Comparative Effectiveness Review
Number 63

Effective Health Care Program

First-Generation Versus Second-Generation Antipsychotics in Adults: Comparative Effectiveness

Executive Summary

Introduction

Antipsychotic medications are used to treat and manage symptoms for several psychiatric disorders and are commonly categorized into two classes. First-generation antipsychotics (FGAs), also known as "typical antipsychotics," were developed in the 1950s. Second-generation antipsychotics (SGAs), also known as "atypical antipsychotics," emerged in the 1980s. To date, FGAs have been classified according to their chemical structure, which includes serotonin-dopamine antagonists and multiacting receptor-targeted antipsychotics, whereas SGAs have been categorized according to their pharmacological properties as dopamine partial agonists. There is ongoing research testing the proposed mechanisms of action within each class with respect to the neurobiology of different psychiatric disorders.[1,2]

According to findings from the 2004–05 Medical Expenditure Panel Survey, an estimated 2 million adult patients in the United States were prescribed an antipsychotic medication, three-quarters of whom were taking an SGA.[3] In 2003, an estimated $2.82 billion were spent in the country on these medications, with SGAs accounting for 93 percent of this expenditure.[3]

Effective Health Care Program

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at **www.effectivehealthcare.ahrq.gov/reports/final.cfm**.

Today, 20 FGAs and SGAs are commercially available in the United States and approved by the U.S. Food and Drug Administration (FDA).

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Effective Health Care

Individuals taking antipsychotics may stop taking their medication for a number of reasons, including adverse events (AEs) and a lack of improvement in their symptoms.[4] As a result, ongoing evaluations of drug efficacy and models of patient decisionmaking are essential.

This Comparative Effectiveness Review (CER) provides a comprehensive synthesis of the evidence examining the benefits and harms associated with the use of FDA-approved FGAs and SGAs. In contrast to previous reviews, this CER focuses on comparisons of individual medications rather than drug classes. This topic is important and timely, given the ongoing debate about the comparative benefits and harms of FGAs and SGAs.[5] Moreover, the focus of this report complements other recent reviews investigating different SGAs,[6] the off-label use of antipsychotics,[7] and FGAs versus SGAs in the pediatric population.[8] The focus of this report is adults age 18 to 64 years with schizophrenia, schizophrenia-related psychoses, and bipolar disorder. This age group is the normal demographic in which these illnesses have been shown to be prevalent. The illnesses are discussed in more detail in the sections that follow.

Key Questions

The following Key Questions (KQs) were investigated in the report:

1. For adults (age 18 to 64 years) with schizophrenia, schizophrenia-related psychoses, or bipolar disorder, what are the comparative efficacy and effectiveness of FGAs versus SGAs for improving core illness symptoms? The following core symptoms were considered:
   - Schizophrenia or related psychoses: positive (i.e., delusions and hallucinations) and negative (i.e., passive or apathetic social withdrawal and blunted affect) symptoms, general psychopathology (i.e., preoccupation, lack of insight, and motor retardation), and global ratings and total scores.
   - Core illness symptoms for bipolar disorder: mood, motor activity or energy, sleep, speech, behavior, and mood stability.
2. For adults (age 18 to 64 years) with schizophrenia, schizophrenia-related psychoses, or bipolar disorder, what is the comparative effectiveness of FGAs versus SGAs for improving functional outcomes and decreasing health care system utilization?
   - Functional outcomes include any of the following: employment or personal earnings, social relatedness or functioning, encounters with the legal system, sexual function or dysfunction, functional capacity, and living situation.
   - Health care system utilization includes: time to hospitalization or rehospitalization because of mental illness and all other causes, rates of hospitalization or rehospitalization, mean hospital bed days, length of hospitalization stay, rates of emergency department visits, attendance in day care programs, and use of ancillary caseworkers.
3. For adults (age 18 to 64 years) with schizophrenia, schizophrenia-related psychoses, or bipolar disorder, do FGAs and SGAs differ in medication-associated AEs and safety? AEs included:
   - Overall AEs.
   - Specific AEs:
     - *Major:* mortality, cerebrovascular disease-related events, development of diabetes mellitus, diabetic ketoacidosis, neuroleptic malignant syndrome, seizures, tardive dyskinesia, cardiomyopathies and cardiac arrhythmias, agranulocytosis, suicide-related behaviors, and death by suicide.
     - *General:* extrapyramidal symptoms (EPS), weight changes, agitation, constipation, sedation, elevated cholesterol, AEs related to prolactin elevations, galactorrhea or bloody galactorrhea, hypotension, and metabolic changes (including changes in glucose levels, triglycerides, lipids, and the risk of developing diabetes).
   - Study withdrawals and time to withdrawal because of AEs.
   - Persistence and reversibility of AEs.
4. For adults (age 18 to 64 years) with schizophrenia, schizophrenia-related psychoses, or bipolar disorder, what is the comparative effectiveness of FGAs versus SGAs for the following other outcomes:
   - Medication adherence and persistent use (and associated dosing and time to discontinuation of treatment).
   - Patient insight into illness.
   - Health-related quality of life.
   - Patient satisfaction.

- Comorbidity: endpoints of victimization, homelessness, and substance abuse.
- Patient-reported outcomes.
- Ability to obtain and retain employment and succeed in job duties.
- Concomitant use of other medications, especially those used to treat EPS.
- Patient preferences.

5. For adults (age 18 to 64 years) with schizophrenia, schizophrenia-related psychoses, or bipolar disorder, what are the comparative effectiveness and risks of FGAs versus SGAs in subgroups defined by the following variables?
   - Disorder subtypes.
   - Sex.
   - Age group (18–35 years, 36–54 years, and 55–64 years).
   - Race.
   - Comorbidities.
   - Drug dosage.
   - Followup period.
   - Treatment of a first episode versus treatment in the context of previous episodes (previous exposure to antipsychotics).
   - Treatment resistance.

