Aripiprazole is an atypical antipsychotic (AAP) that acts as a partial agonist at dopamine D_2 and serotonin 5-HT_{1A} receptors and as an antagonist at 5-HT_{2A} receptors. Aripiprazole tablets are available in 2 mg, 5 mg, 10 mg, 15 mg, 20 mg, and 30 mg strengths; however, the recommended dose as an adjunct to ADT for treating MDD ranges from 2 mg to 15 mg administered once a day.

Indication under review
As an adjunct to antidepressants for the treatment of major depressive disorder (MDD) in adult patients who had an inadequate response to prior antidepressant treatments during the current episode
Listing criteria requested by sponsor
As per indication

TABLE 2: KEY CHARACTERISTICS OF ARIPIPRAZOLE AND OTHER ATYPICAL ANTIPSYCHOTIC DRUGS USED FOR MAJOR DEPRESSIVE DISORDER

	Aripiprazole	Other AAPs (as a drug class)
Mechanism of Action	Effects may be mediated through a dose-dependent combination of partial agonist activity at D_2 and 5-HT_{1A} receptors and antagonist activity at 5-HT_{2A} receptors	5-HT_{2A} receptor antagonists
Health Canada Indication	As an adjunct to ADT for the treatment of MDD in adult patients who had an inadequate response to prior ADTs during the current episode	Not indicated as an adjunct to ADT for the treatment of MDD; widely used off-label as an adjunct to ADT for MDD
Route of Administration	Oral	Oral
Recommended Dose	2 mg/d to 15 mg/d	Examples: Quetiapine: 150 mg/d or 300 mg/d; Olanzapine: 6 mg/d to 12 mg/d; Risperidone: 1 mg/d to 3 mg/d
Serious Adverse Effects / Safety Issues	SAEs were rare. Examples are: EPS-related AEs; Increased mortality in elderly patients	SAEs were rare. Similar to aripiprazole

5-HT = 5-hydroxytryptamine (serotonin); ADT = antidepressant therapy; AE = adverse event; D = dopamine; EPS = extrapyramidal symptoms; MDD = major depressive disorder; SAE = serious adverse event.
Source: Spielmans et al.;[14] cost-effectiveness analysis report in submission.[15]

2. OBJECTIVES AND METHODS

2.1 Objectives

To evaluate the beneficial and harmful effects of aripiprazole (Abilify) (oral tablets available in 2 mg, 5 mg, 10 mg, and 15 mg strengths) as an adjunct to antidepressants for the treatment of MDD in adult patients who had an inadequate response to prior antidepressant treatments during the current episode.

2.2 Methods

Studies selected for inclusion in the systematic review included the pivotal studies supporting the Health Canada indication provided in the manufacturer's submission to CADTH Common Drug Review, as well as those meeting the selection criteria presented in Table 3.

The literature search was performed by an information specialist using a peer-reviewed search strategy.

Published literature was identified by searching the following bibliographic databases: MEDLINE (1946–) with in-process records and daily updates through Ovid; Embase (1974–) through Ovid; PsycINFO (1987-present) through Ovid, and PubMed. The search strategy consisted of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concepts were Abilify (aripiprazole) and depression.

No methodological filters were applied. Where possible, retrieval was limited to the human population. Retrieval was not limited by publication year or by language. Conference abstracts were excluded from the search results.

The initial search was completed on February 25, 2014. Regular alerts were established to update the search until the meeting of the Canadian Drug Expert Committee on June 18, 2014. Regular search updates were performed on databases that do not provide alert services.

Grey literature (literature that is not commercially published) was identified by searching relevant websites from the following sections of the CADTH Grey Matters checklist (http://www.cadth.ca/en/resources/finding-evidence-is/grey-matters):

- Health Technology Assessment Agencies
- Health Economics
- Clinical Practice Guidelines
- Drug and Device Regulatory Approvals
- Advisories and Warnings
- Drug Class Reviews
- Databases (free)

TABLE 3: INCLUSION CRITERIA FOR THE SYSTEMATIC REVIEW

Patient Population	Adults who had an inadequate response to prior ADT Subgroups by: • Patient's age at diagnosis of MDD • Family history or genetic predisposition • Environmental risk or aggravating factors • Number of previous depression episodes • Baseline symptom severity score (e.g., MADRS or HAM-D17 total score) • Previous ADT drug class (e.g., SSRIs, SNRIs, etc.) • Duration of current episode
Intervention	Aripiprazole at recommended doses in combination with antidepressants used to treat MDD
Comparators	In combination with ADTs to treat MDD (e.g., SNRIs, SSRIs, NRIs, TCAs, MAOIs, other ADTs) • Other AAPs • Add-on with different ADT • Psychotherapy (e.g., CBT, mindfulness, or IPT) • Neurostimulation therapies (e.g., ECT) • Placebo
Outcomes	**Efficacy outcomes:** • HRQoL measured by a validated scale (e.g., SF-36 Health Survey, EQ-5D) • Function/disability measured by a validated scale (e.g., SDS) • Response/remission • Hospitalizations for depression • Symptom severity score rated by patients (e.g., BDI, PHQ-9, IDS-SR) • Symptom severity score rated by physician (e.g., HAM-D17, MADRS) **Harms outcomes:** • Mortality (all-cause and suicide) • Suicidality (ideation/attempts) • SAE • WDAE • AE • Notable AEs: EPS, weight gain, sexual dysfunction, metabolic syndrome
Study Design	Published and unpublished DB RCTs

AAP = atypical antipsychotic; ADT = antidepressant therapy; AE = adverse event; BDI = Beck Depression Inventory; CBT = cognitive behavioural therapy; DB = double-blind; ECT = electroconvulsive therapy; EPS = extrapyramidal symptoms; EQ-5D = EuroQol 5-Dimensions Questionnaire; HAM-D17 = Hamilton Depression Rating Scale – 17 items; HRQoL = health-related quality of life; IDS-SR = Inventory of Depressive Symptomatology–Self-Report; IPT = interpersonal psychotherapy; MADRS = Montgomery–Åsberg Depression Rating Scale; MAOI = monoamine oxidase inhibitor; MDD = major depressive disorder; NRI = norepinephrine reuptake inhibitor; PHQ-9 = Patient Health Questionnaire; RCT = randomized controlled trial; SAE = serious adverse event; SDS = Sheehan Disability Scale; SF-36 = Short Form (36) Health Survey; SNRI = serotonin-norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor; TCA = tricyclic antidepressant; WDAE = withdrawal due to adverse event.

Google and other Internet search engines were used to search for additional Web-based materials. These searches were supplemented by reviewing the bibliographies of key papers and through contacts with appropriate experts. In addition, the manufacturer of the drug was contacted for information regarding unpublished studies.

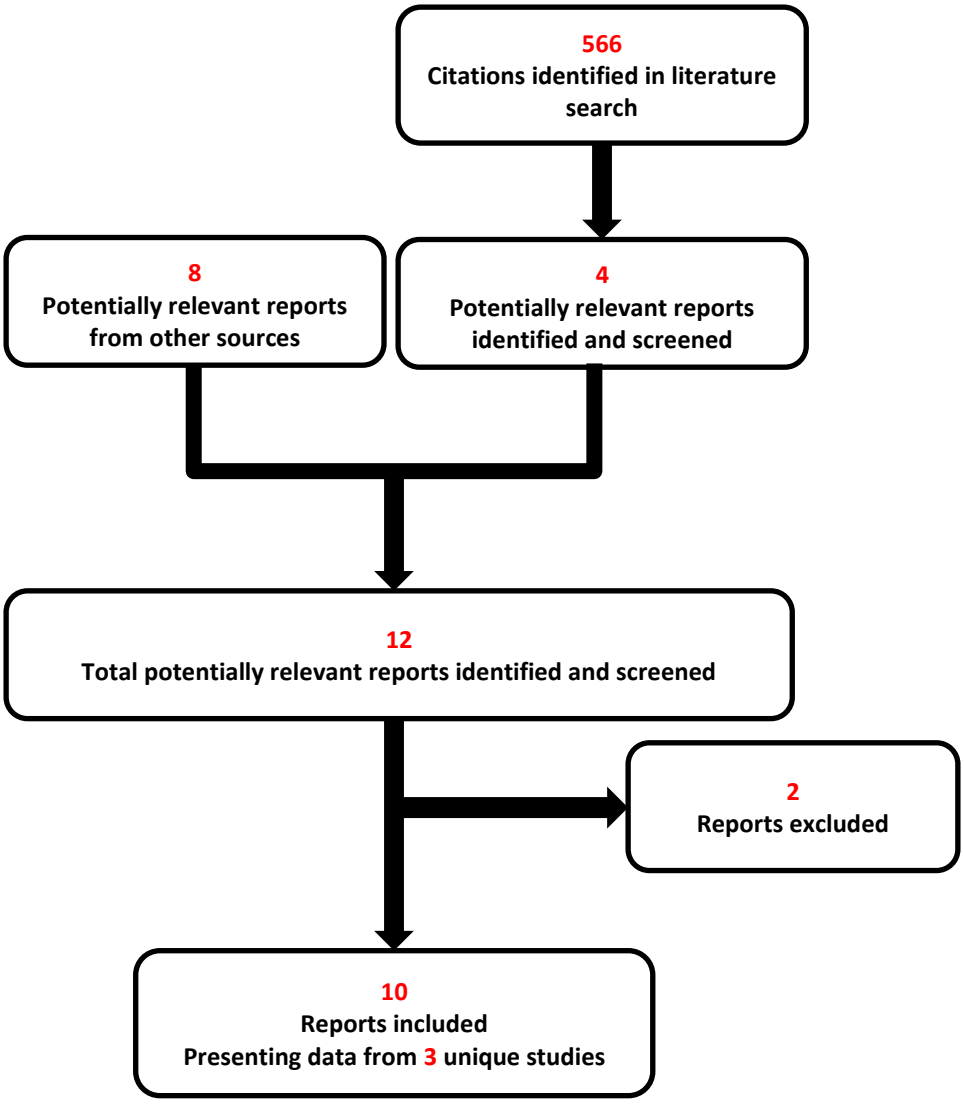
Two CADTH Common Drug Review (CDR) clinical reviewers independently selected studies for inclusion in the review based on titles and abstracts, according to the predetermined protocol. Full-text articles of all citations considered potentially relevant by at least one reviewer were acquired. Reviewers independently made the final selection of studies to be included in the review, and differences were resolved through discussion. Included studies are presented in Table 4; excluded studies (with reasons) are presented in APPENDIX 3: EXCLUDED STUDIES.

3. RESULTS

3.1 Findings From the Literature

A total of three studies were identified from the literature for inclusion in the systematic review (Figure 1). The included studies are summarized in Table 4 and described in Section 3.2. A list of excluded studies is presented in APPENDIX 3: EXCLUDED STUDIES.

FIGURE 1: QUOROM FLOW DIAGRAM FOR INCLUSION AND EXCLUSION OF STUDIES



QUOROM = Quality of Reporting of Meta-analyses.

TABLE 4: DETAILS OF INCLUDED STUDIES

		Study 139	Study 163	Study 165
DESIGNS & POPULATIONS	**Study Design**	DB RCT		
	Locations	24 study sites in the United States	34 study sites in the United States	36 study sites in the United States
	Randomized (N)	362	381	349
	Inclusion Criteria	For phase A (screening): • Adult patients (18 to 65 years old) with a diagnosis of MDE, as defined by DSM-IV-TR criteria • Current MDE ≥ 8 weeks in duration • Current MDE with an inadequate response to at least one and no more than 3 ADTs; response defined as less than a 50% reduction in symptoms assessed by ATRQ; an adequate trial of ADT was defined as treatment for at least 6 weeks (at least 3 weeks for combination treatments) at a minimum dose as specified in the ATRQ.		
		For phase B (run-in): • Patients had a HAM-D17 total score > 18 at the end of phase A • Discontinued all prohibited psychotropic medication during phase A • Able to be rated reliably on the psychiatric scales required by the protocol.		
		For phase C (DB, randomized treatment): • HAM-D17 response < 50% reduction from start of phase B • HAM-D17 score ≥ 14 at the end of phase B • CGI-I score ≥ 3 at the end of phase B.		
	Exclusion Criteria	• Current Axis I (DSM-IV-TR) diagnosis, such as delirium or dementia • Clinically significant current Axis II (DSM-IV-TR) diagnosis such as borderline personality disorder • Any psychotic symptom in the current depressive episode • Inadequate response (< 50% reduction) to more than 3 ADTs during the current depressive episode • Previously not responding to aripiprazole treatment.		
DRUGS	**Intervention**	(As an adjunct to ATD)[a] • Aripiprazole 2 mg/d to 20 mg/d P.O.		
	Comparator(s)	(As an adjunct to ATD)[a] • Placebo		
DURATION	**Phase**			
	Screening (phase A)[b]	7 to 28 d		
	Run-in (phase B)[c]	8 weeks		
	Double-blind (phase C)	6 weeks		
	Follow-up	None		
OUTCOMES	**Primary End Point**	Mean change from end of phase B to end of phase C on the MADRS total score		
	Key Secondary End Point	Mean change from end of phase B to end of phase C in SDS		
	Other End Points	Response on MADRS;[d] remission on MADRS;[e] CGI-S ; CGI-I; HAM-D17 ; IDS-SR; QIDS-SR; Q-LES-Q		

		Study 139	Study 163	Study 165
NOTES	**Publications**	Berman et al. 2007[23]	Marcus et al. 2008[24]	Berman et al. 2009[25]

ADT = antidepressant therapy; ATRQ = Massachusetts General Hospital Antidepressant Treatment Response Questionnaire; DB = double-blind; CGI-I = Clinical Global Impression–Improvement; CGI-S = Clinical Global Impression–Severity of Illness; DSM-IV-TR = *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, Text Revision; HAM-D17 = Hamilton Depression Rating Scale – 17 items; IDS-SR = Inventory of Depressive Symptomatology–Self-Report; MADRS = Montgomery–Åsberg Depression Rating Scale; MDE = major depressive episode; P.O. = orally; QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Report; Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire; RCT = randomized controlled trial; SDS = Sheehan Disability Scale.

[a] ADT was assigned by the investigator with guidance to limit any one ADT to 40% of the patients entering phase B. Percentages are based on the patients who were randomized. For ADTs that were potent CYP2D6 inhibitors (fluoxetine and paroxetine), the dose range was 2 mg to 15 mg per day. For all other ADTs, the dose range was 2 mg to 20 mg per day.

[b] The screening phase A was to enrol patients with a diagnosis of major depressive episode, as defined by DSM-IV-TR criteria, and to have patients discontinue current medication prior to study treatment initiation.

[c] In phase B, patients were assigned placebo plus an open-label marketed ADT that included escitalopram, fluoxetine, paroxetine CR, sertraline, or venlafaxine XR. Due to manufacturing and distribution issues, availability of paroxetine CR was limited; patients had an option of being assigned to treatment with paroxetine (immediate-release formulation). However, no patients received the immediate-release formulation of paroxetine in this study. The choice of ADT was determined by considering each patient's antidepressant treatment history and, unless it was clinically warranted to do otherwise based on the opinion of the study physician, excluding those ADTs with which a patient had reported either an inadequate response or lifetime intolerance.

[d] Response on MADRS: defined as at least 50% reduction from end of phase B in MADRS total score at the end of phase C.

[e] Remission on MADRS: defined as a MADRS total score of 10 or lower and at least 50% reduction from end of phase B in MADRS total score at the end of phase C.

Note: In addition to the three published articles, seven additional reports were included: two US Food and Drug Administration review reports,[26,27] Health Canada review reports,[28] three clinical study reports,[11-13] and submission binder.[15]

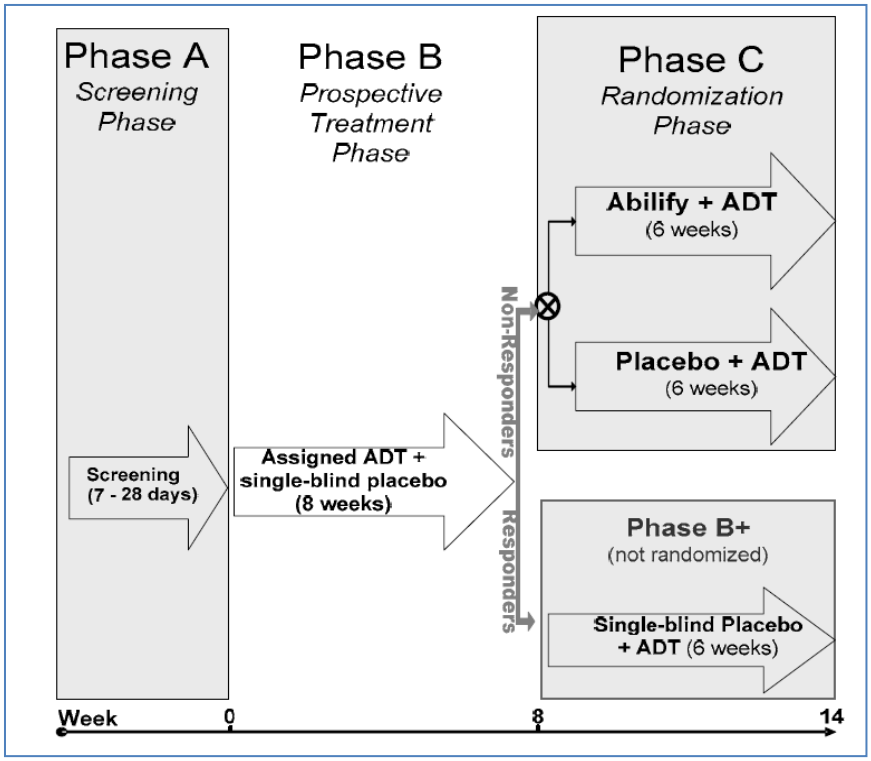
Source: Study 139 Clinical Study Report,[11] Study 163 Clinical Study Report,[12] and Study 165 Clinical Study Report.[13]

3.2 Included Studies

3.2.1 Description of studies

Three studies were identified to meet the inclusion criteria for the review. (In this review, study CN138-139, study CN138-163, and study CN165 are simplified as study 139,[11] study 163,[12] and study 165,[13] respectively.) All three identically designed studies included three phases: a screening phase (phase A); an eight-week open-label prospective ADT plus placebo phase (phase B); followed by a six-week double-blind randomized phase (phase C) for patients who had an incomplete response at the end of phase B (Figure 2). All were multi-centre studies with sites in the US. Sample sizes ranged from 349 to 381 participants. All three studies examined the efficacy and safety of aripiprazole versus placebo as an adjunct to ADT for the treatment of MDD in adults who had an inadequate response to prior ADT during the current episode.

FIGURE 2: STUDY DESIGN OF THE THREE INCLUDED PHASE 3 PLACEBO-CONTROLLED TRIALS



ADT = antidepressant therapy.
Patients who responded to single-blind placebo in combination with ADT at the end of phase B were not randomized but continued on their current single-blind treatment for six weeks (phase B+); however, these patients were not included in the statistical analyses for the outcomes presented in this review.
Source: Study 139 Clinical Study Report,[11] Study 163 Clinical Study Report,[12] and Study 165 Clinical Study Report.[13]

3.2.2 Populations

a) Inclusion and exclusion criteria

The key selection criteria for phase A included patients with a diagnosis of MDE, as defined by *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, Text Revision (DSM-IV-TR) criteria; current depressive episode of eight weeks or more in duration; patients who had reported a history (for the current depressive episode) of an inadequate response (defined as < 50% reduction in depressive symptom severity as assessed by the Massachusetts General Hospital Antidepressant Treatment Response Questionnaire) to at least one and no more than three adequate trials of ADTs (defined as treatment for at least six weeks [or at least three weeks for combination treatments] at a minimum dose as specified in the Antidepressant Treatment Response Questionnaire).

The selection criteria for phase B included patients with a Hamilton Depression Rating Scale, 17-item version (HAM-D_{17}) total score > 18 at the end of phase A. As well, patients must have discontinued all prohibited psychotropic medication during phase A and must have been able to be rated reliably on the psychiatric scales required by the protocol.

Patients who completed phase B and met criteria for an incomplete response (HAM-D17 response < 50% reduction from start of phase B, HAM-D17 score ≥ 14 at the end of phase B, and Clinical Global Impression–Improvement [CGI-I] score ≥ 3 at the end of phase B) were eligible for randomization into phase C. Patients were excluded if they had a current Axis I diagnosis, a clinically significant current Axis II diagnosis, or any psychotic symptom during the current depressive episode. Other exclusion criteria included an inadequate response (< 50% reduction) to more than three ADTs during the current depressive episode, or previous treatment with and lack of response to aripiprazole.

b) Baseline characteristics

Demographic and baseline (i.e., at the end of phase B) characteristics of the randomized controlled trials (RCTs) are shown in Table 5.