Methods

In general, we followed methodologically rigorous methods for systematic reviews as described in recent standards documents.[7,8] Detailed information on the reports prepared by Evidence-based Practice Centers can be found on the Agency for Healthcare Research and Quality Web site at www.effectivehealthcare.ahrq.gov.

Literature Search

We conducted comprehensive searches in the following electronic databases: MEDLINE®, Embase, PsycINFO, International Pharmaceutical Abstracts, CINAHL, ProQuest® Dissertations and Theses–Full Text, Cochrane Central Register of Controlled Trials (CENTRAL), and Scopus™.The searches are up to date to July 2011. For the questions on AEs, we also searched the U.S. National Library of Medicine's TOXLINE® and the MedEffect™ Canada Adverse Drug Reaction Database.

We hand-searched proceedings for the Annual Convention of the American Psychiatric Association (2008–10), the International College of Neuropsychopharmacology (2008–10), and the International Society for Bipolar Disorders (2008–10). We searched clinical trials registers, contacted experts in the field, and contacted authors of relevant studies. In addition, we reviewed the reference lists of reviews and guidelines and searched for articles citing the studies that met our inclusion criteria using Scopus™ Citation Tracker.

Study Selection

Two reviewers independently screened titles and abstracts to determine if an article met the broad inclusion criteria for study design, population, interventions, and comparators. We independently rated each article as "include," "exclude," or "unclear." We retrieved the full text of studies identified as "include" or "unclear." Two reviewers independently reviewed each article using a priori eligibility criteria and a standardized form. We resolved discrepancies through discussion and consensus or by third-party adjudication.

We included studies if they: were randomized (RCTs) or nonrandomized controlled trials (nRCTs), or prospective or retrospective cohort studies with a followup of 2 years or greater; included adults age 18 to 64 years with schizophrenia or related psychoses or bipolar disorder; and compared a commercially available FDA-approved FGA with an FDA-approved SGA.

Quality Assessment and Rating the Body of Evidence

Two reviewers independently assessed the methodological quality of included studies and resolved disagreements through discussion and consensus or third-party adjudication. We assessed RCTs and nRCTs using the Cochrane Collaboration's Risk of Bias tool.[5] We assessed cohort studies using the Newcastle-Ottawa Scale.[7] A priori, the research team developed decision rules regarding application of the tools.

Two reviewers independently evaluated the overall strength of evidence (SoE) using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach used by Evidence-based Practice Centers and resolved discrepancies through discussion. We examined the following four major domains: risk of bias (low, medium, or high), consistency (inconsistency not present, inconsistency present, unknown, or not applicable), directness (direct or indirect), and precision

(precise or imprecise). We assigned an overall evidence grade of "high," "moderate," "low," or "insufficient."

We graded core illness symptoms in the categories of positive symptoms, negative symptoms, general psychopathology, and global ratings and total scores. We provided a grade for each different scale that was used. We graded the following AEs, which were deemed to be most clinically important a priori: diabetes mellitus, mortality, tardive dyskinesia, and major metabolic syndrome. These outcomes were identified a priori as being the most clinically important for decisionmaking.

Data Extraction

Two reviewers independently extracted data using standardized data extraction forms and resolved discrepancies through discussion and consensus or by third-party adjudication. We extracted information on study characteristics, population, interventions and dosing regimens, outcomes assessed, results, and funding source. When studies incorporated multiple relevant treatment arms or multiple followup periods, we extracted data from all groups for the longest followup data. When there were multiple reports of the same study, we referenced the primary or most relevant study and extracted only additional data from companion reports.

Data Analysis

We presented evidence tables for all studies and a qualitative description of results. We conducted meta-analyses using random effects models to answer the KQs when studies were sufficiently similar in terms of design, population, interventions, and outcomes. We presented results separately for the conditions of interest (schizophrenia or schizophrenia-related psychoses and bipolar disorder). Within each condition, we presented results separately for each individual comparison of FGA versus SGA. We quantified statistical heterogeneity using the I-squared (I^2) statistic.

Applicability

We assessed the applicability of the body of evidence using the PICOTS format (population, intervention, comparator, outcomes, timing of outcome measurement, and setting). We reported factors that may potentially limit the applicability of the results. These included patient characteristics (e.g., age, diagnostic criteria, severity of illness, comorbidities, concomitant medications, inpatient or outpatient status) and study characteristics (e.g., length of followup).

Results

Description of Included Studies

The searches identified 9,411 unique study reports. A total of 125 primary publications and 146 companion publications were included. The studies included 121 RCTs, 2 nRCTs, and 2 retrospective cohort studies. The studies were published between 1974 and 2010. The majority of studies were multicenter (n = 70, 56 percent) and involved inpatients (n = 62, 50 percent), and they were conducted more often in North America than elsewhere (n = 57, 46 percent). The number of participants in the studies ranged from 10 to 95,632 (median = 86 [interquartile range (IQR), 36 to 300]). The average age of study participants ranged from 21 to 50 years (median = 37 years [IQR, 33 to 41]). The length of followup ranged from <1 day to 22 years (median = 8 weeks [IQR, 6 to 26 weeks]). Seventy percent of studies (n = 88) had some form of support from the pharmaceutical industry.

Overall, 113 studies examined schizophrenia or schizophrenia-related psychoses, 11 studies examined bipolar disorder, and 1 study included both. A total of 22 and 6 drug comparisons were made for schizophrenia and bipolar disorder, respectively (Table A).

Methodological Quality of Included Studies

None of the 123 RCTs and nRCTs was rated as having a low risk of bias. The majority of the trials (n = 78, 63 percent) had an unclear risk of bias; the remaining trials (n = 45, 37 percent) had a high risk of bias. In the majority of cases, trials were assessed as having unclear risk of bias due to unclear reporting with respect to sequence generation, concealment of allocation, and methods of blinding. The most common reasons for trials to be assessed as having high risk of bias were lack of blinding and inadequate handling or reporting of outcome data.

Data were collected retrospectively in both cohort studies. The methodological quality of the cohort studies was good.

Results of Included Studies

The results are presented by the KQs they address. Within each KQ, we present results by condition and comparison. Tables with a summary of findings for efficacy and safety are presented below. It is important to note that lack of statistical significance does not equate to equivalence or noninferiority, nor does statistical significance equate to clinical significance.

Table A. Comparisons examined in the included studies	
Comparison	**n**
Schizophrenia or Schizophrenia-Related Psychoses	
Chlorpromazine vs. clozapine	12
Chlorpromazine vs. olanzapine	1
Chlorpromazine vs. quetiapine	1
Chlorpromazine vs. ziprasidone	1
Fluphenazine vs. olanzapine	2
Fluphenazine vs. quetiapine	1
Fluphenazine vs. risperidone	1
Haloperidol vs. aripiprazole	8
Haloperidol vs. asenapine	1
Haloperidol vs. clozapine	11[a]
Haloperidol vs. olanzapine	35[a]
Haloperidol vs. quetiapine	11[a]
Haloperidol vs. risperidone	39[b]
Haloperidol vs. ziprasidone	9[c]
Perphenazine vs. aripiprazole	1
Perphenazine vs. olanzapine	2
Perphenazine vs. quetiapine	1
Perphenazine vs. risperidone	2
Perphenazine vs. ziprasidone	1
Trifluoperazine vs. clozapine	1
Thioridazine vs. clozapine	1
Thioridazine vs. risperidone	1
Bipolar Disorder	
Chlorpromazine vs. clozapine	1
Haloperidol vs. aripiprazole	2
Haloperidol vs. olanzapine	2
Haloperidol vs. quetiapine	1
Haloperidol vs. risperidone	5
Haloperidol vs. ziprasidone	1

n = number of studies; nRCT = nonrandomized controlled trial
[a]Includes 1 cohort study.
[b]Includes 1 cohort study and 1 nRCT.
[c]Includes 1 nRCT.
Note: n = 125.

KQ1: Core Illness Symptoms

The findings for core illness symptoms are presented for each condition in Table B. Comparisons and outcomes for which there was insufficient SoE to draw a conclusion (e.g., evidence from single trials) are not displayed in the tables. The SoE comparing individual FGAs and SGAs was insufficient to draw conclusions for the following comparisons: chlorpromazine versus olanzapine, quetiapine, and ziprasidone; fluphenazine versus olanzapine, quetiapine, and risperidone; haloperidol versus asenapine; perphenazine versus aripiprazole, olanzapine, quetiapine, risperidone, and ziprasidone; trifluoperazine versus clozapine.

Table B. Summary of the strength of evidence for core illness symptoms (KQ1)			
Outcome	**Comparison**	**SoE**	**Summary (Number of Studies)**
Schizophrenia and Schizophrenia-Related Psychoses			
Positive symptoms	Haloperidol vs. aripiprazole	Low	No significant difference for PANSS (2 RCTs).
	Haloperidol vs. clozapine	Low	No significant difference for PANSS (2 RCTs).
	Haloperidol vs. olanzapine	Low	No difference for PANSS (14 RCTs) or SAPS (2 RCTs).
	Haloperidol vs. quetiapine	Low	No significant difference for PANSS (4 RCTs).
	Haloperidol vs. risperidone	Low	No difference for PANSS (20 RCTs) or SAPS (2 RCTs).
Negative symptoms	Haloperidol vs. aripiprazole	Moderate	Significant difference favoring aripiprazole for PANSS (3 RCTs).
	Haloperidol vs. clozapine	Low	No significant difference for PANSS (2 RCTs) or SANS (2 RCTs).
	Haloperidol vs. olanzapine	Moderate	Significant difference favoring olanzapine for PANSS (14 RCTs) and SANS (5 RCTs).
	Haloperidol vs. quetiapine	Low	No significant difference for PANSS (4 RCTs).
	Haloperidol vs. risperidone	Low to moderate	Significant difference favoring risperidone for SANS (moderate SoE, 4 RCTs). No significant difference for PANSS (low SoE, 20 RCTs).
	Haloperidol vs. ziprasidone	Low	No significant difference for PANSS (2 RCTs).
General psychopathology	Haloperidol vs. clozapine	Low	No significant difference for PANSS (2 RCTs).
	Haloperidol vs. olanzapine	Low	Significant difference favoring olanzapine for HAM–D (moderate SoE, 3 RCTs) and MADRS (moderate SoE, 6 RCTs). No difference for ABS (low SoE, 2 RCTs), ACES (low SoE, 2 RCTs), CDS–S (low SoE, 3 RCTs), HAM–A (low SoE, 2 RCTs), or PANSS (low SoE, 10 RCTs).
	Haloperidol vs. quetiapine	Low to moderate	No significant difference for CDS–S (2 RCTs) or PANSS (4 RCTs).
Global ratings and total scores	Chlorpromazine vs. clozapine	Moderate	Significant difference favoring clozapine for BPRS (6 RCTs).
	Haloperidol vs. aripiprazole	Low	No significant difference for BPRS (3 RCTs) or CGI–S (5 RCTs).
	Haloperidol vs. clozapine	Low	No difference for BPRS (4 RCTs) or PANSS (3 RCTs).
	Haloperidol vs. olanzapine	Low to moderate	Significant difference favoring olanzapine for CGI–S (moderate SoE, 7 RCTs) and PANSS (moderate SoE, 14 RCTs). No difference for BPRS (low SoE, 13 RCTs) or CGI–I (low SoE, 2 RCTs).
	Haloperidol vs. quetiapine	Low to moderate	Significant difference favoring haloperidol for CGI–S (moderate SoE, 4 RCTs). No difference for BPRS (low SoE, 4 RCTs), CGI–I (low SoE, 3 RCTs), or PANSS (low SoE, 6 RCTs).
	Haloperidol vs. risperidone	Low	No difference for BPRS (13 RCTs), CGI–I (3 RCTs), CGI–S (8 RCTs), or PANSS (20 RCTs).
	Haloperidol vs. ziprasidone	Low	No significant difference for BPRS (4 RCTs), CGI–S (4 RCTs), GAF (3 RCTs), or PANSS (4 RCTs).