Overall, across the included three studies, the demographic characteristics were similar for patients in the placebo and the aripiprazole groups. The mean age of the randomized patients was 44 to 45 years old (range from 19 to 65 years old). Patients were predominantly female (62% to 78%) and Caucasian (87% to 93%). However, more females were randomized to aripiprazole (78%) than to placebo (68%) in study 165.

The baseline psychiatric characteristics of patients were similar in the two treatment groups except for duration of current episode: the placebo-treated patients had a longer duration of current episode than the aripiprazole-treated patients in study 139 and study 163 (median 23.4 months versus 21.0 months, respectively, in study 139, and 19.6 months versus 17.2 months, respectively, in study 163; Table 5). The majority of patients (> 66%) had one previous adequate ADT trial during the current depressive episode, while approximately one-quarter had received two ADTs; there were no differences between treatment groups with respect to the number of previous adequate ADT trials during the current depressive episode in any of the studies. The distribution of ADTs assigned during phase B was similar between the two treatment groups in each study. More than half of patients received either escitalopram or venlafaxine as ADT during phase B, followed by sertraline, fluoxetine, and paroxetine. In addition, treatment groups had similar scores at the end of phase B on HAM-D_{17}; Montgomery–Åsberg Depression Rating Scale (MADRS) total score, and Clinical Global Impression–Severity of Illness (CGI-S) score in all three studies. In terms of severity, overall the patients had moderate MDD at the end of phase B.

TABLE 5: DEMOGRAPHIC CHARACTERISTICS OF RANDOMIZED PATIENTS

Characteristics		Study 139		Study 163		Study 165	
		Placebo N = 178	Aripiprazole N = 184	Placebo N = 190	Aripiprazole N = 191	Placebo N = 172	Aripiprazole N = 177
Age (years)	Mean ± SD	44.1 ± 10.9	46.5 ± 10.6	44.4 ± 10.7	44.6 ± 11.0	45.6 ± 11.3	45.1 ± 10.6
	Median	47.0	48.0	44.5	45.0	47.0	46.0
	Min to Max	21.0 to 64.0	21.0 to 65.0	20.0 to 66.0	19.0 to 67.0	18.0 to 64.0	19.0 to 65.0
Sex n (%)	Male	64 (36.0)	70 (38.0)	62 (32.6)	65 (34.0)	55 (32.0)	39 (22.0)
	Female	114 (64.0)	114 (62.0)	128 (67.4)	126 (66.0)	117 (68.0)	138 (78.0)
Race n (%)	Caucasian	165 (92.7)	161 (87.5)	169 (88.9)	170 (89.0)	149 (86.6)	155 (87.6)
	Black/African-American	10 (5.6)	15 (8.2)	14 (7.4)	14 (7.3)	18 (10.5)	14 (7.9)
Weight (kg)	Mean ± SD	86.4 ± 20.8	84.5 ± 19.5	87.7 ± 23.2	86.5 ± 20.8	88.8 ± 22.8	84.3 ± 21.0
BMI (kg/m^2)	Mean ± SD	30.5 ±7.6	29.5 ±6.7	30.8 ±7.5	30.7±7.4	31.3 ±8.0	30.4 ± 7.2
Duration of current DE (months)	Mean ± SD	43.6 ± 53.5	38.4 ± 58.7	48.5 ± 88.8	43.7 ± 68.0	30.8 ± 36.3	45.7 ± 72.8
	Median	23.4	21.0	19.6	17.2	17.2	18.8
	Min to Max	3.0 to 328.7	1.7 to 474.1	2.1 to 678.8	2.5 to 430.4	1.6 to 236.5	2.1 to 433.1
Age at first DE (years)	Mean ± SD	29.8 ± 13.5	29.9 ± 13.2	26.0 ± 12.6	25.7 ± 13.2	25.8 ± 13.6	25.6 ± 12.4
	Median	27.5	29.0	25.0	22.0	24.0	23.0
	Min to Max	6.0 to 59.0	3.0 to 62.0	2.0 to 58.0	0.0 to 65.0	5.0 to 60.0	2.0 to 57.0
Number of DE	Mean ± SD	3.6 ± 4.0	4.2 ± 4.5	7.3 ± 15.2	6.3 ± 11.8	6.3 ± 10.0	5.3 ± 8.2
	Median	3.0	3.0	3.0	3.0	4.0	3.0
	Min to Max	1.0 to 30.0	0.0 to 40.0	1.0 to 99.0	1.0 to 99.0	0.0 to 99.0	1.0 to 99.0
Number of prior suicide attempts	Mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Median	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Min to Max	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Presence of atypical features n (%)	Yes	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Characteristics		Study 139		Study 163		Study 165	
		Placebo N = 178	Aripiprazole N = 184	Placebo N = 190	Aripiprazole N = 191	Placebo N = 172	Aripiprazole N = 177
Presence of melancholic features n (%)	Yes	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Presence of seasonal pattern n (%)	Yes	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADT at the end of phase B n (%)	Escitalopram	52 (29.2)	55 (29.9)	51 (26.8)	65 (34.0)	52 (30.2)	60 (33.9)
	Fluoxetine	25 (14.0)	26 (14.1)	29 (15.3)	27 (14.1)	25 (14.5)	31 (17.5)
	Paroxetine	14 (7.9)	19 (10.3)	14 (7.4)	13 (6.8)	20 (11.6)	14 (7.9)
	Sertraline	35 (19.7)	37 (20.1)	42 (22.1)	33 (17.3)	30 (17.4)	21 (11.9)
	Venlafaxine XR	52 (29.2)	47 (25.5)	54 (28.4)	53 (27.7)	45 (26.2)	51 (28.8)
Number of previous adequate ADT trials in current DE n (%)	0	0	0	0	0	5 (2.9)	3 (1.7)
	1	118 (66.3)	123 (66.8)	128 (67.7)	135 (71.1)	117 (68.0)	127 (71.8)
	2	46 (25.8)	45 (24.5)	51 (27.0)	49 (25.8)	45 (26.2)	38 (21.5)
	3	14 (7.9)	16 (8.7)	10 (5.3)	5 (2.6)	3 (1.7)	9 (5.1)
	4	0	0	0	1 (0.5)	2 (1.2)	0
	Missing	0	0	1	1	0	0
HAM-D17 total score	Mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Median	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Min to Max	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
MADRS total score	Mean ± SD	26.0 ± 6.5	26.0 ± 6.0	27.0 ± 5.5	25.2 ± 6.2	27.1 ± 5.8	26.6 ± 5.8
	Median	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Min to Max	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SDS mean score	Mean ± SD	5.5 ± 2.4	5.8 ± 2.2	5.5 ± 2.2	5.2 ± 2.4	5.6 ± 2.3	5.4 ± 2.3
	Median	5.7	6.0	5.7	5.3	5.7	5.7
	Min to Max	0.0 to 10.0	0.7 to 10.0	0.0 to 10.0	0.0 to 10.0	0.0 to 10.0	0.0 to 9.7
CGI-S	Mean ± SD	4.1 ± 0.6	4.0 ± 0.6	4.1 ± 0.6	4.0 ± 0.6	4.2 ± 0.6	4.1 ± 0.6
	Median	4.0	4.0	4.0	4.0	4.0	4.0
	Min to Max	3.0 to 6.0	3.0 to 6.0	3.0 to 6.0	3.0 to 6.0	3.0 to 6.0	2.0 to 6.0

Characteristics		Study 139		Study 163		Study 165	
		Placebo N = 178	Aripiprazole N = 184	Placebo N = 190	Aripiprazole N = 191	Placebo N = 172	Aripiprazole N = 177
IDS-SR total score	Mean ± SD	33.3 ± 13.1	33.8 ± 12.8	33.1 ± 11.5	30.9 ± 12.0	32.2 ± 11.2	31.9 ± 11.4
	Median	32.0	33.0	33.0	29.0	31.0	31.0
	Min to Max	3.0 to 74.0	6.0 to 68.0	4.0 to 60.0	2.0 to 65.0	5.0 to 69.0	2.0 to 66.0
QIDS-SR total score	Mean ± SD	12.6 ± 5.0	12.9 ± 4.9	12.6 ± 4.5	11.8 ± 4.5	12.3 ± 4.3	12.4 ± 4.4
	Median	12.5	13.0	13.0	12.0	11.0	12.0
	Min to Max	1.0 to 25.0	2.0 to 26.0	1.0 to 23.0	0.0 to 23.0	3.0 to 25.0	2.0 to 22.0

ADT = antidepressant therapy; BMI = body mass index; CGI-S = Clinical Global Impression–Severity of Illness; DE = depressive episode; HAM-D17 = Hamilton Depression Rating Scale – 17 items; IDS-SR = Inventory of Depressive Symptomatology–Self-Report; MADRS = Montgomery–Åsberg Depression Rating Scale; QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Report; SD = standard deviation; SDS = Sheehan Disability Scale; XR = extended release.
Source: Study 139 Clinical Study Report,[11] Study 163 Clinical Study Report,[12] and Study 165 Clinical Study Report.[13]

3.2.3 Interventions

During the double-blind randomized phase C, patients were randomized to adjunctive aripiprazole or matching placebo. Patients randomized to aripiprazole received one of two dosing ranges based on the ADT they were assigned to in phase B. Patients assigned to escitalopram, sertraline, or venlafaxine extended release (XR) during phase B received 2 mg to 20 mg once daily of aripiprazole, whereas those assigned to fluoxetine, paroxetine, or paroxetine controlled release (CR) during phase B received 2 mg to 15 mg once daily of aripiprazole. Patients randomized to receive aripiprazole started at 5 mg per day, with the possibility of decreasing to 2 mg per day if the original dose was not tolerated. Investigators could increase the dose by up to 5 mg per day each week, to a maximum of 15 mg or 20 mg per day. Allowable aripiprazole doses included 2 mg, 5 mg, 10 mg, 15 mg, and 20 mg per day. No aripiprazole dose increases were allowed after week 4 during the double-blind treatment, although dose reduction for tolerability was permitted at any visit. Open-label ADT was maintained throughout phase C and administered at the same dose as at the end of phase B, which consisted of escitalopram (10 mg to 20 mg per day), fluoxetine (20 mg to 40 mg per day), either paroxetine (20 mg to 40 mg per day) or paroxetine CR (25 mg to 50 mg per day), sertraline (50 mg to 150 mg per day), or venlafaxine XR (37.5 mg to 225 mg per day). The choice of ADT was determined by the investigators after considering each patient's antidepressant treatment history. The use of concomitant medication such as central nervous system medication and medication for dealing with extrapyramidal symptom (EPS)-related adverse events (AEs) were allowed during the studies.

3.2.4 Outcomes

The efficacy and safety outcomes were assessed weekly during phase C. If a patient discontinued prematurely from phase C, the assessment at early termination visit scheduled within 24 hours after the last dose of study medication was used for last observation carried forward (LOCF) imputation.

a) Montgomery–Åsberg Depression Rating Scale (MADRS) total score

The primary outcome in all three studies was the change during phase C (i.e., from end of phase B to the end of phase C) in MADRS total score. MADRS assesses depressive symptomology, particularly change in patients treated with ADTs. This scale is clinician-rated and consists of 10 items; the score ranges from 0 to 60, and higher scores indicate more severe symptoms. The minimal clinically important difference (MCID) for MADRS is a 2-point difference between treatment groups.

b) Sheehan Disability Scale (SDS)

The key secondary outcome was the mean change in score in the Sheehan Disability Scale (SDS) during phase C. The SDS is a short, three-item, self-reported measurement developed to assess the degree to which symptoms of depression, anxiety, panic, and phobia interfere with the patient's work, family, and social life. Each of the items is scored on an 11-point scale (0 to 10); total score ranges from 0 to 30 points, and a higher score indicates poorer function or more severe disability. No MCID was specified.

c) Other outcomes

The HAM-D17 scale is clinician-rated and is most frequently used in efficacy trials. The total score ranges from 0 to 52 (or 53), in which a higher score represents more severe symptoms. No MCID was specified.

Clinical Global Impression–Severity of Illness is a clinician's impression of the patient's illness severity. CGI-S scores range from 1 = "not ill at all" to 7 = "among the most extremely ill." No MCID was specified.

Inventory of Depressive Symptomatology–Self-Report (IDS-SR) is a self-reported 30-item tool that measures depressive symptom severity. Each symptom item is scored on a scale of 0 to 3, with higher scores representing greater symptom severity. No MCID was specified.

Quick Inventory of Depressive Symptomatology–Self-Report (QIDS-SR) is a self-reported 16-item tool that measures depressive symptom severity derived from the IDS. No MCID was specified.

Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) is a 93-item self-reported measure that assesses generic quality of life. Items are scored on a 1 to 5 scale, where higher scores indicate greater enjoyment or satisfaction achieved during the particular activity described in the item. In the included studies, the Q-LES-Q short form, which comprises only 16 items from the fill questionnaire, was used. Three scores were reported, which included Q-LES-Q overall general subscale, satisfaction with medication item, and overall life satisfaction item. No MCID was specified.

Response on MADRS was defined as a ≥ 50% reduction in MADRS total score from the end of phase B, and remission on MADRS was defined as MADRS total score of ≤ 10 and reduction of ≥ 50% from the end of phase B (Appendix 5).

Safety outcomes: Mortality, AEs and serious adverse events (SAEs), potentially clinically relevant abnormalities, changes in weight and body mass index, and sexual dysfunction were reported.

3.2.5 Statistical analysis

Primary efficacy analysis: The mean change in MADRS score during phase C was analyzed based on an LOCF dataset and was assessed by analysis of covariance (ANCOVA), with the end of phase B MADRS total score as covariate, and treatment and study centre as main effects. Superiority testing of MADRS score comparing aripiprazole with placebo was performed. [redacted] The ANCOVA analysis was adjusted with double-blind treatment and study centre as main effects, and MADRS total score at the end of phase B as covariate. Sensitivity analyses were performed with the addition of assigned ADT as main effect. To corroborate the results of the LOCF, the observed case analysis, a longitudinal analysis using direct likelihood estimation, was performed on the mean change from end of phase B in MADRS total score. All secondary continuous outcomes, such as HAM-D_{17}, SDS, IDS-SR, and Q-LES-Q were also analyzed based on the LOCF dataset and assessed by ANCOVA with the end of phase B MADRS total score as covariate, and treatment and study centre as main effects. Response and remission rates were evaluated by the Cochran-Mantel-Haenszel general association test controlling for study centre in the LOCF analyses. If a patient discontinued prematurely during phase C, the assessment at early termination visit scheduled within 24 hours after the last dose of study medication was used for LOCF imputation. Baseline data were not carried forward or averaged with data collected in phase B or phase C, and data from phase B were not carried forward or averaged with phase C on-treatment data to impute missing values for the LOCF dataset.

Subgroup post hoc analyses of change in MADRS total score from end of phase B were performed by gender, ADT, age group, and response on the MADRS at the end of phase B relative to baseline (< 25% improvement versus ≥ 25% improvement). For subgroup analysis, only mean and 95% confidence interval (CI) of the treatment effect were provided (i.e., no *P* values), since these studies were not powered to detect treatment differences in subgroups. Treatment-by-subgroup interaction effects were also assessed at the end of phase C.

No statistical methods were employed to control for multiple testing (or multiplicity) with the secondary outcomes.

a) Analysis populations

Three datasets were analyzed for primary outcomes and all secondary outcomes in all three studies.[11-13] The safety sample comprised all randomized patients who took at least one dose of medication (placebo or aripiprazole) during phase C. The efficacy sample comprised all patients in the safety sample with at least one efficacy evaluation in phase C. The observed cases dataset consisted of the actual observations at a given visit. The LOCF dataset included data recorded at a given visit or, if no observation was recorded at that visit, data carried forward from the previous visit. Baseline data (i.e., at the end of phase B) were not carried forward or averaged with data collected. No pure intention-to-treat (ITT) analysis (i.e., including all RCT patients) was performed. In safety analyses of phase C data, patients were included in the aripiprazole treatment group or the placebo treatment group according to the treatment they initially received, rather than the treatment they were randomized to.

3.3 Patient Disposition

Detailed information on patient disposition in studies 139, 163, and 165 are presented in Table 6. Discontinuation of treatment was lower for the placebo versus aripiprazole groups across all three studies, with respective percentages ranging 10.1% to 14.7% and 12.1% to 16.9%. The most common reason for discontinuation was AEs.

TABLE 6: PATIENT DISPOSITION

	Number of patients (%)					
	Study 139		Study 163		Study 165	
	Placebo	Aripiprazole	Placebo	Aripiprazole	Placebo	Aripiprazole
Screened (phase A)	1044		1151		1147	
Enrolled in phase B	781		830		827	
Completed phase B	622 (79.6)		651 (78.4)		666 (80.5)	
Randomized at phase C	178	184	190	191	172	177
Discontinued from treatment	18 (10.1)	22 (12.1)	28 (14.7)	29 (15.2)	23 (13.4)	30 (16.9)
Lack of efficacy	2 (1.1)	2 (1.1)	3 (1.6)	4 (2.1)	3 (1.7)	2 (1.1)
Adverse event	4 (2.2)	6 (3.3)	2 (1.1)	7 (3.7)	3 (1.7)	11 (6.2)
Patients withdrew consent	4 (2.2)	5 (2.7)	10 (5.3)	3 (1.6)	6 (3.5)	6 (3.4)
Lost to follow-up	4 (2.2)	5 (2.7)	7 (3.7)	5 (2.6)	2 (1.2)	3 (1.7)
Poor/non-compliance	1 (0.6)	2 (1.1)	2 (1.1)	2 (1.0)	4 (2.3)	3 (1.7)
Patients no longer meets study criteria	3 (1.7)	1 (0.5)	4 (2.1)	7 (3.7)	2 (1.2)	5 (2.8)
Other known cause	0 (0.0)	1 (0.5)	0 (0.0)	1 (0.5)	3 (1.7)	0 (0.0)
Completed phase C	160 (89.9)	160 (87.9)	162 (85.3)	162 (84.8)	149 (86.6)	147 (83.1)
Randomized in error[a]	0	2	0	0	0	0
Efficacy[b]	172	181	184	185	169	174
Safety	176	182	190	189	172	176
OC analysis[c]	Varied with different outcomes					
ITT	NR	NR	NR	NR	NR	NR
PP	NR	NR	NR	NR	NR	NR

ITT = intention-to-treat; PP = per-protocol; NR = not reported; OC = observed cases.
[a] Two patients in study 139 were randomized to aripiprazole although they had not completed phase B.
[b] The efficacy sample size varied with the outcomes.
[c] The number of observed cases varied with the outcomes.
Source: Study 139 Clinical Study Report,[11] Study 163 Clinical Study Report,[12] Study 165 Clinical Study Report.[13]

3.4 Exposure to Study Treatments

Detailed information concerning medication exposure is presented in Appendix 4 (Table 15, Table 16 Table 17).

a) Study medication

The mean daily doses of aripiprazole were similar across the three studies at study end point, regardless of the concomitant ADT treatment: 10.7 mg (range 2 mg to 20 mg) in study 163 and study 165, and 11.4 mg (range 2 mg to 20 mg) in study 139. The mean daily dose of aripiprazole in individual ADT groups at end point was as follows: escitalopram 10.9 mg to 11.5 mg; fluoxetine 9.6 mg to 10.1 mg; paroxetine CR 8.9 mg to 10.8 mg; sertraline 10.9 mg to 13.9 mg; and venlafaxine XR 10.2 mg to 11.6 mg.

b) Antidepressant therapy dose

The ADT dose in the placebo group is similar to that in the aripiprazole group in all three studies. The mean daily doses of ADTs at end point for patients who received aripiprazole were as follows: escitalopram 19 mg to 19.8 mg; fluoxetine 36.8 mg to 40 mg; paroxetine CR 48.1 mg to 48.6 mg; sertraline 141.7 mg to 150 mg; and venlafaxine XR 206.6 mg to 213.8 mg. The ADT dose in the placebo group was similar to that in the aripiprazole group in all three studies.

c) Concomitant drug use

[redacted]

3.5 Critical Appraisal

3.5.1 Internal validity

The included three studies were double-blind, multi-centre RCTs. The research objective was clearly defined. The randomization process, including allocation concealment and blinding method, was well described and performed. Overall, the important baseline characteristics were comparable in the two treatment groups. All efficacy outcomes that measured the between-group difference of changes from end of phase B were adjusted for the study centre and baseline assessments.