Table B. Summary of the strength of evidence for core illness symptoms (KQ1) (continued)			
Outcome	**Comparison**	**SoE**	**Summary (Number of Studies)**
Bipolar Disorder			
Mood (mania)	Haloperidol vs. aripiprazole	Low	No significant difference in YMRS (2 RCTs).
	Haloperidol vs. olanzapine	Low	No significant difference in YMRS (2 RCTs).
	Haloperidol vs. risperidone	Low	No significant difference in YMRS (3 RCTs).
Mood (depression)	Haloperidol vs. aripiprazole	Low	No significant difference in MADRS (2 RCTs).
Global ratings and total scores	Haloperidol vs. aripiprazole	Low	No significant difference in CGI–BP (2 RCTs).

ABS = Agitated Behavior Scale; ACES = Agitation-Calmness Evaluation Scale; BPRS = Brief Psychiatric Rating Scale; CDS–S = Calgary Depression Scale for Schizophrenia; CGI–BP = Clinical Global Impression–Bipolar; CGI–I = Clinical Global Impression–Improvement; CGI–S = Clinical Global Impression–Severity; GAF = Global Assessment of Functioning; HAM–A = Hamilton Rating Scale for Anxiety; HAM–D = Hamilton Rating Scale for Depression; KQ = Key Question; MADRS = Montgomery-Asberg Depression Rating Scale; PANSS = Positive and Negative Symptom Scale; RCT = randomized controlled trial; SANS = Scale for the Assessment of Negative Symptoms; SAPS = Scale for the Assessment of Positive Symptoms; SoE = strength of evidence; YMRS = Young Mania Rating Scale

For schizophrenia or related psychoses, seven studies provided data on core illness symptoms for chlorpromazine versus clozapine. No differences were found for positive or negative symptoms or general psychopathology. Clozapine showed benefits for total score (moderate SoE).

Eight studies provided data on core illness symptoms for haloperidol versus clozapine. No significant differences were found for positive symptoms, negative symptoms, or general psychopathology (low SoE). The findings were discordant for total symptom score: no difference was found based on the Brief Psychiatric Rating Scale (BPRS) and Positive and Negative Symptom Scale (PANSS) (low SoE); one study showed benefits for clozapine on the Clinical Global Impression–Improvement (CGI–I) and Clinical Global Impression–Severity (CGI–S) scales (insufficient SoE).

Twenty-seven studies provided data on core illness symptoms for haloperidol versus olanzapine. No differences were found for positive symptoms (low SoE). Olanzapine was favored for negative symptoms (moderate SoE). In terms of general psychopathology, a significant benefit for olanzapine was found based on the Hamilton Rating Scale for Depression (HAM–D), Montgomery-Asberg Depression Rating Scale (MADRS), and Young Mania Rating Scale (YMRS). No differences were observed for the other five scales of general symptoms assessed. The SoE varied across outcomes from insufficient to moderate. Olanzapine was favored for global ratings and total symptom scores based on the CGI–S and PANSS; however, no differences were found for the other four scales assessed. The SoE for these outcomes also varied from insufficient to moderate.

Nine studies provided data on core illness symptoms for haloperidol versus quetiapine. No significant differences were found for positive or negative symptoms, or general psychopathology. A significant difference favoring haloperidol was found for one of the five global ratings (CGI–S) and total symptom scores assessed. The SoE across outcomes ranged from insufficient to moderate.

Thirty-one studies provided data on core illness symptoms for haloperidol versus risperidone. There were no differences for positive symptoms (low SoE). Risperidone was favored for negative symptoms based on the Scale for the Assessment of Negative Symptoms (SANS) (moderate SoE); in contrast, no difference for negative symptoms was found based on PANSS (low SoE). No differences were found for any of the six measures used to assess general psychopathology (low to insufficient SoE). Seven of the global ratings or total symptom scores showed no differences, whereas the Symptom Checklist (SCL–90–R) showed a benefit for risperidone (low to insufficient SoE).

Seven studies provided data on core illness symptoms for haloperidol versus ziprasidone. There were no significant differences in terms of negative symptoms, general psychopathology, global ratings, or total score (low to insufficient SoE). No studies provided data on positive symptoms.

A total of 12 studies included patients with bipolar disorder. The most frequent comparison was haloperidol versus risperidone (five RCTs). No significant differences were found for mood (mania), mood (depression), positive or negative symptoms, or global ratings and total scores (low to insufficient SoE). Two studies compared haloperidol versus olanzapine and found no significant

differences in sleep, mood (mania), mood (depression), or global ratings and total scores (low or insufficient SoE). Two studies compared haloperidol with aripiprazole and found no differences in mood (mania), mood (depression), positive or negative symptoms, or global ratings and total scores (low or insufficient SoE). Single studies compared chlorpromazine versus clozapine and haloperidol versus quetiapine and ziprasidone (insufficient SoE).

KQ2: Functional Outcomes and Health Care System Utilization

The findings for functional outcomes and health care system utilization are presented for each condition and comparison in Table C. We did not assess the SoE for outcomes in KQ2.

Results for functional outcomes were available from nine head-to-head comparisons in studies of patients with schizophrenia or schizophrenia-related psychoses. No significant differences in functional outcomes were observed between groups for any of the comparisons. However, in most cases evidence came from single studies. Results for health care system utilization were available for 10 head-to-head comparisons, and no differences were found for any comparison.

Only one trial comparing haloperidol with olanzapine provided data on functional outcomes in patients with bipolar disorder. Significant differences were found favoring olanzapine in terms of the number of individuals actively working for pay. No differences were found for impairment in household or work activities.

Table C. Summary of evidence for functional outcomes, health care system utilization, and other outcomes (KQ2)

Outcome	Comparison	Summary (Number of Studies)
Schizophrenia and Schizophrenia-Related Psychoses		
Functional outcomes	Fluphenazine vs. quetiapine	No significant difference for sexual dysfunction or improvement on treatment (1 RCT).
	Fluphenazine vs. risperidone	No significant difference for sexual dysfunction or improvement on treatment (1 RCT).
	Haloperidol vs. olanzapine	No significant difference for positive urine toxicology (1 RCT) or sexual dysfunction (1 RCT).
	Haloperidol vs. quetiapine	No significant difference for sexual dysfunction (1 RCT).
	Haloperidol vs. risperidone	No significant difference for economic independence (1 RCT) or attitude regarding drugs (1 RCT).
	Haloperidol vs. ziprasidone	No difference for sexual dysfunction (1 RCT).
	Perphenazine vs. quetiapine	No significant difference in patients with paid employment (1 RCT).
	Perphenazine vs. risperidone	No significant difference in patients with paid employment (1 RCT).
	Perphenazine vs. ziprasidone	No significant difference in patients with paid employment (1 RCT).
Health care system use	Chlorpromazine vs. clozapine	No significant difference in rates of hospitalization or rehospitalization (1 RCT).
	Haloperidol vs. clozapine	No significant difference in mean hospital bed days (1 RCT).
	Haloperidol vs. olanzapine	No significant difference in mean hospital bed days or rates of hospitalization or rehospitalization (1 RCT).
	Haloperidol vs. quetiapine	No significant difference in rates of hospitalization or rehospitalization (1 RCT).
	Haloperidol vs. risperidone	No significant difference in rates of hospitalization or rehospitalization (3 RCTs).
	Haloperidol vs. ziprasidone	No significant difference in rates of hospitalization or rehospitalization (2 RCTs).
	Perphenazine vs. olanzapine	No significant difference in rates of hospitalization or rehospitalization (1 RCT).

Table C. Summary of evidence for functional outcomes, health care system utilization, and other outcomes (KQ2) (continued)		
Outcome	**Comparison**	**Summary (Number of Studies)**
Schizophrenia and Schizophrenia-Related Psychoses (continued)		
Health care system use (continued)	Perphenazine vs. quetiapine	No significant difference in rates of hospitalization or rehospitalization (1 RCT).
	Perphenazine vs. risperidone	No significant difference in rates of hospitalization or rehospitalization (1 RCT).
	Perphenazine vs. ziprasidone	No significant difference in rates of hospitalization or rehospitalization (1 RCT).
Bipolar Disorder		
Functional outcomes	Haloperidol vs. olanzapine	Significant difference favoring olanzapine for number of active workers (i.e., work for pay) (1 RCT). No difference in impairment in household or work activities (1 RCT).

KQ = Key Question; RCT = randomized controlled trial

KQ3: Medication-Associated AEs and Safety

The findings for the AEs that were deemed most clinically important are summarized in Table D. The SoE comparing individual FGAs and SGAs was insufficient to draw conclusions for the following outcomes and comparisons: tardive dyskinesia (chlorpromazine vs. clozapine and ziprasidone; haloperidol vs. clozapine, olanzapine, quetiapine, and ziprasidone); mortality (chlorpromazine vs. clozapine and ziprasidone; haloperidol vs. risperidone; thioridazine vs. clozapine and risperidone); diabetes mellitus (haloperidol vs. olanzapine; perphenazine vs. olanzapine, quetiapine, risperidone, and ziprasidone); and metabolic syndrome (haloperidol vs. clozapine; perphenazine vs. olanzapine, quetiapine, risperidone, and ziprasidone).

Two trials each provided data on mortality for chlorpromazine versus clozapine and haloperidol versus aripiprazole; no significant differences were found, although the length of followup of the trials for the latter comparison was only 24 hours. For metabolic syndrome, two trials provided data for haloperidol versus olanzapine and showed no significant difference in incidence of metabolic syndrome. The SoE for these comparisons was low, suggesting that further research may change the results and change our confidence in the results.

Data were also recorded for general measures of AEs and specific AEs by physiological system (e.g., cardiovascular, endocrine); these outcomes were not assessed for SoE. For general measures of AEs, significant differences were found in the incidence of patients with AEs and withdrawals due to AEs for several comparisons. Most often, the comparison included haloperidol, and the risk was consistently higher for the FGA. The most frequently reported AEs with significant differences were in the category of EPS, and they most often involved a comparison with haloperidol. In the vast majority of cases, the SGA had the preferred AE profile for EPS.

Table D. Summary of the strength of evidence for medication-associated adverse events and safety (KQ3)			
Adverse Event	**Comparison**	**SoE**	**Summary (Number of Studies)**
Mortality	Chlorpromazine vs. clozapine	Low	No significant difference (2 RCTs, length of followup: 52 and 208 wks).
	Haloperidol vs. aripiprazole	Low	No significant difference (2 RCTs, length of followup: 24 hrs for both).
Metabolic syndrome	Haloperidol vs. olanzapine	Low	No significant difference (2 RCTs, length of followup: 6 and 12 wks).

KQ = Key Question; RCT = randomized controlled trial; SoE = strength of evidence

We were unable to adequately examine persistence and reversibility of AEs due to the relatively short followup of the included studies: study followup periods averaged 8 weeks. It is unclear whether AE persistence and reversibility of several significant AEs could be reasonably examined during this time period (e.g., metabolic conditions, body mass index or weight, and cardiovascular measures).