The limitations of the studies include the lack of an active comparator. Although other AAPs such as quetiapine or risperidone have not been approved by Health Canada for this indication, they are commonly used off-label as adjunctive therapy to ADTs in clinical practice. Patients were required to discontinue ADT upon enrolment in phase A of the studies; however, there was very little description of how ADTs (and AAPs if the patient was on one prior to screening) were tapered. The duration of the screening period (phase A), which would have also acted as the wash-out period, ranged from 7 to 28 days; it is possible that this period was insufficient to properly taper patients' prior ADT. As well, patients were not randomized to ADT (phase B, initiation of ADT for the study) or blinded to treatment; choice of ADT and dosing were based on investigator decision. It did not appear that efficacy outcomes were adjusted for ADT in the statistical analysis, and, therefore, differences in assigned ADT may have affected the results. As well, patients who were ADT-experienced could have had preconceived notions regarding the efficacy and safety of ADT, which could influence assessments on subjective scales. A post hoc analysis of pooled data from all three studies[29] found that the mean improvement in MADRS total score was statistically significantly greater with adjunctive aripiprazole than with adjunctive placebo for both between-class and within-class switch ADT groups. The potential impact of the open-label ADT on the effect measurement, such as HAM-D_{17}, SDS, or Q-LES-Q, is not clear and needs to be further addressed.

In addition, the eligibility of patients included in phase C (double-blind randomized phase) of the studies was based on HAM-D17 response, while the primary outcome and the response and remission rates were reported based on MADRS score. Hence, one symptom severity scale (HAM-D_{17}) was used to decide inclusion in the study, while a separate one (MADRS) was used to assess the primary outcome, which could complicate the interpretation the results. HAM-D17 and MADRS score are highly correlated, but they are not the same questionnaire. The clinical expert involved in the review noted that outcomes on the HAM-D17 may appear better for sedating drugs (versus non-sedating drugs) because the scale has several questions related to improved sleep. Conversely, the MADRS has only one sleep-related question and may show less benefit with a sedating drug. At the lower doses of aripiprazole used in MDD, the drug acts more as a dopamine agonist and is less sedating than at higher doses, or than other AAPs used as adjunctive therapy in MDD. Consequently, using the MADRS as the primary outcome measure may be somewhat more likely to bias findings in favour of aripiprazole than using the HAM-D_{17}. In fact, aripiprazole was statistically and clinically superior to placebo on the MADRS in all three studies but, although the observed findings with the HAM-D17 were statistically significant across studies, their clinical significance was uncertain. As well, because remission and response to treatment are defined using both of these scales, they may differ when using the HAM-D17 instead of the MADRS. Nevertheless, the studies were not designed to specifically detect differences between study groups on the HAM-D_{17}, and both scales are considered valid and acceptable as primary outcomes for the US Food and Drug Administration and Health Canada. According to the manufacturer, the rationale for using different scales for baseline entry (HAM-D_{17}) and primary outcomes assessment (MADRS) was to minimize rater bias.[30]

Patients in the aripiprazole group used more medication to manage EPS-related AEs. The impact of these concomitant drugs on the efficacy observed in the studies is unknown, and actual EPS-related AEs in the aripiprazole group could be higher than reported in reality.

The statistical analysis of the efficacy outcomes was not based on the pure ITT analysis; only efficacy sample and observation case analysis were performed. Although only a small portion of patients were excluded from the efficacy sample analysis (< 3.5% for most outcomes), these exclusions could introduce bias in favour of the aripiprazole or placebo, especially for SDS or Q-LES-Q efficacy analyses, which excluded > 6% of patients.

Depressive symptoms and quality of life were assessed with a number of scales, several of which had multiple subscales. As a result, there were numerous statistical tests performed on the secondary end points, but no statistical method employed to control the type 1 error rate. This increased the risk of finding a statistically significant difference between groups on the secondary end points due to chance.

Finally, the improvement with aripiprazole according to the clinician-rated MADRS in all three studies did not translate into similar improvements with aripiprazole with the patient-rated IDS-SR or QIDS-SR. This discrepancy may be due to the fact that the IDS-SR has been shown to be less sensitive to change than the clinician-rated IDS in short-term studies.[31] In addition, the pivotal studies included in this review may lack power to detect between-group differences on the IDS-SR.

3.5.2 External validity

The generalizability of the findings could be limited because there were no data for patients older than 65 years of age. Patients with established medical comorbidities, which are common in MDD, and patients who had failed more than three prior ADTs were excluded. As well, some other commonly recommended ADTs, such as mirtazapine and bupropion, were not used in the studies. Finally, according to the Canadian Network for Mood and Anxiety Treatments clinical guidelines for the management of MDD in adults, acute treatment for MDD typically lasts 8 to 12 weeks, and maintenance treatment lasts 6 to 24 months or longer.[1] The randomized phase in these trials was six weeks. The findings from these trials may not be generalizable to long-term treatment (more than six weeks).

3.6 Efficacy

Only those efficacy outcomes identified in the review protocol are reported in Section 2.2 (Table 3). See APPENDIX 4: DETAILED OUTCOME DATA for detailed efficacy data. Although the clinical expert involved in the review helped identify factors to define clinically relevant subpopulations, such as patient's age at diagnosis of MDD, family history or genetic predisposition, environmental risk or aggravating factors, number of previous depression episodes, or duration of current episode, there were no data reported for these subpopulations.

3.6.1 Health-related quality of life

In the included studies, Q-LES-Q was reported in three scores (i.e., Q-LES-Q overall general subscore, satisfaction with medication item, and overall life satisfaction item) (Table 7 and Table 9). Statistically significant differences between the treatment groups in favour of aripiprazole in mean changes during phase C in the Q-LES-Q overall life satisfaction item were reported in all three studies. [redacted] Statistically significant between-group differences in mean changes during phase C in the Q-LES-Q overall general subscore in favour of aripiprazole were also observed in study 163 and study 165, [redacted] No statistically significant differences were observed between the groups in Q-LES-Q satisfaction with medication item in any of the studies.

3.6.2 Function/disability — Sheehan Disability Scale

Functional capacity was measured using the SDS in all three included studies. In study 139, the mean change during phase C in SDS mean score was not statistically significantly different between the treatment groups (Table 7), although statistical significance in favour of aripiprazole was demonstrated on social life [redacted] $P = 0.03$) and family life scores [redacted] $P = 0.017$) (Table 9). In study 163, there was a statistically significant difference between groups in favour of aripiprazole in the mean change during phase C in SDS mean score [redacted] Statistical significance in favour of aripiprazole was also demonstrated on individual social life or family life item scores (–0.81, $P = 0.002$ for both). In study 165, a statistically significant difference between the treatment groups in favour of aripiprazole was reported only for family life (–0.57, $P = 0.037$). None of the studies showed superiority for aripiprazole compared with placebo in the work/school item (Table 9).

3.6.3 Hospitalizations for depression

[redacted]

3.6.4 Response and remission

Response was defined by at least 50% decrease in the MADRS total score during phase C. Statistically significantly more patients in aripiprazole groups demonstrated response as compared with those receiving placebo during phase C in all three studies (Table 7). The response rates (aripiprazole versus placebo) were 34% versus 24%, 32% versus 17%, and 47% versus 27% in study 139, study 163, and study 165, respectively. The number needed to treat for response ranged from 5 to 11 across the studies.

Remission was defined by at least 50% reduction in the MADRS total score during phase C and a MADRS total score of ≤ 10 at study end point. Statistically significantly more patients in the aripiprazole group than in the placebo group achieved remission during phase C in all three studies (Table 7). The remission rates (aripiprazole versus placebo) were 26% versus 16%, 25% versus 15%, and 37% versus 19% in study 139, study 163, and study 165, respectively. The number needed to treat ranged from 6 to 10 across the studies. Data on treatment response and remission using HAM-D17 during phase C were not reported.[30]

3.6.5 Montgomery–Åsberg Depression Rating Scale total score and Hamilton Depression Rating Scale score

Mean change from end of phase B to the end of phase C in the MADRS total score was the primary outcome in the included trials. The mean treatment-group difference (95% CI) in changes in MADRS total score during phase C was statistically significantly in favour of aripiprazole versus placebo: –3.01 (–4.66 to –1.37), –2.84 (–4.53 to –1.15), and –3.73 (–5.44 to –2.02) in studies 139, 163, and 165, respectively (Table 7). All exceeded the MCID of 2.

Subgroup analyses of mean change during phase C were performed by ADT received during phase C, and on baseline MADRS response (< 25% versus ≥ 25% improvement in MADRS score). The results consistently favoured aripiprazole across all subgroups in all three studies (Appendix 4, Table 10).

[redacted] (Table 7). [redacted]

3.6.6 Other symptom scales

Across all three studies, no statistically significant treatment-group difference was observed in the mean change in the IDS-SR total score during phase C (Table 7). Similar results were found for the QIDS-SR total score during phase C (Table 9).

A statistically significant difference between the treatment groups in favour of aripiprazole in the mean change in the CGI-S score during phase C was observed in all three studies (Table 7).

TABLE 7: EFFICACY OUTCOMES

	Study 139		Study 163		Study 165	
Outcomes During Phase C	**Placebo**	**Aripiprazole**	**Placebo**	**Aripiprazole**	**Placebo**	**Aripiprazole**
Efficacy sample size[a]	N = 172	N = 181	N = 184	N = 185	N = 169	N = 174
Patients in MADRS response						
n (%)	41 (23.8)	61 (33.7)	32 (17.4)	60 (32.4)	45 (26.6)	81 (46.6)
RR (95% CI)	1.45 (1.04 to 2.01) *P* = 0.027		1.86 (1.27 to 2.71) *P* < 0.001		1.74 (1.31 to 2.32) *P* < 0.001	
ARR (95% CI)	0.10 (0.00 to 0.19) *P* = 0.04		0.15 (0.06 to 0.24) *P* = 0.0007		0.20 (0.10 to 0.30) *P* < 0.0001	
NNT	11		7		5	
Patients in MADRS remission						
n (%)	27 (15.7)	47 (26.0)	28 (15.2)	47 (25.4)	32 (18.9)	64 (36.8)
RR (95% CI)	1.70 (1.13 to 2.56) *P* = 0.011		1.66 (1.09 to 2.54) *P* = 0.016		1.95 (1.36 to 2.80) *P* < 0.001	
ARR (95% CI)	0.10 (0.02 to 0.19) *P* = 0.02		0.10 (0.02 to 0.19) *P* = 0.02		0.18 (0.09 to 0.26) *P* = 0.0004	
NNT	10		10		6	
MADRS total score						
At the end of phase B[b] (mean ± SE)	25.65 ± 0.51	25.88 ± 0.48	26.55 ± 0.44	24.59 ± 0.44	26.72 ± 0.53	26.22 ± 0.52
Change during phase C (mean ± SE)	–5.77 ± 0.67	–8.78 ± 0.63	–5.65 ± 0.64	–8.49 ± 0.66	–6.39 ± 0.74	–10.12 ± 0.74
Between-group difference in changes Mean difference (95% CI), *P* value	–3.01 (–4.66 to –1.37) *P* < 0.001		–2.84 (–4.53 to –1.15) *P* = 0.001		–3.73 (–5.44 to –2.02) *P* < 0.001	
HAM-D17 score						
At the end of phase B[b] (mean ± SE)	[redacted]		[redacted]		[redacted]	
Change during phase C (mean ± SE)	[redacted]		[redacted]		[redacted]	
Between-group difference in changes Mean difference (95% CI), *P* value	[redacted]		[redacted]		[redacted]	
SDS mean score	N = 164	N = 167	N =168	N = 180	N = 160	N = 160
At the end of phase B[b] (mean ± SE)	5.35 ± 0.20	5.69 ± 0.19	5.35 ± 0.19	5.06 ± 0.19	5.89 ± 0.22	5.67 ± 0.22
Change during phase C (mean ± SE)	–0.65 ± 0.19	–1.11 ± 0.18	–0.73 ± 0.18	–1.31 ± 0.17	–0.80 ± 0.20	–1.22 ± 0.21
Between-group difference in changes Mean difference (95% CI), *P* value	–0.46 (–0.93 to 0.01) *P* = 0.055		–0.57 (–1.02 to –0.13) *P* = 0.012		–0.42 (–0.88 to 0.04) *P* = 0.075	

	Study 139		Study 163		Study 165	
Outcomes During Phase C	**Placebo**	**Aripiprazole**	**Placebo**	**Aripiprazole**	**Placebo**	**Aripiprazole**
Q-LES-Q (short form) overall general subscore	N = 164	N = 167	N =168	N = 180	N = 160	N = 160
At the end of phase B[b] (mean ± SE)	[redacted]	[redacted]	[redacted]	[redacted]	44.26 ± 1.33	43.75 ± 1.34
Change during phase C (mean ± SE)	[redacted]	[redacted]	[redacted]	[redacted]	5.16 ± 1.39	9.81 ± 1.41
Between-group difference in changes Mean difference (95% CI), *P* value	[redacted]		[redacted]		4.65 (1.50 to 7.81) *P* = 0.04	
CGI–Severity of Illness score						
At the end of phase B[b] (mean ± SE)	4.11 ± 0.05	4.08 ± 0.04	4.0 7± 0.04	4.02 ± 0.04	4.20 ± 0.05	4.10 ± 0.05
Change during phase C (mean ± SE)	–0.64 ± 0.08	–1.03 ± 0.08	–0.63 ± 0.08	–1.10 ± 0.08	–0.69 ± 0.10	–1.14 ± 0.10
Between-group difference in changes Mean difference (95% CI), *P* value	–0.39 (–0.59 to –0.18) *P* < 0.001		–0.48 (–0.68 to –0.27) *P* < 0.001		–0.44 (–0.67 to –0.22) *P* < 0.001	
IDS-SR total score						
At the end of phase B[b] (mean ± SE)	34.04 ± 1.10	34.43 ± 1.03	32.34 ± 0.94	30.27 ± 0.93	33.04 ± 1.06	32.72 ± 1.05
Change during phase C (mean ± SE)	–5.16 ± 0.81	–6.95 ± 0.76	–4.55 ± 0.73	–6.03 ± 0.73	–5.36 ± 0.86	–6.93 ± 0.85
Between-group difference in changes Mean difference (95% CI), *P* value	–1.79 (–3.77 to 0.19) *P* = 0.076		–1.47 (–3.36 to 0.42) *P* = 0.076		–1.57 (–3.54 to 0.40) *P* = 0.118	

ARR = absolute risk reduction; CGI = Clinical Global Impression; CI = confidence interval; HAM-D17 = Hamilton Depression Rating Scale – 17 item version; IDS-SR = Inventory of Depressive Symptomatology–Self-Report; MADRS = Montgomery–Åsberg Depression Rating Scale; NNT=number needed to treat; Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire; RR = relative risk; SDS = Sheehan Disability Scale; SE = standard error.

[a] Sample size for all outcomes except for SDS or Q-LES-Q; all continuous outcome values expressed as mean ± SE.

[b] These values were equivalent to "baseline" for phase C.

Source: Study 139 Clinical Study Report,[11] Study 163 Clinical Study Report,[12] and Study 165 Clinical Study Report.[13]

3.7 Harms

Only those harms identified in the review protocol are reported below (Table 3). See APPENDIX 4: DETAILED OUTCOME DATA for detailed harms data. Harms data from the included studies are reported as treatment-emergent adverse events (TEAEs) that occurred in at least 5% of patients in either of the treatment groups. As well, serious AEs, mortality, and withdrawals due to adverse events, and notable AEs identified for the review following discussion with the clinical expert are reported.

3.7.1 Adverse events

The TEAEs were analyzed in the studies with the safety population set. Fewer placebo-treated patients (62.5% to 63.2% patients) versus aripiprazole-treated patients (80.7% to 82.0%) experienced AEs across all three studies. Most AEs were mild to moderate in severity. The TEAEs that occurred more frequently in the aripiprazole group than in the placebo group (at least twice the rate in placebo) were akathisia (23.1%), restlessness (14.3%), upper respiratory infection (8.2%), insomnia (7.7%), and blurred vision (6.6%) in study 139; akathisia (25.9%), fatigue (10.1%), restlessness (9.5%), insomnia (7.4%), tremor (6.3%), and constipation (5.3%) in study 163; akathisia (18.2%), restlessness (12.5%), blurred vision (7.4%), and somnolence (5.7%) in study 165.

3.7.2. Serious adverse events

During phase C, the overall incidence of SAEs was low in all three studies. In study 139, five patients reported SAEs; three (1.7%) in the placebo group and two (1.1%) in the aripiprazole group. The SAEs consisted mainly of infections, injuries, and exostosis. In study 163, one patient from the aripiprazole group reported an SAE (cellulitis); no SAEs were reported in the placebo group. In study 165, two patients reported SAEs (arterial occlusive disease and suicidal ideation), one (0.6%) in each treatment group.

3.7.3 Withdrawal due to adverse events

In study 139, there were seven (1.9%) patients who discontinued study therapy during phase C because of an AE: three (1.7%) were in the placebo group and four (2.2%) in the aripiprazole group. In study 163, nine (2.4%) patients discontinued study therapy due to an AE: two (1.1%) were in the placebo group and seven (3.7%) in the aripiprazole group. In study 165, 13 (3.7%) patients discontinued study therapy because of an AE: three (1.7%) in the placebo group and 10 (5.7%) in the aripiprazole group. Most frequently reported AEs that led to discontinuation were akathisia, fatigue, or blurred vision in the aripiprazole group, and oral pain, pharyngolaryngeal pain, or depression in the placebo group.

3.7.4 Mortality

No deaths were reported in the three studies during phase C.

3.7.5 Notable harms

After consultation with the clinical expert involved in the review, the following notable harms (i.e., AEs with special interest clinically) were identified: EPS including akathisia, weight gain, sexual dysfunction, and metabolic syndrome. Patients who received aripiprazole ([redacted] experienced more EPS-related AEs than those in the placebo group ([redacted] across all three studies). Likewise, potentially clinically important weight gain (≥ 7% increase in body weight) occurred more frequently among patients treated with aripiprazole (2.1% to 3.4%) than among those who received placebo (0.5% to 2.8%) across all three studies. Sexual dysfunction was reported in [redacted] who received aripiprazole [redacted]. No cases defined as metabolic syndrome were reported in any of the included studies.

TABLE 8: HARMS

	Study 139		Study 163		Study 165	
AEs (≥ 5% PTS IN EITHER GROUP)	Placebo	Aripiprazole	Placebo	Aripiprazole	Placebo	Aripiprazole
Safety sample size (N)	176	182	190	189	172	176
Patients with ≥ 1 AE, n (%)	110 (62.5)	149 (81.9)	120 (63.2)	154 (81.5)	118 (68.6)	142 (80.7)
Most common AEs, n (%)						
Akathisia	8 (4.5)	42 (23.1)	8 (4.2)	49 (25.9)	6 (3.5)	32 (18.2)
Headache	19 (10.8)	11 (6.0)	20 (10.5)	17 (9.0)	14 (8.1)	15 (8.5)
Restlessness	6 (3.4)	26 (14.3)	1 (0.5)	18 (9.5)	6 (3.5)	22 (12.5)
Insomnia	4 (2.3)	14 (7.7)	3 (1.6)	14 (7.4)	9 (5.2)	15 (8.5)
Nausea	9 (5.1)	5 (2.7)	8 (4.2)	10 (5.3)	10 (5.8)	7 (4.0)
Fatigue	6 (3.4)	11 (6.0)	7 (3.7)	19 (10.1)	8 (4.7)	16 (9.1)
SAEs						
Patients with ≥ 1 SAE, n (%)	3 (1.7)	2 (1.1)	0	1 (0.5)	1 (0.6)	1 (0.6)
Most common SAEs						
Infection	2 (1.2)	2 (1.1)	0	1 (0.5)	0	0
WDAEs						
WDAEs, n (%)	4 (2.2)	6 (3.3)	2 (1.1)	7 (3.7)	3 (1.7)	11 (6.2)
Most common reasons						
Akathisia	0	1 (0.5)	0	2 (1.1)	0	2 (1.1)
Fatigue	0	0	0	2 (1.1)	0	0
Vision blurred	0	0	0	2 (1.1)	0	0
Deaths						
Number of deaths, n (%)	0	0	0	0	0	0
Notable harms						
Metabolic syndrome	NR	NR	NR	NR	NR	NR
EPS[a]	17 (9.7)	50 (27.5)	[redacted]	[redacted]	[redacted]	[redacted]
Sexual dysfunction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weight gain[b]	5 (2.8)	5 (2.7)	1 (0.5)	4 (2.1)	3 (1.7)	6 (3.4)

AE = adverse event; EPS = extrapyramidal symptoms; NR = not rated; pts = patients; SAE = serious adverse event; WDAE = withdrawal due to adverse event.

[a] EPS including akathisia.

[b] Potentially clinically relevant weight gain (an increase of at least 7% from end of phase B).

Source: Study 139 Clinical Study Report,[11] Study 163 Clinical Study Report,[12] and Study 165 Clinical Study Report.[13]

4. DISCUSSION

4.1 Summary of Available Evidence

The evidence for this review was derived from double-blind, randomized, placebo-controlled trials evaluating aripiprazole (2 mg to 20 mg daily) as adjunctive therapy to ADT in patients with MDD who had an inadequate response to prior ADT. Overall, the important baseline characteristics were comparable in the two treatment groups. The dropout rates were comparable between two treatment groups in all three studies, although slightly higher for the aripiprazole groups. Patient-reported outcomes and valid symptom severity scales were evaluated; however, MCIDs have not been reported for most of the measures, making it challenging to interpret the findings with respect to clinical relevance. Other key limitations potentially limiting the internal validity or generalizability of the findings from these studies include the following. Patients were required to discontinue ADT upon enrolment in phase A of the studies; however, there was very little description of the tapering of ADTs (and AAP if the patient was receiving one prior to screening). The duration of the screening period (phase A), which would have also acted as the wash-out period, ranged from 7 to 28 days; this period may have been insufficient to properly taper patients' prior ADT. As well, patients were not randomized to ADT (phase B, initiation of ADT for the study) or blinded to treatment; choice of ADT was based on investigator decision. Efficacy outcomes do not appear to have been adjusted for ADT in the statistical analysis; therefore, differences in assigned ADT may have affected the results. As well, patients who were ADT-experienced could have had preconceived notions regarding the efficacy and safety of ADT, which could influence assessments on subjective scales. Nevertheless, the decision concerning which ADT to administer and the open-label design of phase B may reflect circumstances in clinical practice. [redacted] The eligibility of patients in the randomized phase of each study was based on HAM-D17, while the primary outcome and the response and remission rates were reported based on MADRS score. The use of different scores could complicate the interpretation of the response and remission rate, although the rationale to use the different scores provided by the manufacturer was to minimize rater bias.[30] The analysis was not based on ITT analysis, although only a small proportion of patients were excluded from the efficacy sample analysis (< 3.5% to 6%). The generalizability of the findings could also be compromised as a result of the exclusion of patients older than 65, patients with inadequate response to three or more previous ADTs, and other commonly recommended first-line ADTs such as mirtazapine and bupropion, as well as the relatively short trial duration (six weeks). There was no adjustment for the multiple statistical testing performed on the secondary outcomes, thereby increasing the risk of a false-positive outcome. Interpretation of the secondary outcomes is further complicated by the lack of established MCID for a number of the instruments used (such as the Q-LES-Q and SDS). Thus, the clinical and statistical significance of the observed between-group differences remains uncertain.

4.2 Interpretation of Results

4.2.1 Efficacy

The evidence reviewed comes from three identically designed placebo-controlled RCTs. The comparative effectiveness of aripiprazole versus other active comparators (primarily other AAPs) as adjunctive therapy to ADTs for MDD is unknown due to a lack of head-to-head studies.

In terms of health-related quality of life, a statistically significantly greater improvement in patients receiving aripiprazole compared with placebo during phase C in the Q-LES-Q overall general subscore was observed in study 163 and study 165. A statistically significantly greater improvement in aripiprazole compared with placebo during phase C in the Q-LES-Q overall life satisfaction item was also observed in all three studies. However, the clinical importance of these findings is uncertain because an MCID has not been reported for treatment differences in the Q-LES-Q.

A statistically significant difference between treatment groups, in favour of aripiprazole, in the overall SDS mean score was demonstrated in only one of the studies (study 163). A statistically significant difference between the treatment groups, in favour of aripiprazole, was demonstrated in SDS–family life item score in all three studies, and in SDS–social life item score in study 139 and study 163. However, whether these improvements were clinically meaningful is not clear, and, therefore, the effect of adjunctive aripiprazole on functional capacity and disability remains to be fully determined.

Treatment with aripiprazole showed superiority over placebo for MADRS response and remission rates in all three studies. The findings were consistent with those reported in previous meta-analysis.[14,32,33] An important limitation of these findings is that the durability of response and remission (i.e., over the long-term) is unknown. Response and remission were not outcomes in the 52-week extension study (Appendix 6: SUMMARY OF OTHER STUDIES).[34] In the absence of head-to-head trial data for aripiprazole versus other adjunctive AAP comparators, the manufacturer conducted a network meta-analysis (NMA) based on RCTs [redacted].[15,16] [redacted] The NMA methodology was appropriate though was limited by the heterogeneity between the studies and patients included in the analysis.

Symptom severity was assessed using both investigator- (MADRS and HAM-D17 total scores) and patient-rated (QIDS-SR and IDS-SR total scores) scales. Results from the three included studies demonstrated statistically significantly superior efficacy of aripiprazole versus placebo in alleviating depressive symptoms — based on changes in the MADRS and HAM-D17 total scores from baseline — in patients over a six-week treatment period. As well, aripiprazole achieved a clinically significantly greater improvement compared with placebo for the change from baseline MADRS score in all three studies (mean treatment differences of –2.84 to –3.73; MCID = 2).[35,36] [redacted].[17] [redacted])[37] [redacted] (Appendix 5). [redacted]. No statistically significant treatment-group difference in the IDS-SR total score or QIDS-SR score was found at the end of the RCT. The main limitation of the evidence is that the clinical improvement with aripiprazole evaluated with clinician-rated MADRS total score in all three studies did not translate into similar improvements with the patient-rated IDS. This finding may be due to the fact that the IDS-SR has been shown to be less sensitive to change than the clinician-rated IDS in short-term studies. The concerns regarding the impact of aripiprazole and other AAPs as adjunctive therapy in improving overall well-being was also raised in a systematic review by Spielmans et al.[14] According to the Canadian Network for Mood and Anxiety Treatments, the goal of treatment for MDD should be remission (i.e., resolution of depressive

symptoms).[20] The cut-off point for remission reportedly ranged from < 8 to < 10 (MADRS score) (Appendix 7). The clinical expert involved in this review also indicated that the lower cut-off (< 8 points) is a better goal for evidence of remission, since relapse is higher if symptoms are not completely resolved.

4.2.2 Harms

There were no deaths reported in the three trials. Overall, TEAEs were more common with aripiprazole than with placebo across the three studies (80.7% to 82.0% versus 62.5% to 63.2%). The most common TEAEs reported among aripiprazole-treated patients included akathisia, restlessness, fatigue, insomnia, blurred vision, somnolence, and constipation. More patients in the aripiprazole group ([redacted] reported EPS-related AEs than in placebo group ([redacted] Weight gain (of more than 7%) was infrequent and comparable in both treatment groups, but was assessed only over a six-week duration. Sexual dysfunction was reported in [redacted] who received aripiprazole. Overall, the incidence of SAEs was infrequent and comparable in both groups. SAEs included infections, injuries, exostosis, cellulitis, arterial occlusive disease, and suicidal ideation. More patients in the aripiprazole group discontinued treatment due to AEs compared with those in the placebo group (5.9% versus 1.7%). The most common AEs leading to discontinuation in the aripiprazole group were akathisia (1.3%) and fatigue (0.7%). The overall safety profile is comparable to other AAPs.[14,16] Among AAPs, there appears to be an increased risk of akathisia associated with aripiprazole and of weight gain associated with olanzapine.[14]

5. CONCLUSIONS

In the three included double-blind randomized placebo-controlled trials, MADRS remission and response rates as well as changes in MADRS and HAM-D17 scores demonstrated statistically significantly greater improvements with adjunctive aripiprazole compared with placebo. A statistically significant improvement in health-related quality of life (by Q-LES-Q overall life satisfaction score and CGI) was observed in all three studies. As well, improvement in function/disability (by SDS) was statistically significant in some of the subscores in favour of aripiprazole in some studies. However, the clinical significance of the between-group differences in health-related quality of life and functional capacity is uncertain because of the lack of an MCID for these. In terms of patient-reported symptoms, measured with IDS-SR or QIDS-SR, no statistically significant benefit was reported with aripiprazole compared with placebo. Serious adverse events were infrequent and too few to draw conclusions. Common adverse events associated with aripiprazole were EPS-related AEs, such as akathisia. Other TEAEs such as upper respiratory infection, insomnia, and blurred vision also occurred more frequently with aripiprazole than placebo. Akathisia, somnolence, [redacted], and insomnia were the common reasons for discontinuing treatment with aripiprazole. In summary, compared with placebo, adding aripiprazole to ADT statistically significantly improved symptoms on investigator-rated scales and increased response and remission rates in patients with MDD who had inadequate response to ADT treatment alone. However, the findings from the three studies do not clearly indicate whether aripiprazole is superior to placebo in improving health-related quality of life or patients' functioning. [redacted].

The main limitations of the body of evidence for aripiprazole as adjunctive therapy to ADT in MDD included the following: an active comparator was lacking; clinician-rated MADRS did not translate into similar improvements with aripiprazole with the patient-rated IDS-SR or QIDS-SR; and data demonstrating clear superiority over placebo for health-related quality of life or function/disability outcomes were lacking.

APPENDIX 1: PATIENT INPUT SUMMARY

This section was summarized by CADTH Common Drug Review (CDR) staff based on the input provided by patient groups. It has not been systematically reviewed. It has been reviewed by the submitting patient groups.

1. Brief Description of Patient Group(s) Supplying Input

Mood Disorders Society of Canada (MDSC) is a national voluntary consumer/patient-controlled health charity. Its goal is to improve access to treatment, inform research, and shape program development and government policies to improve the quality of life for people affected by mood disorders. MDSC does not have corporate members. MDSC has joint working relationships with a large number of collaborators, including the Canadian Pharmacists Association, Canadian Medical Association, Canadian Institutes for Health Research, members of Canada's academic research community, the Mental Health Commission of Canada, the Neurological Health Charities Canada, Canadian Alliance on Mental Illness and Mental Health, and other national and local mental health and chronic disease non-governmental organizations. MDSC has no core funding from any level of government and relies on project funding, donations, and unrestricted educational grants from pharmaceutical companies, charitable foundations, and others. Bristol-Myers Squibb Canada played no role in the compilation of this submission.

2. Condition- and Current Therapy-Related Information

Information was collected from personal histories of key MDSC staff living with depression/major depressive disorder (MDD) and from a literature review. MDSC hosts a national online discussion forum containing more than 18,000 posts from persons with mental illness and their families and caregivers. MDSC has an in-depth knowledge of both disorders and the patient perspective on the needs for innovative new medicines to address mental illness.

According to the MDSC Quick Facts publication, depression is the leading cause of years lived with disability and is the fourth leading cause of disability and premature death in the world. By 2020, it is predicted that depression will become the second leading cause of disability in the world after heart disease. Approximately 7.9% of Canadians will experience depression in their lifetime. Women are twice as likely as men to experience depression and 1.5 times more likely to be hospitalized with depression as men. The majority (80%) of individuals with depression respond well to treatment, although approximately 90% of depressed individuals never seek treatment. Depression is on par with smoking as a predictor of mortality.

Depression can be described as a case of too much or too little: an individual may be sleeping little or sleeping too much, have gained or lost weight, be highly agitated or sluggish and inert, be extremely sad or very bad tempered, or both. Patients with depression may feel a loss of interest in the pleasures of life, work, family and friends; be unable to concentrate and make decisions; feel negative, anxious, trapped, unable to act, despairing, guilty, and unworthy. They may experience fatigue and an overall loss of energy. They may be suicidal – expressing thoughts and sometimes planning. They may feel numb and may experience unexplained aches and pains.

Depression can be diagnosed if an individual has experienced at least five of the symptoms previously described for a period of two weeks or longer. Physical symptoms associated with depression include headaches, back pain, muscle and joint pain, chest pain, digestive problems and pain, exhaustion and fatigue, sleeping problems, change in appetite or weight, and dizziness or light-headedness. The most common symptoms experienced by Canadians include lack of motivation, loss of ability to enjoy

favourite activities, difficulty concentrating, and feeling of isolation. As symptoms vary by individual, so too do activities that individuals cannot undertake while depressed.

Current therapies include pharmaceuticals, self-help and peer support, cognitive behavioural therapy, and psychotherapy. These therapies are reasonably effective; however, many of the currently available pharmaceutical therapies have severe to moderate side effects. The medications psychiatrists most associate with helping reduce the functional impairment of depression are serotonin-norepinephrine reuptake inhibitors and selective serotonin reuptake inhibitors. Sometimes “current therapies” become less effective for patients on a long-term basis, and they therefore need access to newer medications to maintain control of their illness.

Challenges faced by caregivers include educating families and friends about depression and its symptoms and treatments, and taking care of themselves. Caregivers also have to face the stigma of mental illness that exists in our society. Caregivers must always be alert to ensure their loved one is taking medications as prescribed and noting any real or perceived side effects. Patients experiencing side effects due to their medication may refuse to continue to take it, and then the fear of “recurring” becomes an enormous burden.

3. Related Information About the Drug Being Reviewed

Abilify has been used by patients in Canada; however, the newest and best medications for any mood disorder, including depression, should be available to the community for future needs. Based on MDSC’s extensive national knowledge developed over the past 12 years, many persons living with depression want a medication that addresses their illness without side effects that can be almost as discomforting as the illness itself. MDSC expects that patients living with depression could enjoy a better quality of life and a more stable adherence to their medication schedule if weight gain, in particular, was not of concern.

Patients living with depression should never be exposed to medications with more serious side effects than those associated with currently available medications. New therapies should be at least no worse than those currently available. If new therapies are at least as good as those already available or even only marginally better by virtue of less impact from side effects, then they should be made available. New therapies are another option for the treatment of MDD, and they can give new hope to the patients/consumers struggling to gain control of their lives.

APPENDIX 2: LITERATURE SEARCH STRATEGY

OVERVIEW	
Interface:	Ovid
Databases:	Embase 1974 to present MEDLINE Daily and MEDLINE 1946 to present MEDLINE In-Process & Other Non-Indexed Citations PsycINFO 1987 to present **Note:** Subject headings have been customized for each database. Duplicates between databases were removed in Ovid.
Date of Search:	February 25, 2014
Alerts:	Weekly search updates until June 18, 2014
Study Types:	No search filters were applied
Limits:	No date or language limits were used Human filter was applied Conference abstracts were excluded

SYNTAX GUIDE	
/	At the end of a phrase, searches the phrase as a subject heading
.sh	At the end of a phrase, searches the phrase as a subject heading
MeSH	Medical Subject Heading
fs	Floating subheading
exp	Explode a subject heading
*	Before a word, indicates that the marked subject heading is a primary topic; or, after a word, a truncation symbol (wildcard) to retrieve plurals or varying endings
#	Truncation symbol for one character
?	Truncation symbol for one or no characters only
adj	Requires words are adjacent to each other (in any order)
adj#	Adjacency within # number of words (in any order)
.ti	Title
.ab	Abstract
.ot	Original title
.hw	Heading word; usually includes subject headings and controlled vocabulary
.pt	Publication type
.po	Population group [PsycInfo only]
.rn	CAS registry number
.nm	Name of substance word
pmez	Ovid database code; MEDLINE In-Process & Other Non-Indexed Citations, MEDLINE Daily and Ovid MEDLINE 1946 to Present
oemezd	Ovid database code; Embase 1974 to present, updated daily

MULTI-DATABASE STRATEGY	
1	(abilify* or aripiprazole* or abilitat* or discmelt or HSDB7320 or HSDB 7320 or OPC31 or OPC 31 or OPC14597 or OPC 14597).ti,ab,hw,ot.
2	129722-12-9.rn,nm.
3	1 or 2
4	(depression* or depressive).ti,ab,hw.
5	3 and 4
6	5 use psyf
7	5 use pmez
8	*aripiprazole/
9	(abilify* or aripiprazole* or abilitat* or discmelt or HSDB7320 or HSDB 7320 or OPC31 or OPC 31 or OPC14597 or OPC 14597).ti,ab.
10	8 or 9
11	*major depression/
12	*treatment resistant depression/
13	(depression* or depressive).ti,ab.
14	11 or 12 or 13
15	10 and 14
16	15 use oemezd
17	6 or 7 or 16
18	conference abstract.pt.
19	17 not 18
20	remove duplicates from 19
21	exp animals/
22	exp animal experimentation/ or exp animal experiment/
23	exp models animal/
24	nonhuman/
25	exp vertebrate/ or exp vertebrates/
26	animal.po.
27	or/21-26
28	exp humans/
29	exp human experimentation/ or exp human experiment/
30	human.po.
31	or/28-30
32	27 not 31
33	20 not 32

OTHER DATABASES	
PubMed	Same MeSH, keywords, limits, and study types used as per MEDLINE search, with appropriate syntax used.
Trial registries (Clinicaltrials.gov and others)	Same keywords, limits used as per MEDLINE search.

Grey Literature

Dates for Search:	February 2014
Keywords:	Abilify, aripiprazole, depression, depressive
Limits:	No date or language limits used

Relevant websites from the following sections of the CADTH grey literature checklist, "Grey matters: a practical tool for evidence-based searching" (http://www.cadth.ca/en/resources/finding-evidence-is/grey-matters) were searched:

- Health Technology Assessment Agencies
- Health Economics
- Clinical Practice Guidelines
- Drug and Device Regulatory Approvals
- Advisories and Warnings
- Drug Class Reviews
- Databases (free)
- Internet Search.

APPENDIX 3: EXCLUDED STUDIES

Reference	Reason for Exclusion
Fava et al.[38]	Study design not of interest
Clinical study report CN138-164[39]	Study design not of interest

APPENDIX 4: DETAILED OUTCOME DATA

TABLE 9: EFFICACY: SUMMARY OF CHANGE FROM BASELINE TO END OF PHASE C, EFFICACY SAMPLE (LAST OBSERVATION CARRIED FORWARD)

	Study 139			Study 163			Study 165		
Variable	Pbo	Arip	Between-group difference (95% CI) *P* value	Pbo	Arip	Between-group difference (95% CI) *P* value	Pbo	Arip	Between-group difference (95% CI) *P* value
Sample size[a]	N = 172	N = 181	(Arip – pbo)	N = 184	N = 185	(Arip – pbo)	N = 169	N = 174	(Arip – pbo)
SDS									
SDS–work/school	N = 130	N = 127		N = 136	N = 142		N = 126	N = 115	
At the end of phase B,[b] mean ± SE	4.51 ± 0.28	5.37 ± 0.26		4.55 ± 0.27	4.18 ± 0.28		5.44 ± 0.32	5.23 ± 0.34	
Change at the end of phase C, mean ± SE	–0.72 ± 0.26	–0.71 ± 0.24	0.01 (–0.62 to 0.64) *P* = 0.976	–0.40 ± 0.22	–0.54 ± 0.23	–0.14 (–0.69 to 0.41) *P* = 0.615	–0.67± 0.26	–0.75 ± 0.28	–0.08 (–0.67 to 0.51) *P* = 0.792
SDS–social life	N = 164	N = 167		N = 170	N = 181		N = 160	N = 160	
At the end of phase B,[b] mean ± SE	5.85 ± 0.23	6.00 ± 0.22		5.56 ± 0.21	5.42 ± 0.20		6.13 ± 0.24	5.90 ± 0.24	
Change at the end of phase, mean ± SE	–0.76 ± 0.22	–1.35 ± 0.20	**–0.59 (–1.11 to –0.06) *P* = 0.030**[c]	–0.61 ±0.20	–1.41± 0.20	**–0.81 (–1.31 to –0.30) *P* = 0.002**[c]	–0.65 ± 0.23	–1.18 ± 0.24	–0.53 (–1.06 to 0.00) *P* = 0.052
SDS–family life	N = 164	N = 167		N = 170	N = 181		N = 160	N = 160	
At the end of phase B,[b] mean ± SE	5.43 ± 0.22	5.64 ± 0.20		5.65 ± 0.20	5.30 ± 0.20		5.93 ± 0.24	5.72 ± 0.25	
Change at the end of phase, mean ± SE	–0.50 ± 0.21	–1.12 ± 0.19	**–0.62 (–1.12 to –0.11) *P* = 0.017**[c]	–0.96 ± 0.20	–1.77 ± 0.20	**–0.81 (–1.33 to –0.30) *P* = 0.002**[c]	–0.82 ± 0.23	–1.39 ± 0.24	**–0.57 (–1.10 to –0.03) *P* = 0.037**[c]
QIDS-SR total score									
At the end of phase B,[b] mean ± SE	[redacted]	[redacted]		12.33 ± 0.36	11.60 ± 0.36		12.81 ± 0.40	12.96 ±0.40	

	Study 139			Study 163			Study 165		
Variable	**Pbo**	**Arip**	**Between-group difference (95% CI) *P* value**	**Pbo**	**Arip**	**Between-group difference (95% CI) *P* value**	**Pbo**	**Arip**	**Between-group difference (95% CI) *P* value**
Sample size[a]	**N = 172**	**N = 181**	**(Arip – pbo)**	**N = 184**	**N = 185**	**(Arip – pbo)**	**N = 169**	**N = 174**	**(Arip – pbo)**
Change at the end of phase, mean ± SE	[redacted]	[redacted]	[redacted]	–1.80 ± 0.31	–2.30 ± 0.30	–0.50 (–1.29 to 0.29) *P* = 0.213	–2.13 ± 0.34	–2.83 ± 0.34	–0.70 (–1.48 to 0.08) *P* = 0.077
Q-LES-Q (short form)									
Satisfaction with medication (item 15)	[redacted]	[redacted]		[redacted]	[redacted]		N = 153	N = 146	
At the end of phase B,[b] mean ± SE	[redacted]	[redacted]		[redacted]	[redacted]		3.05 ± 0.09	3.11 ± 0.09	
Change at the end of phase, mean ± SE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	0.08 ± 0.09	0.20 ± 0.09	0.12 (–0.09 to 0.33) *P* = 0.262
Overall life satisfaction (item 16)	N = 161	N = 166		N = 170	N = 181		N = 161	N = 160	
At the end of phase B,[b] mean ± SE	[redacted]	[redacted]		[redacted]	[redacted]		2.55 ±0.08	2.63 ±0.08	
Change at the end of phase, mean ± SE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	0.32 ± 0.08	0.64 ± 0.08	**0.33 (0.13 to 0.52) *P* = 0.001[c]**

Arip = aripiprazole; CI = confidence interval; pbo = placebo; QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Report; Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire; SDS = Sheehan Disability Scale; SE=standard error.