KQ4: Other Outcomes

The findings for other outcomes are presented for each condition and comparison in Table E. We did not assess the SoE for outcomes in KQ4.

Results for other outcomes were available for 19 head-to-head comparisons in studies of patients with schizophrenia or schizophrenia-related psychoses. Few significant differences were found across the comparisons and outcomes examined. For all significant findings, the SGA was preferred. The most commonly reported other outcome was response rate. A significant difference in response rates based on three studies was found favoring clozapine compared with chlorpromazine. Olanzapine was favored over haloperidol for remission (3 trials) and response rates (14 trials). Significant differences were found favoring aripiprazole over haloperidol for caregiver satisfaction (one trial) and patient satisfaction (one trial). Risperidone was favored over haloperidol for relapse rates (six trials). Olanzapine was favored over perphenazine for time to all-cause medication discontinuation (one trial). Health-related quality of life was evaluated for the following comparisons, and no significant differences were found: haloperidol versus olanzapine, quetiapine, risperidone, and ziprasidone (one trial each); perphenazine versus aripiprazole, olanzapine, quetiapine, risperidone, and ziprasidone (one trial each).

Table E. Summary of the evidence for other outcomes (KQ4)

Comparison	Summary[a] (Number of Studies)
Schizophrenia and Schizophrenia-Related Psychoses	
Chlorpromazine vs. clozapine	Significant difference favoring clozapine for response rates (3 RCTs). No difference in remission rates (2 RCTs).
Chlorpromazine vs. olanzapine	No significant difference in response rates (1 RCT).
Chlorpromazine vs. quetiapine	No significant difference in response rates (1 RCT).
Chlorpromazine vs. ziprasidone	No significant difference in response rates (1 RCT).
Fluphenazine vs. olanzapine	No significant difference in response rates (1 RCT).
Fluphenazine vs. quetiapine	No significant difference in response rates (1 RCT).
Fluphenazine vs. risperidone	No significant difference in response rates (1 RCT).
Haloperidol vs. aripiprazole	No significant difference in response rates (5 RCTs) or medication adherence (1 RCT). Difference favoring aripiprazole for caregiver and patient satisfaction (1 RCT).
Haloperidol vs. asenapine	No significant difference in response rates (1 RCT).
Haloperidol vs. clozapine	No significant difference in relapse (1 RCT), response (2 RCTs) or remission (1 RCT) rates or patient satisfaction (1 RCT).
Haloperidol vs. olanzapine	Significant difference favoring olanzapine for response rates (14 RCTs) and remission rates (3 RCTs). No significant difference in medication adherence (1 RCT), patient insight into illness (1 RCT), or HRQoL (5 RCTs).
Haloperidol vs. quetiapine	No significant difference in response rates (6 RCTs), remission rates (1 RCT), or HRQoL (1 RCT).
Haloperidol vs. risperidone	Significant difference favoring risperidone for relapse rates (6 RCTs). No significant difference in remission rates (2 RCTs), response rates (16 RCTs), medication adherence (3 RCTs), patient satisfaction (1 RCT), or HRQoL (2 RCTs).
Haloperidol vs. ziprasidone	No significant difference in response rates (6 RCTs), remission rates (3 RCTs), or HRQoL (2 RCTs).
Perphenazine vs. aripiprazole	No significant difference in response rates (1 RCT) or HRQoL (1 RCT).

Table E. Summary of the evidence for other outcomes (KQ4) (continued)	
Comparison	**Summary[a] (Number of Studies)**
Schizophrenia and Schizophrenia-Related Psychoses (continued)	
Perphenazine vs. olanzapine	No significant difference in HRQoL (1 RCT). Significant difference favoring olanzapine in time to all-cause medication discontinuation (1 RCT).
Perphenazine vs. quetiapine	No significant difference in HRQoL (1 RCT).
Perphenazine vs. risperidone	No significant difference in time to all-cause medication discontinuation (1 RCT) or HRQoL (1 RCT).
Perphenazine vs. ziprasidone	No significant difference in HRQoL (1 RCT).
Bipolar Disorder	
Haloperidol vs. aripiprazole	Significant difference in favor of haloperidol for relapse rates (1 RCT). No difference in remission (1 RCT) or response (2 RCTs) rates.
Haloperidol vs. olanzapine	No difference for relapse (1 RCT), response (1 RCT), or remission rates (1 RCT). Significant difference favoring haloperidol for HRQoL mental summary score (1 RCT). Significant difference favoring olanzapine for HRQoL physical summary score (1 RCT).
Haloperidol vs. quetiapine	No significant difference in response or remission rates (1 RCT).
Haloperidol vs. risperidone	No difference in response rates (1 RCT).
Haloperidol vs. ziprasidone	Significant difference favoring haloperidol for response rates (1 RCT). No difference for remission rates (1 RCT).

HRQoL = health-related quality of life; KQ = Key Question; RCT = randomized controlled trial
[a]Response rates were defined by authors of the primary studies and may have varied across trials.

Results for other outcomes were available for three head-to-head comparisons in studies of patients with bipolar disorder. Significant differences were found for health-related quality of life in one trial comparing haloperidol versus olanzapine: haloperidol was favored for the mental summary score, and olanzapine was favored for the physical summary score. One study showed a significant difference favoring haloperidol compared with ziprasidone for response rates.

KQ5: Subgroups

A total of 41 studies compared outcomes for predefined subgroups. Among the studies of patients with schizophrenia and schizophrenia-related psychoses, data were most often available for race and treatment resistance. The race most often examined was Asian. No notable differences were observed for the subgroups compared with the overall findings.