[a] Sample size for all outcomes except SDS and Q-LES-Q.

[b] Baseline means baseline of phase C (or at end of phase B).

[c] Bolding indicates statistically significant.

Source: Study 139 Clinical Study Report,[11] Study 163 Clinical Study Report,[12] and Study 165 Clinical Study Report.[13]

TABLE 10: MONTGOMERY–ÅSBERG DEPRESSION RATING SCALE TOTAL SCORE — ADJUSTED MEAN CHANGE DURING PHASE C BY SUBGROUP

	Placebo		Aripiprazole		Treatment Comparison[a] Aripiprazole – Placebo	
	N	Mean ± SE	N	Mean ± SE	Mean difference (95% CI)	Interaction[b] test *P* value
Study 139						
By ADT						
Escitalopram	[redacted]					
Fluoxetine						
Paroxetine						
Sertraline						
Venlafaxine XR						
By MADRS response at end of phase B						
< 25% improvement from baseline						
≥ 25% improvement from baseline						
Study 163						
By ADT						
Escitalopram						
Fluoxetine						
Paroxetine						
Sertraline						
Venlafaxine XR						
By MADRS response at end of phase B						
< 25% improvement from baseline						
≥ 25% improvement from baseline						
Study 165						
ADT						
Escitalopram						
Fluoxetine						
Paroxetine						
Sertraline						
Venlafaxine XR						
By MADRS response at end of phase B						
< 25% improvement from baseline						
≥ 25% improvement from baseline						

ADT = antidepressant therapy; CI = confidence interval; MADRS= Montgomery–Åsberg Depression Rating Scale; SE = standard error; XR = extended release.

[a] ANCOVA model, with double-blind treatment as main effect and end of phase B assessment as covariate, is used for mean change from end of phase B comparisons. Means, standard errors, treatment differences between aripiprazole and placebo, and 95% CIs for the differences are based on ANCOVA model.

[b] ANCOVA model, with double-blind treatment and subgroup as main effects, end of phase B assessment as covariate, and subgroup-by-treatment as interaction effect.

TABLE 11: INCIDENCE OF TREATMENT-EMERGENT ADVERSE EVENTS OCCURRING IN AT LEAST 5% OF PATIENTS IN EITHER TREATMENT GROUP DURING PHASE C

	Number (%)					
	Study 139		Study 163		Study 165	
	Pbo	Arip	Pbo	Arip	Pbo	Arip
N of pts screened for AEs	176	182	190	189	172	176
Number of patients with ≥ 1 TEAE	110 (62.5)	149 (81.9)	120 (63.2)	154 (81.5)	118 (68.6)	142 (80.7)
Vision blurred	3 (1.7)	12 (6.6)	0	0	3 (1.7)	13 (7.4)
Diarrhea	10 (5.7)	6 (3.3)	0	0	13 (7.6)	10 (5.7)
Dry mouth	11 (6.3)	6 (3.3)	0	0	0	0
Nausea	9 (5.1)	5 (2.7)	8 (4.2)	10 (5.3)	10 (5.8)	7 (4.0)
Constipation			5 (2.6)	10 (5.3)	6 (3.5)	10 (5.7)
Fatigue	6 (3.4)	11 (6.0)	7 (3.7)	19 (10.1)	8 (4.7)	16 (9.1)
Upper respiratory tract infection	7 (4.0)	15 (8.2)	0	0	13 (7.6)	13 (7.4)
Akathisia	8 (4.5)	42 (23.1)	8 (4.2)	49 (25.9)	6 (3.5)	32 (18.2)
Headache	19 (10.8)	11 (6.0)	20 (10.5)	17 (9.0)	14 (8.1)	15 (8.5)
Somnolence	0	0	7 (3.7)	13 (6.9)	1 (0.6)	10 (5.7)
Tremor	0	0	5 (2.6)	12 (6.3)	0	0
Dizziness	0	0	0	0	5 (2.9)	9 (5.1)
Restlessness	6 (3.4)	26 (14.3)	1 (0.5)	18 (9.5)	6 (3.5)	22 (12.5)
Insomnia	4 (2.3)	14 (7.7)	3 (1.6)	14 (7.4)	9 (5.2)	15 (8.5)

AE = adverse event; arip = aripiprazole; pbo = placebo; pts = patients.
Note: Patients may have had more than one AE, but were counted in the overall total only once. Each patient was also counted at most once for a particular AE, even if the AE occurred more than once for the same patient.

TABLE 12: INCIDENCE OF SERIOUS ADVERSE EVENTS DURING PHASE C

	Number (%)					
	Study 139		Study 163		Study 165	
	Pbo	Arip	Pbo	Arip	Pbo	Arip
Number of patients screened for AEs	176	182	190	189	172	176
Number of patients with ≥ 1 SAE	3 (1.7)	2 (1.1)	0	1 (0.5)	1 (0.6)	1 (0.6)
Cellulitis, staphylococcal	0	1 (0.5)	0	0	0	0
Pneumonia	0	1 (0.5)	0	0	0	0
Cellulitis, other than staphylococcal	1 (0.6)	0	0	1 (0.5)	0	0
Staphylococcal abscess	1 (0.6)	0	0	0	0	0
Injury, poisoning and procedural complications	0	0	0	0	0	0
Contusion	1 (0.6)	0	0	0	0	0
Exostosis	1 (0.6)	0	0	0	0	0
Physical assault	1 (0.6)	0	0	0	0	0
Psychiatric disorders	0	0	0	0	0	1 (0.6)
Suicidal ideation	0	0	0	0	0	1 (0.6)
Vascular disorders	0	0	0	0	1 (0.6)	0
Arterial occlusive disease	0	0	0	0	1 (0.6)	0

AE = adverse event; arip = aripiprazole; pbo = placebo; SAE = serious adverse event.
Note: Patients may have had more than one SAE, but were counted in the overall total only once.

TABLE 13: ADVERSE EVENTS LEADING TO DISCONTINUATION OF TREATMENT OCCURRING DURING PHASE C THERAPY, SAFETY SAMPLE

	Study 139		Study 163		Study 165	
	Pbo	Arip	Pbo	Arip	Pbo	Arip
Number of patients screened for AEs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients with ≥ 1 AE leading to discontinuation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Akathisia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Somnolence	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Restless legs syndrome	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Coordination abnormal	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sedation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Psychiatric disorders	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Anxiety	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Insomnia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Libido decreased	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Suicidal ideation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Depression	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Fatigue	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Gait disturbance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Chest pain	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pain	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Eye disorders	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Vision blurred	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hyperhidrosis	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Rash	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Urticaria	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Alopecia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Vascular disorders	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hypertension	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hematochezia	[redacted]	[redacted]	[redacted]	[redacted]		
Nausea	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Balance disorder	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Disturbance in attention	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Arthralgia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Muscle twitching	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

AE = adverse event; arip = aripiprazole; pbo = placebo.
Note: Patients may have more than one AE, but were counted in the overall total only once. Each patient was also counted at most once for a particular AE, even if the AE occurred more than once for the same patient.

TABLE 14: INCIDENCE OF TREATMENT-EMERGENT EXTRAPYRAMIDAL SYMPTOM-RELATED ADVERSE EVENTS DURING PHASE C

	Study 139		Study 163		Study 165	
	Pbo	Arip	Pbo	Arip	Pbo	Arip
Number of patients screened for AEs						
Number of patients with ≥ 1 treatment-emergent EPS-related AE						
Akathisia						
Psychomotor hyperactivity						
Dystonic events						
Muscle spasms						
Dystonia						
Muscle contractions involuntary						
Muscle rigidity						
Tremor						
Extrapyramidal disorder						
Residual events						
Muscle twitching						
Cogwheel rigidity						
Dyskinetic events						
Dyskinesia						

AE = adverse event; arip = aripiprazole; EPS = extrapyramidal symptom; pbo = placebo.
Note: Patients may have more than one AE, but were counted in the overall total only once. Each patient was also counted at most once for a particular AE, even if the AE occurred more than once for the same patient.

TABLE 15: ARIPIPRAZOLE DOSE IN RANDOMIZED PHASE C

	Aripiprazole Dose (mg/d) at End Point[a]			
Dose by ADT	N (%)	Mean	Min	Max
Study 139				
All ADT	182 (100)	11.4	2.0	20.0
Escitalopram	55 (100)	10.9	2.0	20.0
Fluoxetine	26 (100)	10.1	2.0	15.0
Paroxetine[b]	18 (100)	9.8	2.0	15.0
Sertraline	36 (100)	13.9	2.0	20.0
Venlafaxine XR	47 (100)	11.6	2.0	20.0
Study 163				
All ADTs	[redacted]	[redacted]	[redacted]	[redacted]
Escitalopram	[redacted]	[redacted]	[redacted]	[redacted]
Fluoxetine	[redacted]	[redacted]	[redacted]	[redacted]
Paroxetine[b]	[redacted]	[redacted]	[redacted]	[redacted]
Sertraline	[redacted]	[redacted]	[redacted]	[redacted]
Venlafaxine XR	[redacted]	[redacted]	[redacted]	[redacted]
Study 165				
All ADT	[redacted]	[redacted]	[redacted]	[redacted]
Escitalopram	[redacted]	[redacted]	[redacted]	[redacted]
Fluoxetine	[redacted]	[redacted]	[redacted]	[redacted]
Paroxetine[b]	[redacted]	[redacted]	[redacted]	[redacted]
Sertraline	[redacted]	[redacted]	[redacted]	[redacted]
Venlafaxine XR	[redacted]	[redacted]	[redacted]	[redacted]

ADT = antidepressant therapy; XR = extended release.
[a] Reflects the last non-missing dose taken within the 6-week randomization phase.
[b] Paroxetine was allowed instead of paroxetine controlled release (CR) if the latter was not available due to distribution and manufacturing limitations.

TABLE 16: ANTIDEPRESSANT THERAPY DOSE IN RANDOMIZED PHASE C

	ATD Dose (mg/d) at End Point[a]							
	Placebo				Aripiprazole			
Study 139								
ADT	[redacted]							
Escitalopram	[redacted]							
Fluoxetine	[redacted]							
Paroxetine[b]	[redacted]							
Sertraline	[redacted]							
Venlafaxine XR	[redacted]							
Study 163								
Escitalopram	[redacted]							
Fluoxetine	[redacted]							
Paroxetine[b]	[redacted]							
Sertraline	[redacted]							
Venlafaxine XR	[redacted]							
Study 165								
Escitalopram	[redacted]							
Fluoxetine	[redacted]							
Paroxetine[b]	[redacted]							
Sertraline	[redacted]							
Venlafaxine XR	[redacted]							

ADT = antidepressant therapy; XR = extended release.
[a] Reflects the last non-missing dose taken within the 6-week randomization phase.
[b] Paroxetine was allowed instead of paroxetine controlled release (CR) if the latter was not available due to distribution and manufacturing limitations.

TABLE 17: SUMMARY OF CONCOMITANT MEDICATIONS FOR POTENTIAL TREATMENT OF EXTRAPYRAMIDAL SYMPTOMS DURING PHASE C

	Number (%) of Patients					
	Study 139		Study 163		Study 165	
ADT/Class of drug	Pbo	Arip	Pbo	Arip	Pbo	Arip
All ADT	[redacted]					
Any EPS medication[a]	[redacted]					
Escitalopram	[redacted]					
Any EPS medication	[redacted]					
Fluoxetine	[redacted]					
Any EPS medication	[redacted]					
Paroxetine	[redacted]					
Any EPS medication	[redacted]					
Sertraline	[redacted]					
Any EPS medication	[redacted]					
Venlafaxine XR	[redacted]					
Any EPS medication	[redacted]					

ADT = antidepressant therapy; arip = aripiprazole; EPS = extrapyramidal symptoms; pbo = placebo; XR = extended release.
[a] EPS medication included benztropine and propranolol.

APPENDIX 5: VALIDITY OF OUTCOME MEASURES

Aim

To summarize the validity of the outcome measures used in the included studies.

Findings

Outcome measures Hamilton Depression Rating Scale – 17 items (HAM-D17), Montgomery–Åsberg Depression Rating Scale (MADRS), Quick Inventory of Depressive Symptomology–Self-Report (QIDS-SR), Inventory of Depressive Symptomology–Self-Report (IDS-SR), Clinical Global Impression–Severity of Illness (CGI-S), Sheehan Disability Scale (SDS), and Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) are briefly summarized in Table 18.

TABLE 18: VALIDITY AND MINIMAL CLINICALLY IMPORTANT DIFFERENCE OF OUTCOME MEASURES

Instrument	Type	MCID	References
HAM-D17	HAM-D17 addresses both somatic and psychological symptoms of depression. The scale is clinician-rated, in which ratings are made on the basis of a clinical interview and additional available information, such as family report. Items are either rated on a 5-point scale (0 to 4 spectrum) or a 3-point scale (0 to 2 spectrum), in which increasing scores represent increasing severity of symptoms. Scores on the 17 items are summed to obtain a total score of 52, or 53 in some versions.	2 to 3[a]	Zimmerman et al. 2005[40] Bagby et al. 2004[41] Montgomery and Möller 2009[17]
MADRS	MADRS assesses depressive symptomology, particularly change in patients treated with antidepressants. This scale is clinician-rated and consists of 10 items. Each item is rated on a 0 to 6 scale, resulting in a maximum total score of 60 points, in which higher scores are indicative of greater depressive symptomology.	2	Zimmerman et al. 2004[35] Lam et al. 2005[36]
QIDS-SR	QIDS-SR is a 16-item tool derived from the Inventory of Depressive Symptomatology (IDS) that measures depressive symptom severity. The responses are converted from the 16 items into 9 DSM-IV symptom criterion domains. Each symptom item is scored on a scale of 0 to 3, with higher scores representing greater symptom severity. The total score ranges from 0 to 27.	Unspecified	Rush et al. 2003[42]
IDS-SR	IDS-SR is a 30-item tool that measures depressive symptom severity. The IDS-SR is scored by summing the responses to 28 of 30 items. Each symptom item is scored on a scale of 0 to 3, with higher scores representing greater symptom severity. The total score ranges from 0 to 84. It was suggested that IDS-SR was less sensitive to change than clinician-rated IDS.	Unspecified	Rush et al. 1986[43] Rush et al. 1996[44] Biggs et al. 2000[45] Rush et al. 2003[42] Corruble et al. 1999[31]
CGI-S	The CGI-S is a component of the CGI instrument and assesses the clinician's impression of the patient's illness severity; it is often used both before and after treatment. Scores range from 1 = "not ill at all" to 7 = "among the most extremely ill."	Unspecified	Guy 2000[46]
SDS	The SDS is a short, 3-item, self-reported measure developed to assess the degree to which symptoms of	Unspecified	Lam et al. 2005[36] Sheehan et al.

Instrument	Type	MCID	References
	depression, anxiety, panic, and phobia interfere with the patient's work, family, and social life. Each of the items is scored on a 0 to 10 scale, where 0 indicates no impairment, 1 to 3 mild impairment, 4 to 6 moderate impairment, 7 to 9 marked impairment, and 10 extreme impairment. The items may also be summed into a total measure of global impairment, ranging from 0 to 30 points.		1996[47]
Q-LES-Q-SF	Q-LES-Q-SF is a 16-item self-reported measure that assesses generic quality of life. Items are scored on a 1 to 5 scale, where higher scores indicate greater enjoyment or satisfaction achieved during the particular activity described in the item.	Unspecified	Lam et al. 2005[36] Endicott et al. 2003[48]

CGI = Clinical Global Impression; CGI-S = Clinical Global Impression–Severity of Illness; HAM-D17 = Hamilton Rating Scale for Depression – 17 items; IDS = Inventory of Depressive Symptomatology; IDS-SR = Inventory of Depressive Symptomatology–Self-Report; MADRS = Montgomery–Åsberg Depression Rating Scale; MCID = minimal clinically important difference; QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Report; Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire; SDS = Sheehan Disability Scale.

[a] Reported values are not MCIDs. For clinical trials, the National Institute for Health and Care Excellence (NICE) recommends a 3-point difference between drug and placebo groups as a criterion for clinical significance.[37] In a separate report, Montgomery and Möller suggested that a difference of 2 points between antidepressant and placebo might be clinically relevant.[17]

Hamilton Rating Scale for Depression

The HAM-D is the most frequently utilized outcome measure in clinical trials of MDD and is considered by many to be the standard for assessment of depression.[40] While numerous versions of this scale exist, the 17-item scale is the version most frequently used in efficacy trials.[40] The scale is clinician-rated; ratings are made on the basis of a clinical interview and additional available information, such as family report.[36] As a measure of the severity of depression symptoms, the HAM-D17 addresses both somatic and psychological symptoms of depression.[49] Items are either rated on a 5-point scale (0 to 4 spectrum) or a 3-point scale (0 to 2 spectrum), on which increasing scores represent increasing severity of symptoms.[41] Scores on the 17 items are summed to obtain a total score of 52, or 53 in some versions.[50] Since the number of response options varies between items, certain items contribute more to the total score than others.[41] The HAM-D17 scoring instructions indicate that a total score ranging from 0 to 7 indicates that the patient is in the normal range (no depression), a score ranging from 8 to 13 indicates "mild depression," 14 to 18 indicates "moderate depression," 19 to 22 indicates "severe depression," and a total score of 23 or greater indicates "very severe depression."[40]

While many of the psychometric properties of the HAM-D17 are adequate and consistently meet established criteria, some psychometric and conceptual flaws have also been identified.[41] Reliability coefficients for internal consistency as well as interrater and test-retest reliability are generally good for the overall scale, as are the internal reliability estimates for the individual items of the scale. Although numerous items have weak interrater and retest reliability at the item level, the use of structured interview guides may increase the reliability of individual items and the total scale.[41] The content validity of the HAM-D17 is poor, as there is only partial overlap between the content of this scale and the *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition (DSM-IV) symptom inclusion diagnostic criteria for MDD.[40] Some symptoms on the HAM-D17 are not official DSM-IV criteria and, while some such symptoms are recognized as associated with depression (e.g., psychic anxiety), the link to depression is more tenuous for other symptoms (e.g., loss of insight, hypochondriasis).[41] Conversely, important features of DSM-IV criteria for depression, such as concentration difficulties, feelings of worthlessness, and reverse vegetative symptoms, are either buried within complex items or not

captured at all.[41] The convergent validity of the HAM-D17 has been shown to be adequate, as this scale has moderate to high correlation with many other depression scales.[41] Similarly, the discriminant validity of this scale has been shown to be adequate.[41] Several meta-analyses have also found the HAM-D17 to be more sensitive to change (responsive) in patients' conditions compared with other depression scales, such as the Beck Depression Inventory.[51,52] However, the multi-dimensional nature of the HAM-D17 may somewhat reduce its sensitivity in detecting changes in depression severity over time.[53] For instance, the full HAM-D17 scale has been shown to be less sensitive than unidimensional subscales of its items.[54] Frequently used subscales include the Core (includes items related to core depressive symptoms [depressed mood, guilt, suicidal behaviours and ideation, work/activities, and psychomotor retardation]) and the Maier (includes core items in addition to items related to anxiety and agitation). Overall, some of the psychometric properties of the HAM-D17 are adequate, yet some inherent psychometric and conceptual flaws remain.[41]

Two clinically important outcomes on the HAM-D17 are frequently reported in efficacy trials: response, defined as a 50% reduction from baseline HAM-D17 total score; and remission, generally defined as a score of 7 or less on the HAM-D17 total score.[40] Since a consensus panel recommended in 1991 that the cut-off for remission be a score of ≤ 7,[55]10 this level has been widely adopted in clinical trials. However, more recent evidence has suggested that, based on a narrow definition of remission in DSM-IV, which requires an absence of clinically significant symptoms of depression, the optimal cut-off should be ≤ 2 on the HAM-D17 total score.[40] The level of ≤ 7 was found to be appropriate when using a broader definition of remission.[40] For clinical trials, the National Institute for Health and Care Excellence (NICE) recommended a 3-point difference between drug and placebo groups as a criterion for clinical significance,[18] although no justification for this figure was provided.[50] In the updated NICE guidelines,[56] there was no mention of what constituted a clinically significant difference. A separate report by Montgomery and Möller suggested a difference of two points between antidepressant and placebo might be clinically relevant,[17] although, similar to the NICE guidelines, it appears that this figure was opinion-based. Therefore, a formally derived minimal clinically important difference (MCID) or an evidence-based clinically significant difference for HAM-D17 were not identified.