The only subgroup available for analysis in studies of patients with bipolar disorder was disorder subtype, specifically bipolar I and bipolar II. The results were consistent with the overall findings. A significant difference favored haloperidol compared with ziprasidone for core illness symptoms (YMRS) in patients with bipolar I disorder.

Results in the Context of Other Research

The results of this review are similar in some respects to another recent systematic review of SGAs versus FGAs, although the present review is broader in scope in terms of medications included, patient populations, and outcomes.[9] There were a number of methodological differences between the previous review and this one. The previous review included antipsychotics not approved by the FDA, restricted the analysis to only double-blinded trials, included only studies examining optimum SGA dosage and oral route of administration, and pooled data across efficacy outcome measures. The differences in the methodologies may have led to slightly different conclusions regarding individual SGAs.

The previous review compared nine SGAs (six of which are included in this report) with FGAs for overall efficacy (total symptom scores); positive, negative, and depressive symptoms; relapse; quality of life; EPS; weight gain; and sedation. The authors reported that the overall efficacy of the FDA-approved SGAs clozapine, olanzapine, and risperidone was better than that of FGAs. In terms of global ratings and total symptom scores, we found that clozapine was more efficacious than chlorpromazine but not haloperidol. We found that olanzapine performed better

than haloperidol on one of the three total symptom scores assessed. We found no differences between haloperidol and risperidone for the five total symptom scores reported. The previous review found that SGAs were not superior to FGAs regarding the negative symptoms. We found no difference in negative symptoms for haloperidol versus clozapine; however, we found evidence that olanzapine was more efficacious than haloperidol for negative symptoms, whereas the evidence for risperidone compared with haloperidol was mixed. In general, the findings for AEs were consistent between reviews, showing poorer safety profiles with respect to EPS for FGAs (specifically haloperidol) and more weight gain among the SGAs (in particular, olanzapine and risperidone).

One of the unique features of our review was the SoE assessments, which provide information on how confident we can be in the results of existing studies and how likely it is that the estimates of treatment effects will change with future research. In most cases, the SoE was insufficient or low, highlighting the likelihood that future research will change the estimates of effect and the need for a stronger evidence base to inform clinical practice.

Applicability

This report included studies that compared an individual FGA with an individual SGA. Placebo-controlled studies or studies comparing an FGA versus another FGA, or an SGA versus another SGA, were not included. Therefore, the evidence is focused on the comparative effectiveness of FGAs versus SGAs, but not on their effectiveness and safety compared with placebo or other active agents. Overall, there were 20 head-to-head comparisons across the relevant studies; however, within most comparisons there were few studies.

The focus of our review was adults age 18 to 64 years with schizophrenia, schizophrenia-related psychoses, or bipolar disorder. The average age across studies ranged from 21 to 50 years (median = 37 years [IQR, 33 to 41]). Most studies were highly selective in patient enrollment and included patients who (1) met strict diagnostic criteria for case definition, (2) had few comorbidities, and (3) used few or no concomitant medications. Older adults and the most seriously ill patients were underrepresented. Such highly selective criteria may increase the likelihood of drug benefit and decrease the likelihood of AE occurrence. Almost half the studies involved hospitalized patients (inpatient treatment) (62 of 125 studies) or mixed inpatient and outpatient populations (26 studies); relatively few studies examined only outpatient treatment populations (19 studies). As such, we judge the results of this report to be applicable to patients in outpatient and inpatient treatment settings.

Another factor that restricts the applicability is the limited duration of followup. Despite our efforts to identify long-term safety data from observational studies, only two retrospective cohort studies provided data for the minimum 2-year followup period.

Limitations of Existing Evidence

Inconsistency in treatment comparisons, outcomes, outcome measurement, and patient populations across studies makes it difficult to draw firm clinical conclusions. Few studies compared the same antipsychotic medications and dosage using similar measures; various scales and surrogate measures were used to assess efficacy for different outcomes and AEs. Consensus is needed regarding outcomes and measures used to assess outcomes. Additionally, functional outcomes and symptomatic outcomes (e.g., sedation, restlessness) were rarely and unequally reported throughout the trial reports, even though these outcomes are often vital to patient compliance.

A key limitation and challenge in synthesizing and interpreting this body of evidence is the issue of heterogeneous patient populations across and within studies, which is in part driven by the complex nature of these disorders and their course over time. The studies we included had very mixed populations with respect to disorder subtypes, comorbid drug or alcohol use, treatment resistance, and number of previous episodes. These variables may create differential response to treatment, and this has been the basis for recommendations around personalized medicine in this area.[8] We conducted extensive subgroup and sensitivity analyses to explore these varying features. The results of subgroup analyses should be interpreted as hypothesis generating rather than hypothesis confirming. Our findings may provide some information to make treatment decisions for individual patients but need to be confirmed in future research.

An additional limitation and challenge of synthesis in this area is that characteristics of the research may have changed over time, including drug doses (e.g., lower doses of FGAs in more recent studies) and patient populations (e.g., fewer patients already exposed to FGAs or proven treatment resistant to FGAs in recent studies).

An important limitation of this review and other systematic reviews is the design and quality of the primary included studies. The majority of studies providing data for this report were RCTs (n = 123); however, most were designed as superiority trials, often with an a priori hypothesis that the SGA would be more efficacious.[10] The individual studies and, in many cases, the pooled results may not have sufficient power to detect equivalence or noninferiority between drugs. Further, all of the included trials had an unclear risk of bias (n = 78, 63 percent) or high risk of bias (n = 45, 37 percent). Of note, few trials (n = 20) reported blinding study investigators and participants (26 percent had unclear reporting), which is important in interpreting the results because lack of blinding has been shown to produce exaggerated treatment effects.[9]

Future Research

More longitudinal research is needed on the long-term comparative effectiveness of FGAs versus SGAs. Only two cohort studies were identified for this review that examined serious AEs with long-term antipsychotic use; these studies examined only two serious events: tardive dyskinesia and mortality rates. The SoE for these AEs was insufficient to draw conclusions. Studies examining the naturalistic and long-term efficacy, and particularly the safety, of antipsychotics over the course of several years and across a number of important AEs are required. Further, consensus is needed on the most important comparisons of FGAs versus SGAs for future studies; the most frequent FGA in the studies to date was haloperidol.