Montgomery–Åsberg Depression Rating Scale

Next to the HAM-D17, the MADRS is the most commonly used outcome measure in antidepressant efficacy trials and has been used with increasing frequency during the past decade.[35] The main purpose of this scale is to assess depressive symptomology, particularly change in patients treated with antidepressants.[35] While the HAM-D17 includes items that address somatic symptoms, the MADRS focuses on the psychological symptoms of depression (e.g., sadness, tension, and pessimistic thoughts).[49] This scale is clinician-rated and consists of 10 items; each item is rated on a 0-6 scale, resulting in a maximum total score of 60 points, with higher scores indicative of greater depressive symptomology.[36] The MADRS scoring instructions indicate that a total score ranging from 0 to 6 indicates that the patient is in the normal range (no depression), a score ranging from 7 to 19 indicates "mild depression," 20 to 34 indicates "moderate depression," a score of 35 and greater indicates "severe depression," and a total score of 60 or greater indicates "very severe depression."[57] There is evidence that an improvement of two points or more on the MADRS is considered clinically relevant.[58]

The psychometric properties of the MADRS scale have been evaluated in numerous studies and compared with those of other scales, such as the HAM-D17. The MADRS has high internal consistency, slightly higher than that of the HAM-D17.[53] The clinician interrater reliability of this scale was also acceptable on individual items as well as the total score.[59] With respect to its content validity, most of the items are highly related to the core concept of depression. However, similar to the HAM-D17, not all

of the core criteria symptoms used in the DSM-IV are assessed by the MADRS, and therefore neither scale is completely adequate to define the severity of depression or remission.[53] There is a high degree of correlation between scores of the MADRS and other measures, such as the HAM-D17 and the HAM-D_6, showing good convergent validity.[49,53,59] The MADRS has also shown high ability to discriminate between various levels of depression severity.[53] Studies have repeatedly found the MADRS to have greater sensitivity to treatment-related change compared with the HAM-D17;[59-61] however, at least one study found its sensitivity to be lower than that of the HAM-D17.[62] This high capability of the MADRS to detect change in patients' conditions over time may be related to its more uniform structure compared with the HAM-D17.[63] Overall, the MADRS has been found to have sound psychometric properties and be at least comparable to, if not somewhat exceeding, the HAM-D17 on certain psychometric aspects.

No consensus has emerged regarding a cut-off value on the MADRS for defining remission in clinical trials.[64]18 Criterion scores to identify patients who have experienced remission have ranged from 6 through 12 in various trials.[65] However, one recent study that intended to establish an empirically-based cut-off for remission concluded that, based on a narrow definition of remission, the optimal MADRS cut-off was ≤ 4 points. On the basis of a less conservative definition of remission, the recommended cut-off was ≤ 9 points.[64] There is evidence that a MADRS score of < 10 is a valid cut-off for remission.[66]

Inventory of Depressive Symptomatology–Self-Report and Quick Inventory of Depressive Symptomatology–Self-Report

The IDS is a 30-item tool that measures depressive symptom severity. The IDS is available in both self-report (IDS-SR) and clinician-rated formats (IDS-C).[43] The 30 items include DSM-IV diagnostic criterion items for MDD. Both IDS-SR and IDS-C are scored by summing the responses to 28 of 30 items (i.e., only appetite and weight increase *or* appetite and weight decrease is scored for a given rating). Each symptom item is scored on a scale of 0 to 3, with higher scores representing greater symptom severity. The total score ranges from 0 to 84.[43] The IDS-SR has been validated in numerous studies and has been shown to measure depression in a manner consistent with the most widely used assessments.[44] Furthermore, the IDS-SR has been shown to have high internal consistency in patients with MDD (Cronbach's alpha ranging from 0.83 to 0.92).[45]

The QIDS is a 16-item tool, derived from the IDS, that measures depressive symptom severity. The QIDS is available in both self-report (QIDS-SR) and clinician-rated formats (QIDS-C).[42] The QIDS was constructed by selecting the IDS items that assess DSM-IV criterion diagnostic symptoms. The QIDS-SR has a recall period — the period during which patients are asked to rates their symptoms — of seven days. The responses are converted from the 16 items into 9 DSM-IV symptom criterion domains: sad mood; concentration; self-criticism; suicidal ideation; interest; energy/fatigue; sleep disturbance (initial, middle, and late insomnia or hypersomnia); decrease/increase in appetite/weight; and psychomotor agitation/retardation. Each symptom item is scored on a scale of 0 to 3, with higher scores representing greater symptom severity. The total score ranges from 0 to 27. The QIDS-SR scoring instructions indicate that a total score ranging from 0 to 5 indicates that the patient is in the normal range (no depression), a score ranging from 6 to 10 indicates "mild depression," 11 to 15 indicates "moderate depression," 16 to 20 indicates "severe depression," and a total score of 21 or greater indicates "very severe depression."[42] The QIDS-SR was found to have high internal consistency (Cronbach's alpha = 0.86), and scores were highly correlated with IDS-SR_{30} (0.96) and HAM-D24 (0.86) total scores. An MCID for either the IDS-SR and QIDS-SR has not been specified.

Clinical Global Impression–Severity of Illness and –Improvement

The CGI scale consists of three components: Severity of Illness (CGI-S), Improvement (CGI-I), and the Efficacy Index (CGI-E). Scores on the Severity of Illness subscale range from 1 = "not ill at all" to 7 = "among the most extremely ill." The Improvement subscale also goes from 1 = very much improved to 7 = very much worse. The Efficacy Index involves locating a rating on a matrix of therapeutic versus adverse events. Score range from 0 = marked improvement and no adverse events to 4 = unchanged or worse and adverse events outweigh therapeutic effects. The CGI instrument does not yield a global score, as each component of the CGI is rated separately. The distribution of scores and normality of CGI items were examined at the first and last week (eight weeks apart) in three clinical trials with a total of 175 patients with schizophrenia, depression, or anxiety disorders. Scores on the Improvement and Therapeutic Effects subscales were highly correlated (r = 0.90). However, there was only a moderate correlation(r = –0.47 to –0.66) between changes in the Severity of Illness and Improvement subscales. Severity of Illness was moderately correlated with adverse events rating. Test-retest reliability was calculated by correlating the ratings of each item at the first visit with all respective ratings at subsequent visits. These test-retest correlations were rather low: for Severity of Illness, reliability values ranged from 0.20 to 0.81. Another study found relatively good reliability scores for the CGI-S ratings (0.66 and 0.41 for physicians and nursing staff, respectively). An eight-week clinical trial involving 116 patients have shown that panic disorder, depression, anticipatory anxiety, and panic frequency each had positive significant relationships with clinician ratings of severity on the CGI scale. Although the CGI scale has been widely used in psychopharmacology trials, there have been only a few studies of its psychometric characteristics, and these studies differ widely in their assessment of the usefulness and reliability of the scale. It has been suggested that the positive qualities of the CGI scale could be enhanced by interrater training and more highly structured anchor points for each CGI item. Other studies suggested that the CGI is unreliable and that some items are inappropriate constructed and are of doubtful clinical significance.[46] An MCID has not been specified.

Sheehan Disability Scale

The SDS is a short, three-item self-reported measure developed to assess the degree to which symptoms of depression, anxiety, panic, and phobia interfere with the patient's work, family, and social life.[36] Each of the items is scored on an 11-point scale (0 to 10), on which 0 indicates no impairment, 1 to 3 mild impairment, 4 to 6 moderate impairment, 7 to 9 marked impairment, and 10 extreme impairment. Scores exceeding 5 points on any of the items are indicative of functional impairment and heightened risk of mental disorder.[36] The items may also be summed to a total measure of global impairment, ranging from 0 to 30 points.[36] This scale is a simple scale with only three questions and is regarded as similar to a global impression scale. There is some evidence that the SDS is a sensitive measure of disability for patients with psychiatric disorders in primary care. One study evaluated this scale in a sample of 1,001 psychiatric patients in primary care and found that an higher score (5 or more) was associated with an increased risk of psychiatric impairment.[47] Also, more than 80% of patients with a diagnosis of a mental disorder were shown to have an elevated SDS score.[47] An MCID has not been specified.

Quality of Life Enjoyment and Satisfaction Questionnaire

The Q-LES-Q short form (Q-LES-Q-SF) is a 16-item self-reported measure that assesses generic quality of life. This scale was first developed in a population of outpatients with depression and has since become a widely used for measuring quality of life in patients with mood and anxiety disorders.[36] Items are scored on a 5-point scale (1 to 5), in which higher scores indicate greater enjoyment or satisfaction achieved during the particular activity described in the item.[48] The scoring of the Q-LES-Q-SF involves summing the first 14 items to generate an overall general subscore, while "satisfaction with medication" and "overall life satisfaction" subscores are stand-alone items derived from items 15 and 16 of the Q-LES-Q-SF, respectively. The raw overall general score ranges from 14 to 70 and is converted to a percentage of the maximum score possible.[36,48] Evidence supports the strong psychometric properties of the Q-LES-Q-SF, as it has shown to be a reliable and valid (clinical assessment of quality of life with an internal control).[67] One study that assessed 57 adults with psychiatric diagnosis demonstrated a test-retest coefficient of 0.93 and significant correlations (0.41 to 0.81) between the majority of items and the total score and other measures used in the study.[67] Responsiveness parameters revealed that the Q-LES-Q-SF was an 80% sensitive and 100% specific measure.[67] An MCID has not been specified for the individual scale scores or summary score.

APPENDIX 6: SUMMARY OF OTHER STUDIES

Objective
To provide a summary of extension study 164.[34]

Methods
Study 164 was a 52-week, open-label, uncontrolled study that assessed the long-term safety and tolerability of aripiprazole as an adjunct to antidepressant therapy (ADT) in patients with major depressive disorder MDD and with an incomplete response to one or more ADT. *De novo* patients as well as patients who completed studies 139 or 163 (randomized to either aripiprazole or placebo) had the option to enter extension study 164. Patients from study 165 were not eligible, as the study was initiated later than the other studies and an adequate sample size had already been achieved for long-term safety assessment.

De novo patients had had an inadequate response to one to four ADT trials (baseline Montgomery–Åsberg Depression Rating Scale [MADRS] total score > 10). In addition to the ADTs permitted during studies 139 and 163, *de novo* patients could receive mirtazapine, bupropion, or duloxetine. Concomitant use of the following psychotropic drugs was prohibited: neuroleptics, anticonvulsants, antidepressants (other than continued antidepressant therapy), mood stabilizers, opioid analgesics, and stimulants and barbiturates (except for migraine). All patients, regardless of prior exposure to aripiprazole in double-blind studies, started open-label adjunctive treatment with aripiprazole (5 mg per day). Based on the clinical judgment of the investigator with respect to tolerability and therapeutic efficacy, dose adjustments were made within the range of 2 mg to 30 mg per day for patients receiving venlafaxine extended release, escitalopram, mirtazapine or sertraline, or within the range of 2 mg to 15 mg per day for patients receiving fluoxetine, paroxetine, duloxetine, or bupropion.

Rollover patients (i.e., those continuing on from studies 139 and 163) continued to receive the same final dose of ADT that was prescribed in the previous studies. *De novo* patients were permitted to use these same ADTS, but could also receive bupropion sustained release (300 mg to 400 mg per day), bupropion extended release (150 mg to 450 mg per day), bupropion (300 mg to 450 mg per day), duloxetine (40 mg to 60 mg per day), or mirtazapine (15 mg to 45 mg per day) (Table 19). During the open-label treatment, patients were not permitted to switch ADT. Dose adjustments of ADT were permitted within the recommended dose range but were not to be made within the same week as the aripiprazole dose adjustment.

The primary outcome was the assessment of safety of adjunctive aripiprazole. The secondary outcome was the maintenance of efficacy of adjunctive aripiprazole using the Clinical Global Impression–Severity of Illness (CGI-S) scale. The study characteristics are summarized in Table 19. No statistical testing was performed, and the sample size was not based on statistical power considerations.

TABLE 19: CHARACTERISTICS OF STUDY 164

Objective	To assess the long-term safety and tolerability of aripiprazole augmentation compared with adjunctive placebo to ADT in MDD patients with incomplete response to one or more ADT alone
Population	**The included patients** were men and women, aged 18 years and older, who met the DSM-IV-TR criteria for a major depressive episode. **Rollover patients (prior aripiprazole and prior placebo) from study 139 or study 163:** • completed randomized phase C (received study drug), or • completed 14 weeks (phase B plus phase C) on ADT alone ("responders" to the prospective ADT) and did not meet criteria for remission (defined as MADRS score ≤ 10) at the week 14 visit. • In the opinion of the investigator, had the potential to benefit from adjunctive aripiprazole. ***De novo* patients:** • diagnosis of a single or recurrent, non-psychotic episode of MDD, as defined by DSM-IV-TR, with a current depressive episode ≥ 8 weeks in duration • currently taking allowable ADT at an adequate dose for ≥ 6 weeks by the end of the screening phase and reported a history in the current depressive episode of an inadequate response to ≥ 1 and < 4 adequate ADTs • MADRS total score > 10 at the end of screening phase and, in the opinion of the investigator, had residual symptoms that may have benefited from pharmacologic modification. **All patients were excluded from the trial for the following:** • current Axis I diagnosis of delirium or dementia • amnestic or other cognitive disorder • schizophrenia or other • psychotic disorder • bipolar I or II disorder • eating disorder • clinically significant current Axis II diagnosis of borderline, antisocial, paranoid, schizoid, schizotypal, or histrionic personality disorder • suicide risk.
Interventions	**Rollover patients:** aripiprazole 5 mg/d (modifications permitted in the range of 2 to 30 mg/d to optimize therapeutic benefit) once daily for up to 52 weeks as adjunct to their assigned open-label ADT (venlafaxine XR 150 to 225 mg/d , escitalopram 10 to 20 mg/d, paroxetine CR 37.5 to 50 mg/d, paroxetine 20 to 40 mg/d [if paroxetine CR was not available], fluoxetine 20 to 40 mg/d, sertraline 100 to 150 mg/d) from either study 139 or study 163. ***De novo* and prior placebo patients:** aripiprazole 5 mg/d (modifications permitted in the range of 2 to 30 mg/d to optimize therapeutic benefit) once daily for up to 52 weeks as adjunct to their current ADT (venlafaxine XR 150 to 225 mg/d, escitalopram 10 to 20 mg/d, paroxetine 20 to 40 mg/d, paroxetine CR 37.5 to 50 mg/d, fluoxetine 20 to 40 mg/d, mirtazapine 15 to 45 mg/d, bupropion SR 300 to 400 mg/d twice daily, bupropion XL 150 to 450 mg/d, or sertraline or duloxetine 40 to 60 mg/d)
Outcomes	**Efficacy outcome:** • CGI-S at weeks 1, 2, 4, 6, 8, 14, 20, 26, 32, 38, 44, and 52 **Safety outcomes:** • adverse events • serious adverse events • withdrawals due to adverse events • body weight (weeks 26 and 52).
Design	Open-label, uncontrolled study

ADT = antidepressant therapy; CGI-S = Clinical Global Impressions–Severity of Illness; CR = controlled release; DSM-IV-TR = *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, Text Revision; MADRS = Montgomery–Åsberg Depression Rating Scale; MDD = major depressive disorder; XL = extended release.

Results

Table 20 summarizes the baseline characteristics of study participants. A total of 1,076 patients were enrolled in the study. Of these, 994 patients were included in the safety sample (n = 274, prior aripiprazole; n = 429, prior placebo; n = 291, *de novo*), and 975 patients were included in the efficacy sample (n = 267, prior aripiprazole; n = 423, prior placebo; n = 285, *de novo*). Completion rates were 27%, 35%, and 33% for the prior aripiprazole, prior placebo, and de novo patients, respectively. As seen in Table 20, the majority of patients were female, Caucasian, and had a body mass index > 30). Both CGI-S and MADRS scores were greater in *de novo* patients than in prior aripiprazole and prior placebo patients at baseline.

TABLE 20: BASELINE AND DEMOGRAPHIC CHARACTERISTICS OF SAFETY SAMPLE OF STUDY 164

	Prior aripiprazole (n = 274)	Prior placebo (n = 429)	*De novo* (n = 291)
Age, mean (SD)	[redacted]	[redacted]	[redacted]
Female, n (%)	[redacted]	[redacted]	[redacted]
Race			
White	[redacted]	[redacted]	[redacted]
Black	[redacted]	[redacted]	[redacted]
Asian	[redacted]	[redacted]	[redacted]
Other	[redacted]	[redacted]	[redacted]
Weight (kg), mean (SD)	[redacted]	[redacted]	[redacted]
BMI (kg/m^2), mean (SD)	[redacted]	[redacted]	[redacted]
CGI-S score, mean (SD)	[redacted]	[redacted]	[redacted]
MADRS total score, mean (SD)	[redacted]	[redacted]	[redacted]

BMI = body mass index; CGI-S = Clinical Global Impression–Severity of Illness; MADRS = Montgomery–Åsberg Depression Rating Scale; SD = standard deviation.
Source: Study 164 Clinical Study Report.[39]

As seen in Table 21, baseline CGI-S scores were highest in *de novo* patients and lowest in the prior aripiprazole patients. Changes from baseline were also highest in *de novo* patients and lowest in the prior aripiprazole patients at all time points. Larger decreases from baseline were generally seen in all patients as time progressed.

TABLE 21: CHANGE FROM BASELINE IN CLINICAL GLOBAL IMPRESSION–SEVERITY OF ILLNESS SCORE IN STUDY 164

	Prior aripiprazole (n = 267)	Prior placebo (n = 423)	*De novo* (n = 285)
CGI-S score, mean (SE)			
Baseline	[redacted]	[redacted]	[redacted]
Change from baseline at week 6	[redacted]	[redacted]	[redacted]
Change from baseline at week 26	[redacted]	[redacted]	[redacted]
Change from baseline at week 52	[redacted]	[redacted]	[redacted]

CGI-S = Clinical Global Impression–Severity of Illness; SE = standard error.
Source: Study 164 Clinical Study Report.[39]

As seen in Table 22, change from baseline in CGI-S score at week 52 was greatest in the *de novo* group, and lowest in the prior aripiprazole group, regardless of ADT.

TABLE 22: CHANGE FROM BASELINE IN CLINICAL GLOBAL IMPRESSION—SEVERITY OF ILLNESS SCORE AT WEEK 52 BY ANTIDEPRESSANT THERAPY (OBSERVED RESULTS) IN STUDY 164

	Prior Aripiprazole (n = 267)	Prior Placebo (n = 423)	*De novo* (n = 285)
ADT	**CGI-S score at week 52, mean (SE)**		
Escitalopram			
Venlafaxine			
Sertraline			
Fluoxetine			
Paroxetine CR or paroxetine			
Bupropion XL or bupropion SR			
Duloxetine			
Mirtazapine			

ADT= antidepressant therapy; CR = controlled release; CGI-S = Clinical Global Impression–Severity of Illness; SE = standard error; SR = sustained release; XL = extended release.
Source: Study 164 Clinical Study Report.[39]

As seen in Table 23, adverse events (AEs) that occurred at an incidence of 5% or more were most frequent among *de novo* patients (94.5%) compared with prior placebo patients (92.1%) and prior aripiprazole patients (79.9%). The most common AEs were akathisia, fatigue, and weight increase. The proportion of patients who experienced serious adverse events (SAEs) was highest among prior placebo patients (4.9%) compared with *de novo* patients (4.1%) and aripiprazole patients (1.8%). SAEs that occurred in two or more patients were suicidal ideation, intentional overdose, depression, chest pain, cholecystitis, and myocardial infarction. All other SAEs occurred in one patient each. The proportion of patients who withdrew from the study due to AEs was greatest among *de novo* patients (25.4%) compared with prior placebo patients (20.3%) and prior aripiprazole patients (19.7%). The most common reasons for withdrawals due to AEs were weight increase (3.2%) and akathisia (3.1%). Other AEs that led to discontinuation included suicidal ideation (three patients), intentional overdose (one patient), self-injurious ideation (one patient), tardive dyskinesia (two patients), and dyskinesia (three patients). Increases in body weight were greatest in the prior placebo patients at week 26 and week 52 compared with *de novo* and prior aripiprazole patients. There were no reported deaths in the study.

TABLE 23: SAFETY RESULTS IN STUDY 164

	Prior Aripiprazole (n = 274)	Prior Placebo (n = 429)	*De novo* (n = 291)
Adverse events,[a] n (%)	219 (79.9)	[redacted]	[redacted]
Akathisia	34 (12.4)	[redacted]	[redacted]
Fatigue	22 (8.0)	[redacted]	[redacted]
Weight increase	30 (10.9)	[redacted]	[redacted]
Restlessness	19 (6.9)	[redacted]	[redacted]
Insomnia	11 (4.0)	[redacted]	[redacted]
Somnolence	20 (7.3)	[redacted]	[redacted]
Headache	7 (2.6)	[redacted]	[redacted]
Upper respiratory tract infection	22 (8.0)	[redacted]	[redacted]
Nausea	7 (2.6)	[redacted]	[redacted]
Dizziness	9 (3.3)	[redacted]	[redacted]
SAEs	5 (1.8)	[redacted]	[redacted]
WDAEs	54 (19.7)	[redacted]	[redacted]

SAE = serious adverse event; WDAE = withdrawal due to adverse event.
[a] Occurring in at least 5% of patients.
Note: Reporting of adverse events for prior aripiprazole patients includes any adverse events that may have occurred with aripiprazole treatment during the previous double-blind study period.
Source: Study 164 Clinical Study Report.[39]

As seen in Table 24, changes in weight gradually increased over time among all patients at week 26 and week 52. Increases in body weight were greatest in the prior placebo patients, and lowest in the *de novo* patients.

TABLE 24: CHANGE IN BODY WEIGHT FROM BASELINE IN STUDY164

Body weight (kg), mean (SE)			
Baseline	[redacted]	[redacted]	[redacted]
Change from baseline at week 26	[redacted]	[redacted]	[redacted]
Change from baseline at week 52	[redacted]	[redacted]	[redacted]

SE = standard error.
Source: Study 164 Clinical Study Report.[39]

Interpretation

Baseline characteristics between treatment groups were generally similar between rollover (both aripiprazole and placebo) and *de novo* patients, with the exception of higher CGI-S and MADRS mean scores among *de novo* patients, indicating a greater disease severity at baseline. Baseline demographics indicated the majority of patients were obese (body mass index ≥ 30). It should be noted that the inclusion criteria for the *de novo* patients was inconsistent with studies 139 and 163, in which those with an inadequate response to more than three ADTs were excluded.

The efficacy results revealed that disease severity, as measured by the CGI-S, gradually decreased over time in all patients. Greater decreases were seen in the *de novo* patients. Analyses by type of ADT were performed, although given the small and varying sample sizes between groups at week 52, it remains uncertain whether differences in exposure to different ADTs had an effect on the efficacy results. Because of a weak study design (i.e., non-randomized open-label with no control) and uncertainty concerning whether the sample size was adequately powered, results should be interpreted with caution. Although, the primary objective was to measure long-term safety and tolerability, there was no justification why the CGI-S scale was used to measure efficacy instead of the HAM-D17, which is the most frequently used outcome measure in clinical trials of MDD and is considered by many to be the standard for assessment of depression.[40]

Over the 52-week duration, the most common AEs associated with aripiprazole were akathisia, fatigue, and weight increase. Overall, the incidence of AEs, SAEs, and withdrawals due to AEs appeared to be lower among prior aripiprazole patients, which may suggest greater tolerability after prior exposure to treatment. However, without a control group, it is difficult to interpret safety results or to attribute AEs solely to adjunctive aripiprazole treatment, as some AEs may have resulted from the long-term use of the ADTs. All patients, regardless of whether they received adjunctive aripiprazole in the prior double-blind studies, were retitrated to aripiprazole (5 mg per day) at entry into open-label treatment, which may have had an impact on the patients receiving a stable dose in the previous studies. Furthermore, patients received a variety of different ADTs (based on investigator judgment), and the relative benefit of each ADT was not assessed.

Body weight gradually increased in all patients, with the greatest increases seen in the prior aripiprazole group, and lowest increase in the *de novo* patients. For reasons previously stated, it is difficult to interpret and explain differences between groups.

Conclusion

Given the limitations of the study design, long-term efficacy and safety of aripiprazole remain uncertain. Common AEs included akathisia, fatigue, and weight increase over time with treatment. Although disease severity, measure with the CGI-S, appeared to decrease over time, and aripiprazole appeared to be generally well tolerated over 52 weeks, a rigorously designed double-blind, placebo-controlled study is necessary to establish long-term efficacy and safety as adjunctive treatment to ADT in MDD.

APPENDIX 7: SUMMARY OF COMPARATORS

Objective

To summarize evidence for the efficacy and safety of adjunctive atypical antipsychotics (AAPs) for the treatment of major depressive disorder (MDD), including a manufacturer–submitted network meta-analysis.

Findings

A recent meta-analysis by Spielmans et al.[14] included randomized controlled trials (RCTs) comparing adjunctive antipsychotic medication to placebo for treatment-resistant depression in adults. The literature search identified 14 short-term trials of aripiprazole (three trials), olanzapine (five trials), quetiapine (three trials), and risperidone (three trials). Trials were included if they were randomized, acute-phase (i.e., not for relapse prevention or maintenance treatment), placebo-controlled trials. Open-label studies were not excluded from the analysis. Patients were required to have a diagnosis of treatment-resistant depression, defined as a diagnosis with MDD and an inadequate response to at least one course of antidepressant therapy (ADT), prior to enrolment in the study. It should be noted that, aside from aripiprazole, none of the AAP comparators have a Health Canada–approved indication for the adjunctive treatment of MDD. A random effects model was used to pool estimates while incorporating potential heterogeneity. Odds ratios were calculated for categorical measures and were weighted by inverse variance to provide a pooled effect size estimate. For continuous outcomes, effect sizes were computed from means and standard deviations when possible.

A summary of the included study characteristics can be found in Table 25. All comparators, with the exception of olanzapine (fluoxetine, nortriptyline, or venlafaxine only) used various selective serotonin reuptake inhibitors (SSRIs) or serotonin-norepinephrine reuptake inhibitors (SNRIs) ADTs. The mean age ranged from 42.0 to 46.1 years, the majority of patients were female, and the duration ranged from 4 weeks to 12 weeks. As seen in Table 26, the definitions of response and remission varied across studies.

TABLE 25: SUMMARY OF BASELINE CHARACTERISTICS OF INCLUDED STUDIES IN SPIELMANS ET AL. (2013) META-ANALYSIS

Study	ADT	Daily dosage at end point	Total N	Mean age (years)	Female (%)	Duration (weeks)	Prior failed trials
Aripiprazole							
Berman (2007)	Various	Flexible, mean = 11.8 mg	353	45.4	62.8	6	1 to 3 historical, 1 prospective
Berman (2009)	Various	Flexible, mean = 10.7 mg	343	45.3	73.1	6	1 to 3 historical, 1 prospective
Marcus (2008)	Various	Flexible, mean = 11.0 mg	369	44.5	66.7	6	1 to 3 historical, 1 prospective
Quetiapine							
Bauer (2009)	Various	Fixed, 150 or 300 mg	487	45.4	67.6	6	1 historical
El-Khalili (2010)	Various	Fixed, 150 or 300 mg	432	45.5	72.5	6	1 historical
McIntyre (2007)	Various	Flexible, mean = 182 mg	58	44.5	62.0	8	1 trial
Olanzapine							
Corya (2006)	Fluoxetine or venlafaxine	Fixed; olanzapine 6 mg/fluoxetine 25 mg, olanzapine 6 mg/fluoxetine 50 mg, olanzapine 12 mg/fluoxetine 25 mg, or olanzapine 12 mg/fluoxetine 50 mg	344	45.7	72.5	12	1 historical, 1 prospective
Shelton (2001)	Fluoxetine	Flexible, mean modal dose = olanzapine 13.5 mg/fluoxetine 52 mg	20	42.0	75	8	2 historical and 1 prospective
Shelton (2005)	Fluoxetine or nortriptyline	Flexible, mean modal dose = olanzapine 8.5 mg/fluoxetine 35.6 mg	356	42.0	69.4	8	1 historical, 1 prospective
Thase 1 (2007)	Fluoxetine	Fixed; olanzapine 6 mg/fluoxetine 50 mg, olanzapine 12 mg/fluoxetine 50 mg, or olanzapine 18 mg/fluoxetine 50 mg	203	44.1	60.2	8	1 historical, 1 prospective

Study	ADT	Daily dosage at end point	Total N	Mean age (years)	Female (%)	Duration (weeks)	Prior failed trials
Thase 2 (2007)	Fluoxetine	Fixed; olanzapine 6 mg/fluoxetine 50 mg, olanzapine 12 mg/fluoxetine 50 mg, or olanzapine 18 mg/ fluoxetine 50 mg	198	44.9	68.0	8	1 historical, 1 prospective
Risperidone							
Keitner (2009)	Various	Flexible, mean = 1.6 mg	95	45.2	56.7	4	1 prospective
Mahmoud (2007)	Various	Flexible, mean = NR, 1 or 2 mg permitted	268	46.1	73.5	6	1 prospective
Reeves (2008)	Various	Flexible, mean = 1.17 mg	23	44.0	69.6	8	1 prospective

ADT= antidepressant therapy; NR= not reported.

TABLE 26: DEFINITION OF OUTCOMES REPORTED IN SPIELMANS ET AL. (2013) META-ANALYSIS

Study	Response	Remission
	Definition	
Bauer (2009)	≥ 50% decrease in baseline MADRS	MADRS ≤ 8
Berman (2007)	≥ 50% decrease in baseline MADRS	MADRS ≤ 10 and ≥ 50% decrease in baseline MADRS
Berman (2009)	≥ 50% decrease in baseline MADRS	MADRS ≤ 10 and ≥ 50% decrease in baseline MADRS
Corya (2006)	≥ 50% decrease in baseline MADRS	MADRS ≤ 8
El-Khalili (2010)	≥ 50% decrease in baseline MADRS	MADRS ≤ 8
Keitner (2009)	≥ 50% decrease in baseline MADRS	MADRS ≤ 10
Mahmoud (2007)	≥ 50% decrease in HAM-D17	HAM-D17 ≤ 7
Marcus (2008)	≥ 50% decrease in baseline MADRS	MADRS ≤ 10 and ≥ 50% decrease in baseline MADRS
McIntyre (2007)	≥ 50% decrease in HAM-D17	HAM-D17 ≤ 7
Reeves (2008)	NA	
Shelton (2001)	≥ 50% decrease in baseline MADRS	MADRS ≤ 8
Shelton (2005)	≥ 50% decrease in baseline MADRS	MADRS ≤ 8
Thase 1 (2007)	≥ 50% decrease in baseline MADRS	MADRS ≤ 10
Thase2 (2007)	≥50% decrease in baseline MADRS	MADRS ≤ 10

HAM-D17 = Hamilton Rating Scale for Depression – 17items; MADRS = Montgomery–Åsberg Depression Rating Scale; NA = not applicable.
Note: Although study duration was provided, the timing of study end points was not indicated in the publication.

Remission and response rates were reported for each treatment, while safety outcomes of interest such as akathisia and weight gain were reported for two (aripiprazole and olanzapine) and three (aripiprazole, olanzapine, and quetiapine) treatments, respectively (Table 27).

As seen in Table 27, all four treatments had statistically significant effects on remission between 4 and 12 weeks. Effect sizes for remission were largest with adjunctive therapy with risperidone (odds ratio, 2.37, 95% confidence interval [CI], 1.41 to 4.30), although these results are based on a smaller sample size (n = 363) compared with the other treatments. All drugs, with the exception of olanzapine, also had statistically significant effects on response rates. The effect size for response was greatest with adjunctive aripiprazole (odds ratio, 2.07, 95% CI, 1.58 to 2.72). Of the two treatments that analyzed akathisia, the effect size was greatest with adjunctive aripiprazole (odds ratio, 7.47, 95% CI, 5.07 to 11.0). Of the three treatments that analyzed weight gain, the effect size was greatest with adjunctive olanzapine (odds ratio, 16.28, 95% CI, 7.02 to 37.76). In addition, weight gain was also analyzed as a continuous outcome. The mean weight gain in trials of adjunctive aripiprazole, quetiapine, and risperidone was approximately 1 kg, while the mean weight gain resulting from adjunctive olanzapine was 4.20 kg (95% CI, 3.79 to 4.61 kg).

TABLE 27: SUMMARY OF EFFICACY AND SAFETY RESULTS FROM SPIELMANS ET AL. (2013) META-ANALYSIS

	Number of studies	Number of patients	OR (95% CI)	I^2 (95% CI)	NNT/NNH (95% CI)
Remission					
Aripiprazole	3	1,065	2.01 (1.48 to 2.73)	0% (0% to 38.8%)	9 (6 to 18)
Quetiapine	3	977	1.79 (1.33 to 2.42)	0% (0% to 50.5%)	9 (6 to 19)
Olanzapine	5	1,121	1.42 (1.01 to 2.0)	15.2% (0% to 82.4%)	19 (9 to 713)
Risperidone	2	363	2.37 (1.31 to 4.30)	0% (0% to 79.4%)	9 (5 to 35)
Response					
Aripiprazole	3	1,065	2.07 (1.58 to 2.72)	0% (0% to 87.0%)	7 (5 to 12)
Quetiapine	3	977	1.53 (1.17 to 2.0)	0% (0% to 73. 7%)	10 (6 to 26)
Olanzapine	5	1,121	1.30 (0.87 to 1.93)	50. 8% (0% to 81.9%)	17 (NNH 34; NNT 7)[a]
Risperidone	2	363	1.83 (1.16 to 2.88)	0% (0% to 86.5%)	8 (5 to 33)
Akathisia					
Aripiprazole	3	1,085	7.47 (5.07 to 11.0)	0% (0% to 87.2%)	4 (3 to 6)
Quetiapine	NA				
Olanzapine	4	1,079	1.48 (0.96 to 2.30)	5.4% (0% to 85.5%)	28 (NNH 11; NNT 321)[a]
Risperidone	NA				
Weight Gain (dichotomous)					
Aripiprazole[b]	3	1,085	5.91 (2.14 to 16.29)	0% (0% to 63.5%)	29 (10 to 119)
Quetiapine[c]	3	968	2.86 (1.11 to 7.37)	0% (0% to 78.6%)	37 (12 to 594)
Olanzapine[c]	4	1,121	16.28 (7.02 to 37.76)	0% (0% to 47.8%)	9 (5 to 20)
Risperidone	NA				
Weight Gain, kg (continuous)					
	Number of studies	**Number of patients**	**Mean (95% CI)**	**I^2 (95% CI)**	**NNT/NNH (95% CI)**
Aripiprazole	3	1,085	1.05 (0.35 to 1.74)	83.38% (49.64% to 94.51%)	NA
Quetiapine	3	968	0.94 (0.62 to 1.26)	0% (0% to 79.80%)	NA
Olanzapine	4	1,121	4.20 (3.79 to 4.61)	9.80% (0% to 86.21%)	NA
Risperidone	3	NR	1.26 (NR)	42.95% (0% to 82.87%	NA

CI = confidence interval; NA = not applicable; NNT = number needed to treat; NNH = number needed to harm; NR = not reported; OR = odds ratio.

[a] The 95% CI included the possibility of both treatment-related benefit and treatment-related harm.

[b] Weight gain of ≥ 7%.

[c] Weight gain of ≥ 10%.

Discussion

The efficacy results suggested that all included AAPs were more efficacious in terms of response and remission than adjunctive placebo. The response outcome for olanzapine was not statistically significant, as results among the five trials were moderately heterogeneous (I^2 = 50.8%). With the exception of the risperidone trials, which did not assess safety outcomes, adjunctive treatment with aripiprazole or olanzapine demonstrated substantial risk of akathisia, and adjunctive treatment with aripiprazole, quetiapine or olanzapine demonstrated a significant risk of weight gain. The greater risk of weight gain with olanzapine is supported by higher mean weight gain in the continuous outcome analysis.

Three other meta-analyses[32,33,68] comparing adjunctive AAPs for the treatment of MDD were performed between 2007 and 2010. Similar to Spielmans et al.,[14] Papakostas et al.[32] and Nelson and Papakostas[33] analyzed efficacy in terms of dichotomous response and remission outcomes. Safety in Papakostas et al.[32] and Nelson and Papakostas was assessed only by examining dropout rates due to adverse events. Results in Spielmans et al. were generally similar to those in Papakostas et al. and Nelson and Papakostas. The greatest differences between Spielmans et al. and Nelson and Papakostas were in regards to the olanzapine results for remission, which demonstrated a slightly lower odds ratios favouring olanzapine (1.42 versus 1.83). Although the same olanzapine studies were included in both analyses, Spielmans et al. used a more restrictive definition, whereas Nelson and Papakostas used the definitions provided by the authors. Specifically, some olanzapine trials defined remission as MADRS score of 8 or less at two consecutive visits during the study, even if these two consecutive visits did not necessarily occur at study end point. Spielmans et al. calculated the number of participants in remission as those who achieved interim remission minus the number of patients who subsequently relapsed. The other key difference compared with Nelson and Papakostas was the treatment effect of olanzapine for response. Spielmans et al. included all adjunctive placebo comparison groups and used a random effects analysis, while Nelson and Papakostas excluded one comparison group from two olanzapine trials and used a fixed-effects model. As a result, Nelson and Papakostas estimated a significant treatment effect for olanzapine for response, while Spielmans et al. did not. Despite some differences in methodology, results of Spielmans et al. were generally similar and yielded the same conclusions as Komossa et al.[68]

Critical Appraisal

The systematic review and meta-analysis was reported according to preferred reporting items for systematic reviews and meta-analyses guidelines using a specified a priori design. There was duplicate study selection and data extraction. A comprehensive literature search was performed including unpublished data from manufacturer's online clinical trial registries and new drug applications to the US Food and Drug Administration. In this specific publication, a list of both included and excluded studies was provided, although data pertaining to patient baseline characteristics of the included studies were limited. All studies included adult patients with a similar mean age, who were predominantly Caucasian females. Further details of baseline characteristics in 11 of the 14 included studies are provided in Table 2828 for the critical appraisal of the network meta-analysis (NMA) provided by the manufacturer. Results revealed that patient populations were not robustly matched across studies.

Overall, there was a lack of clarity regarding the quality of the included studies, as only three studies clearly described adequate sequence generation, the similarity of placebo and treatment was uncertain in all but one study, and raters were blinded only in one study. A formal analysis for publication bias was not performed given the small number of trials for each drug; hence, publication bias may have slightly enhanced the overall effect size on depression measures.