Short- and long-term evaluations of the effectiveness of FGAs and SGAs with patient subpopulations, including patients with medical and neurological comorbidities, are needed. Further, there is a need for studies investigating how drug dose, age, and other factors, such as comorbidities, influence the occurrence of serious AEs, which would help estimate possible risks in specific patient populations.

Future studies should examine functional naturalistic outcomes that are important to patients. These outcomes include health-related quality of life and other patient-reported outcomes, relationships, academic and occupational performance, and legal interactions.

Conclusions

This report provides a comprehensive synthesis of the evidence on the comparative effectiveness and safety of individual FDA-approved FGAs compared with individual FDA-approved SGAs. The report provides extensive details in terms of study characteristics and methodological features, which may help inform individual treatment decisions.

Numerous studies provided data on core illness symptoms; however, many different scales were used to assess outcomes, which limited the quantitative pooling of data. Few notable differences of clinical importance were identified. The SoE was low or insufficient for most comparisons, suggesting that future research is likely to change the results and change our confidence in the results.

Data on the relative effectiveness for functional outcomes, health care system utilization, and other outcomes were generally sparse. The variety of functional measures assessed across studies precluded firm conclusions regarding the overall effectiveness of individual drugs in terms of patient functioning. Few studies reported on health care system utilization or patient-important outcomes. Where health-related quality of life was assessed, no differences were found.

We included cohort studies with a minimum followup of 2 years in order to identify the AEs of most clinical importance, including diabetes mellitus, mortality, tardive dyskinesia, and major metabolic syndrome. Only two studies with long-term followup were identified; hence, evidence on these important AEs is limited and urgently needed. A variety of AEs associated with numerous physiological systems were reported. The AEs most often reported involved EPS, which occurred more frequently for FGAs, particularly haloperidol, than for SGAs. Long-term longitudinal studies of at least 2-year duration are needed to detect important differences in the relative safety profile of individual FGAs and SGAs.

The evidence for important subgroups was limited. The most frequently examined subgroups were race and treatment resistance. There were no notable differences in outcomes for these subgroups compared with the overall results.

In summary, data on the comparative effectiveness of individual FGAs and SGAs precluded drawing firm conclusions for outcomes that are directly relevant to front-line clinical decisions. Overall, there were few significant differences of clinical importance. Outcomes potentially important to patients were rarely assessed. Finally, data on long-term safety are lacking and urgently needed.

References

1. Meltzer HY, Huang M. In vivo actions of atypical antipsychotic drug on serotonergic and dopaminergic systems. Prog Brain Res. 2008;172:177-97. PMID: 18772033.
2. Lopez-Gil X, Artigas F, Adell A. Unraveling monoamine receptors involved in the action of typical and atypical antipsychotics on glutamatergic and serotonergic transmission in prefrontal cortex. Curr Pharm Des. 2010;16(5):502-15. PMID: 19909228.
3. Aparasu RR, Bhatara V. Antipsychotic use and expenditure in the United States. Psychiatr Serv. 2006;57(12):1693. PMID: 17158480.
4. Ascher-Svanum H, Nyhuis AW, Stauffer V, et al. Reasons for discontinuation and continuation of antipsychotics in the treatment of schizophrenia from patient and clinician perspectives. Curr Med Res Opin. 2010;26(10):2403-10. PMID: 20812791.
5. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1. The Cochrane Collaboration; 2008.
6. Eden J, Levit L, Berg A, et al., eds. Finding What Works in Health Care: Standards for Systematic Reviews. Committee on Standards for Systematic Reviews of Comparative Effectiveness Research, Board on Health Care Services. Washington: The National Academies Press; 2011.
7. Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomized Studies in Meta-Analyses. Department of Epidemiology & Community Medicine, University of Ottawa, Canada. 2009. www.ohri.ca/programs/clinical_epidemiology/oxford.htm.
8. Clark SL, Adkins DE, van den Oord EJ. Analysis of efficacy and side effects in CATIE demonstrates drug response subgroups and potential for personalized medicine. Schizophr Res. 2011;132(2-3):114-20. PMID: 121872442.
9. Leucht S, Corves C, Arbter D, et al. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. Lancet. 2009;373(9657):31-41. PMID: 19058842.
10. Lieberman JA, Stroup TS. The NIMH-CATIE Schizophrenia Study: what did we learn? Am J Psychiatry. 2011;168(8):770-5. PMID: 121813492.

Full Report

This executive summary is part of the following document: Abou-Setta AM, Mousavi SS, Spooner C, Schouten JR, Pasichnyk D, Armijo-Olivo S, Beaith A, Seida JC, Dursun S, Newton AS, Hartling L. First-Generation Versus Second-Generation Antipsychotics in Adults: Comparative Effectiveness. Comparative Effectiveness Review No. 63. (Prepared by the University of Alberta Evidence-based Practice Center under Contract No. 290-2007-10021.) AHRQ Publication No. 12-EHC054-EF. Rockville, MD: Agency for Healthcare Research and Quality; August 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

For More Copies

For more copies of First-Generation Versus Second-Generation Antipsychotics in Adults: Comparative Effectiveness: Comparative Effectiveness Review Executive Summary No. 63 (AHRQ Pub. No. 12-EHC054-1), please call the AHRQ Clearinghouse at 1-800-358-9295 or email ahrqpubs@ahrq.gov.

AHRQ Pub. No. 12-EHC054-1
August 2012