The interpretation of the pooled efficacy and safety results is limited, given the differences in definition of treatment-resistant depression, in definition of response and remission, in doses of comparators, and in timing of outcome assessments (Table 26). The authors indicated that the process by which diagnoses were made was described clearly only in six trials. It remains uncertain whether the unclear or inadequate randomization of the double-blind trials may have affected the study results. Pooled risperidone results for remission and response are based solely on two studies with a small sample size. These results should be interpreted with caution, as the authors highlight discrepancies in data reporting in the larger risperidone study (69% of the total participants).

Moreover, the noticeably greater risk of weight gain with olanzapine may be understated, as the olanzapine studies measured weight gain using a more conservative cut-off of 10%, while the aripiprazole and quetiapine studies measured weight gain greater than 7%.

Summary

Overall, the available evidence suggests that aripiprazole, quetiapine, olanzapine, and risperidone as adjunctive treatment for MDD are efficacious in regards to remission and response. Based on the results of the meta-analysis, there appears to be an increased risk of akathisia with aripiprazole and olanzapine and an increased risk of weight gain with aripiprazole, quetiapine, and olanzapine. Given the limitations of the study design and data-reporting methods of the included studies, the results should be interpreted with caution.

Summary of Network Meta-analysis

Rationale

The manufacturer indicated that the submitted systematic review and NMA were undertaken because the comparative efficacy between the AAPs as adjunctive treatment to ADT in patients with MDD remains to be established in the absence of direct head-to-head trials. Comparative data were needed in order to inform the economic analysis.

Methods

Eligibility criteria

The inclusion criteria for eligibility of trials in the NMA consisted of the following:

Population	[redacted]
Intervention	[redacted] • [redacted]
Comparator	[redacted]
Outcomes	[redacted] • [redacted][a] • [redacted][a]
Study design	[redacted]

[redacted]
[a] [redacted]

[redacted]

Network Meta-analysis and Systematic Review

[redacted]

[redacted]

Results

FIGURE 3: NETWORK OF INCLUDED RANDOMIZED CONTROLLED TRIALS

Figure 3 contained confidential information and was removed at the request of the manufacturer.

Study and Patient Characteristics

[redacted] (Figure 3). [redacted] Spielmans et al.[14] [redacted] Table 28.

[redacted]

[redacted]

[redacted]

TABLE 28: PATIENT BASELINE CHARACTERISTICS OF INCLUDED TRIALS IN THE NETWORK META-ANALYSIS

Study	Mean (SD) number of MDEs	Mean (SD) / median (range) duration of current MDE	Number of prior adequate ADT trials in current MDE	MADRS score (SD)	HAM-D17 score (SD)
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					
[redacted]					

[redacted]

Source: Abilify cost-effectiveness analysis report.[16]

[redacted]
Table 26.

Results of the Network Meta-Analysis

[redacted]
Table 28, [redacted]

[redacted]
Table 29, [redacted]

TABLE 29: ODDS RATIOS FOR REMISSION AND RESPONSE FOR COMPARATOR ATYPICAL ANTIPSYCHOTICS VERSUS ARIPIPRAZOLE FROM THE NETWORK META-ANALYSIS (RANDOM EFFECTS MODEL)

AAP	N	OR vs. aripiprazole	Lower 95% CrI	Upper 95% CrI
[redacted]				
[redacted]				
[redacted]				
[redacted]				
[redacted]				
[redacted]				
[redacted]				
[redacted]				

[redacted]

Critical Appraisal of Network Meta-Analysis

[redacted]
Table 30.

Strengths

[redacted]

[redacted]

[redacted]

Limitations

[redacted] Spielmans et al.[14] [redacted]

[redacted]

[redacted]

Summary

[redacted] The NMA methodology was appropriate though was limited by the heterogeneity between the studies and patients included in the analysis.

TABLE 30: APPRAISAL OF NETWORK META-ANALYSIS USING ISPOR CRITERIA

ISPOR Checklist Item		Details and Comments
1.	Are the rationale for the study and the objectives stated clearly?	[redacted]
2.	Does the methods section include the following? • Eligibility criteria • Information sources • Search strategy • Study selection process • Data extraction • Validity of individual studies	[redacted]
3.	Are the outcome measures described?	[redacted]
4.	Is there a description of methods for analysis/synthesis of evidence? • Description of analyses methods/models • Handling of potential bias/inconsistency • Analysis framework	[redacted]
5.	Are sensitivity analyses presented?	[redacted]
6.	Do the results include a summary of the studies included in the network of evidence? • Individual study data? • Network of studies?	[redacted]
7.	Does the study describe an assessment of model fit?	[redacted]
8.	Are the results of the evidence synthesis presented clearly?	[redacted]
9.	Sensitivity/scenario analyses	[redacted]

ISPOR = International Society of Pharmacoeconomics and Outcomes Research; MDD = major depressive disorder; NMA = network meta-analysis; RCT = randomized controlled trial.

REFERENCES

1. Patten SB, Kennedy SH, Lam RW, O'Donovan C, Filteau MJ, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) clinical guidelines for the management of major depressive disorder in adults. I. Classification, burden and principles of management. J Affect Disord. 2009 Oct;117 Suppl 1:S5-14.
2. Public Health Agency of Canada. What is depression? [Internet]. Ottawa: The Agency; 2013. [cited 2014 Feb 20]. Available from: http://www.phac-aspc.gc.ca/cd-mc/mi-mm/depression-eng.php
3. Gravel R, Beland Y. The Canadian Community Health Survey: mental health and well-being. Can J Psychiatry. 2005 Sep;50(10):573-9.
4. Puyat JH, Marhin WW, Etches D. Estimating the prevalence of depression from EMRs. Can Fam Physician [Internet]. 2013 [cited 2014 Feb 14];54(4):445. Available from: http://www.cfp.ca/content/59/4/445.full.pdf
5. Patten SB, Wang JL, Williams JV, Currie S, Beck CA, Maxwell CJ, et al. Descriptive epidemiology of major depression in Canada. Can J Psychiatry. 2006 Feb;51(2):84-90.
6. World Health Organization. The world health report 2001 - mental health: new understanding, new hope. Geneva: The Organization; 2001.
7. Berndt ER, Koran LM, Finkelstein SN, Gelenberg AJ, Kornstein SG, Miller IM, et al. Lost human capital from early-onset chronic depression. Am J Psychiatry [Internet]. 2000 Jun [cited 2014 Apr 29];157(6):940-7. Available from: http://ajp.psychiatryonline.org/data/Journals/AJP/3713/940.pdf
8. Chisholm D, Diehr P, Knapp M, Patrick D, Treglia M, Simon G, et al. Depression status, medical comorbidity and resource costs. Evidence from an international study of major depression in primary care (LIDO). Br J Psychiatry [Internet]. 2003 Aug [cited 2014 Apr 29];183:121-31. Available from: http://bjp.rcpsych.org/content/183/2/121.full.pdf+html
9. Broadhead WE, Blazer DG, George LK, Tse CK. Depression, disability days, and days lost from work in a prospective epidemiologic survey. JAMA. 1990 Nov 21;264(19):2524-8.
10. Fava M. Diagnosis and definition of treatment-resistant depression. Biol Psychiatry. 2003 Apr 15;53(8):649-59.
11. Clinical study report: CN138163. A multicenter, randomized, double-blind, placebo-controlled study of the safety and efficacy of aripiprazole as adjunctive therapy in the treatment of patients with major depressive disorder [**CONFIDENTIAL** internal manufacturer's report]. Wallingford [CT]: Bristol-Myers Squibb; 2007 Mar 15.
12. Clinical study report: CN138139. A multicenter, randomized, double-blind, placebo-controlled study of the safety and efficacy of aripiprazole as adjunctive therapy in the treatment of patients with major depressive disorder [**CONFIDENTIAL** internal manufacturer's report]. Wallingford [CT]: Bristol-Myers Squibb; 2007 Feb 7.
13. Clinical study report: CN138165. A multicenter,randomized, double-blind, placebo-controlled study of the safety and efficacy of aripiprazole as adjunctive therapy in the treatment of patients with major depressive disorder [**CONFIDENTIAL** internal manufacturer's report]. Wallingford [CT]: Bristol-Myers Squibb; 2008 Sep 25.
14. Spielmans GI, Berman MI, Linardatos E, Rosenlicht NZ, Perry A, Tsai AC. Adjunctive Atypical Antipsychotic Treatment for Major Depressive Disorder: A Meta-Analysis of Depression, Quality of Life, and Safety Outcomes. PLoS Med [Internet]. 2013 [cited 2014 Feb 26];10(3). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3595214/pdf/pmed.1001403.pdf

15. CDR submission: Abilify (aripiprazole) tablets . Company: Bristol-Myers Squibb Canada [**CONFIDENTIAL** manufacturer's submission]. St. Laurent (QC): Bristol-Myers Squibb Canada; 2013 Sep.

16. Pharmacoeconomic evaluation. In: CDR submission: Abilify (aripiprazole) tablets. Company: Bristol-Myers Squibb Canada. [**CONFIDENTIAL** manufacturer's submission]. St.Laurent (QC): Bristol-Myers Squibb Canada; 2013 Sep.

17. Montgomery SA, Möller HJ. Is the significant superiority of escitalopram compared with other antidepressants clinically relevant? Int Clin Psychopharmacol. 2009 May;24(3):111-8.

18. National Collaborating Centre for Mental Health Commissioned by the National Institute for Health and Clinical Excellence. Depression: management of depression in primary and secondary care [Internet].The British Psychological Society and The Royal College of Psychiatrists; 2006. [cited 2014 May 28]. (National clinical practice guideline no. 23). Available from: http://www.nccmh.org.uk/downloads/Depression/cg023fullguideline.pdf

19. Dewa CS, Hoch JS, Goering P, Lin E, Paterson M. Use of antidepressants among Canadian workers receiving depression-related short-term disability benefits. Psychiatr Serv. 2003 May;54(5):724-9.

20. Parikh SV, Segal ZV, Grigoriadis S, Ravindran AV, Kennedy SH, Lam RW, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) clinical guidelines for the management of major depressive disorder in adults. II. Psychotherapy alone or in combination with antidepressant medication. J Affect Disord. 2009 Oct;117 Suppl 1:S15-S25.

21. Lam RW, Kennedy SH, Grigoriadis S, McIntyre RS, Milev R, Ramasubbu R, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) clinical guidelines for the management of major depressive disorder in adults. III. Pharmacotherapy. J Affect Disord. 2009 Oct;117 Suppl 1:S26-S43.

22. [Pr]Abilify* aripiprazole tablets, 2, 5, 10, 15, 20 and 20 mg. Antipyschotic agent [product monograph]. Montreal (QC): Bristol-Myers Squibb Canada; 2013 May 27.

23. Berman RM, Marcus RN, Swanink R, McQuade RD, Carson WH, Corey-Lisle PK, et al. The efficacy and safety of aripiprazole as adjunctive therapy in major depressive disorder: a multicenter, randomized, double-blind, placebo-controlled study. J Clin Psychiatry. 2007 Jun;68(6):843-53.

24. Marcus RN, McQuade RD, Carson WH, Hennicken D, Fava M, Simon JS, et al. The efficacy and safety of aripiprazole as adjunctive therapy in major depressive disorder: a second multicenter, randomized, double-blind, placebo-controlled study. J Clin Psychopharmacol. 2008 Apr;28(2):156-65.

25. Berman RM, Fava M, Thase ME, Trivedi MH, Swanink R, McQuade RD, et al. Aripiprazole augmentation in major depressive disorder: a double-blind, placebo-controlled study in patients with inadequate response to antidepressants. CNS Spectr. 2009 Apr;14(4):197-206.

26. Center for Drug Evaluation and Research, U.S. Food and Drug Administration. Medical review(s). In: Abilify (aripiprazole) tablets. Company: Otsuka Pharmaceutical Co., Ltd. Application no.: 21-436s018. Approval date: 11/16/2007. Rockville (MD): The Center; 2007 [cited 2014 May 2]. (FDA drug approval package). Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2007/021436s018TOC.cfm.

27. Center for Drug Evaluation and Research, U.S. Food and Drug Administration. Statistical review(s). In: Abilify (aripiprazole) tablets. Company: Otsuka Pharmaceutical Co., Ltd. Application no.: 21-436s018. Approval date: 11/16/2007. Rockville (MD): The Center; 2007 [cited 2014 May 2]. (FDA drug approval package). Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2007/021436s018TOC.cfm.

28. Health Canada reviewer's report: Abilify (aripiprazole) [**CONFIDENTIAL** internal report]. In: Ottawa: Therapeutic Products Directorate, Health Canada; 2013.

29. Dunner DL, Laubmeier KK, Manos G, Forbes RA, Baker RA, Berman RM. Beneficial effects of adjunctive Aripiprazole in major depressive disorder are not dependent on antidepressant therapy history: a post hoc analysis of 3 randomized, double-blind, placebo-controlled trials. Prim Care Companion J Clin Psychiatry. 2012;14(6).

30. Bristol-Myers Squibb Canada response to March 17, 2014 CDR request for additional information relating to the Abilify-MDD review [**CONFIDENTIAL** additional manufacturer's information]. Montreal (QC): Bristol-Myers Squibb Canada; 2014 Apr 7.

31. Corruble E, Legrand JM, Zvenigorowski H, Duret C, Guelfi JD. Concordance between self-report and clinician's assessment of depression. J Psychiatr Res. 1999 Sep;33(5):457-65.

32. Papakostas GI, Shelton RC, Smith J, Fava M. Augmentation of antidepressants with atypical antipsychotic medications for treatment-resistant major depressive disorder: a meta-analysis. J Clin Psychiatry. 2007;68(6):826-31.

33. Nelson JC, Papakostas GI. Atypical antipsychotic augmentation in major depressive disorder: a meta-analysis of placebo-controlled randomized trials. Am J Psychiatry [Internet]. 2009 Sep [cited 2014 Feb 26];166(9):980-91. Available from: http://ajp.psychiatryonline.org/data/Journals/AJP/3899/09aj0980.PDF

34. Berman RM, Thase ME, Trivedi MH, Hazel JA, Marler SV, McQuade RD, et al. Long-term safety and tolerability of open-label aripiprazole augmentation of antidepressant therapy in major depressive disorder. Neuropsychiatr Dis Treat [Internet]. 2011 [cited 2014 Apr 22];7:303-12. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3104689

35. Zimmerman M, Posternak MA, Chelminski I. Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. J Psychiatr Res. 2004 Nov;38(6):577-82.

36. Lam RW, Michalak EE, Swinson RP. Assessment scales in depression, mania and anxiety. London: Taylor and Francis; 2005.

37. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med. 2008 Feb;5(2):e45.

38. Fava M, Mischoulon D, Iosifescu D, Witte J, Pencina M, Flynn M, et al. A double-blind, placebo-controlled study of aripiprazole adjunctive to antidepressant therapy among depressed outpatients with inadequate response to prior antidepressant therapy (ADAPT-A Study). Psychother Psychosom. 2012;81(2):87-97.

39. Clinical study report: CN138164. A multicenter, long-term, open-label study to assess the safety and tolerability of aripiprazole as adjunctive therapy in the treatment of outpatients with major depressive disorder [**CONFIDENTIAL** internal manufacturer's report]. Wallingford [CT]: Bristol-Myers Squibb; 2008 Jul 15.

40. Zimmerman M, Posternak MA, Chelminski I. Is the cutoff to define remission on the Hamilton Rating Scale for Depression too high? J Nerv Ment Dis. 2005 Mar;193(3):170-5.

41. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? Am J Psychiatry. 2004 Dec;161(12):2163-77.

42. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry. 2003 Sep 1;54(5):573-83.

43. Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger J, Burns C. The Inventory for Depressive Symptomatology (IDS): preliminary findings. Psychiatry Res. 1986 May;18(1):65-87.
44. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med. 1996 May;26(3):477-86.
45. Biggs MM, Shores-Wilson K, Rush AJ, Carmody TJ, Trivedi MH, Crismon ML, et al. A comparison of alternative assessments of depressive symptom severity: a pilot study. Psychiatry Res. 2000 Nov 20;96(3):269-79.
46. Guy W. Clinical Global Impressions (CGI) scale. 1st ed. In: Handbook of psychiatric measures. Washington (DC): American Psychiatric Association; 2000. p. 100-2.
47. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. Int Clin Psychopharmacol. 1996 Jun;11 Suppl 3:89-95.
48. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. Psychopharmacol Bull. 1993;29(2):321-6.
49. Heo M, Murphy CF, Meyers BS. Relationship between the Hamilton Depression Rating Scale and the Montgomery-Åsberg Depression Rating Scale in depressed elderly: a meta-analysis. Am J Geriatr Psychiatry. 2007 Oct;15(10):899-905.
50. Moncrieff J, Kirsch I. Efficacy of antidepressants in adults. BMJ. 2005 Jul 16;331(7509):155-7.
51. Edwards BC, Lambert MJ, Moran PW, McCully T, Smith KC, Ellingson AG. A meta-analytic comparison of the Beck Depression Inventory and the Hamilton Rating Scale for Depression as measures of treatment outcome. Br J Clin Psychol. 1984 May;23 (Pt 2):93-9.
52. Lambert MJ, Hatch DR, Kingston MD, Edwards BC. Zung, Beck, and Hamilton Rating Scales as measures of treatment outcome: a meta-analytic comparison. J Consult Clin Psychol. 1986 Feb;54(1):54-9.
53. Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D, et al. The Montgomery Äsberg and the Hamilton ratings of depression: a comparison of measures. Eur Neuropsychopharmacol. 2006 Dec;16(8):601-11.
54. Faries D, Herrera J, Rayamajhi J, DeBrota D, Demitrack M, Potter WZ. The responsiveness of the Hamilton Depression Rating Scale. J Psychiatr Res. 2000 Jan;34(1):3-10.
55. Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. Arch Gen Psychiatry. 1991 Sep;48(9):851-5.
56. National Collaborating Centre for Mental Health Commissioned by the National Institute for Health and Clinical Excellence. Depression: the NICE guideline on the treatment and management of depression in adults [Internet].The British Psychological Society and The Royal College of Psychiatrists; 2014. [cited 2014 May 28]. (National clinical practice guideline 90). Available from: http://www.nice.org.uk/nicemedia/live/12329/45896/45896.pdf
57. Müller MJ, Himmerich H, Kienzle B, Szegedi A. Differentiating moderate and severe depression using the Montgomery-Åsberg depression rating scale (MADRS). J Affect Disord. 2003 Dec;77(3):255-60.
58. Montgomery S. Clinically relevant outcomes measures in major depressive disorder [abstract]. Int J Neuropsychopharmacol. 2008;11(Suppl 1):318.
59. Davidson J, Turnbull CD, Strickland R, Miller R, Graves K. The Montgomery-Asberg Depression Scale: reliability and validity. Acta Psychiatr Scand. 1986 May;73(5):544-8.

60. Senra C. Evaluation and monitoring of symptom severity and change in depressed outpatients. J Clin Psychol. 1996 May;52(3):317-24.

61. Mulder RT, Joyce PR, Frampton C. Relationships among measures of treatment outcome in depressed patients. J Affect Disord. 2003 Sep;76(1-3):127-35.

62. Maier W, Philipp M, Heuser I, Schlegel S, Buller R, Wetzel H. Improving depression severity assessment--I. Reliability, internal validity and sensitivity to change of three observer depression scales. J Psychiatr Res. 1988;22(1):3-12.

63. Iannuzzo RW, Jaeger J, Goldberg JF, Kafantaris V, Sublette ME. Development and reliability of the HAM-D/MADRS interview: an integrated depression symptom rating scale. Psychiatry Res. 2006 Nov 29;145(1):21-37.

64. Zimmerman M, Posternak MA, Chelminski I. Defining remission on the Montgomery-Asberg depression rating scale. J Clin Psychiatry. 2004 Feb;65(2):163-8.

65. Zimmerman M, Chelminski I, Posternak M. A review of studies of the Montgomery-Asberg Depression Rating Scale in controls: implications for the definition of remission in treatment studies of depression. Int Clin Psychopharmacol. 2004 Jan;19(1):1-7.

66. Hawley CJ, Gale TM, Sivakumaran T, Hertfordshire Neuroscience Research group. Defining remission by cut off score on the MADRS: selecting the optimal value. J Affect Disord. 2002 Nov;72(2):177-84.

67. Stevanovic D. Quality of Life Enjoyment and Satisfaction Questionnaire-short form for quality of life assessments in clinical practice: a psychometric study. J Psychiatr Ment Health Nurs. 2011 Oct;18(8):744-50.

68. Komossa K, Depping AM, Gaudchau A, Kissling W, Leucht S. Second-generation antipsychotics for major depressive disorder and dysthymia. Cochrane Database Syst Rev. 2010;(12):CD008121-.