Effective Health Care Program

Future Research Needs Paper
Number 13

Future Research Needs for First- and Second-Generation Antipsychotics for Children and Young Adults

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

Future Research Needs Paper

Number 13

Future Research Needs for First- and Second-Generation Antipsychotics for Children and Young Adults

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290 2007 10056 I

Prepared by:
RTI-UNC Evidence-based Practice Center
Research Triangle Park, NC

Investigators:
Robert Christian, M.D.
Lissette Saavedra, Ph.D.
Bradley N. Gaynes, M.D., M.P.H.
Brian Sheitman, M.D.
Roberta C.M. Wines, M.P.H.
Daniel E. Jonas, M.D., M.P.H.
Meera Viswanathan, Ph.D.
Alan R. Ellis, M.S.W.
Carol Woodell, B.S.P.H.
Timothy S. Carey, M.D., M.P.H.

AHRQ Publication No. 12-EHC042-EF
February 2012

This report is based on research conducted by the RTI-UNC Evidence-based Practice Center (EPC) under contract to the Agency for healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290 2007 10056 I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care researchers and funders of research make well-informed decisions in designing and funding research and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of scientific judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical research and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Christian R, Saavedra L, Gaynes BN, Sheitman B, Wines RCM, Jonas DE, Viswanathan M, Ellis AR, Woodell C, Carey TS. Future Research Needs for First- and Second-Generation Antipsychotics for Children and Young Adults. Future Research Needs Paper No. 13. (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290 2007 10056 I.) Rockville, MD: Agency for Healthcare Research and Quality; February 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

An important part of evidence reports is to not only synthesize the evidence, but also to identify the gaps in evidence that limited the ability to answer the systematic review questions. AHRQ supports EPCs to work with various stakeholders to identify and prioritize the future research that are needed by decisionmakers. This information is provided for researchers and funders of research in these Future Research Needs papers. These papers are made available for public comment and use and may be revised.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The evidence reports undergo public comment prior to their release as a final report.

We welcome comments on this Future Research Needs document. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Elisabeth Kato, M.D., M.R.P.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

We want to thank the members of our stakeholder panel who gave generously of their time. Their expertise and enthusiasm helped drive this project. We also thank Jennifer Seida, M.P.H., and her colleagues at the University of Alberta EPC, authors of the 2010 draft AHRQ comparative effectiveness review. Thanks also to Mattias Jonsson, Maria Tobin, and Shannon Brode, M.P.H., from the UNC Sheps Center for Health Services Research, for their work in developing our online prioritization tools; Christiane Voisin, M.S.L.S., also from the UNC Sheps Center for Health Services Research, for library support; and our Task Order Officers at AHRQ, Sonia Tyutyulkova, M.D., Ph.D., and Elisabeth Kato, M.D., M.R.P., for support and guidance. We also express our gratitude to RTI staff Loraine Monroe, Publications Specialist, and Sharon Barrell, Editor.

Contributors

Rebecca Bitsko, Ph.D.
CDC National Center on Birth Defects and Development Disabilities
Atlanta, GA

Teri Brister, Ph.D.
Programs for Young Families
Brandon, MS

Jana Davidson, M.D., F.R.C.
BC Mental Health & Addictions Research Institute
Vancouver, BC Canada

Debra Dihoff, M.A.
National Alliance on Mental Illness
Raleigh, NC

Tiffany Farchione, M.D.
Food and Drug Administration, Division of Psychiatry Products
Silver Spring, MD

Laurence Greenhill, M.D.
NYS Psychiatric Institute
New York, NY

Fay Kagan, M.D.
Hathaway Children's Services
Sylmar, CA 91392

Penelope Knapp, M.D.
Department of Psychiatry and Behavioral Sciences
UC Davis M.I.N.D. Institute
Sacramento, CA

Laurel Leslie, M.D., M.P.H.
Tufts Clinical and Translational Science Institute
Boston, MA

Mark Olfson, M.D., M.P.H.
Columbia University Medical Center
New York, NY

Benedetto Vitiello, M.D.
National Institute of Mental Health
Bethesda, MD

Julie Zito, Ph.D.
University of Maryland School of Pharmacy
Baltimore, MD

Contents

Tables

Figures

Appendixes

Executive Summary

This future research needs (FRN) report is based on an Agency for Healthcare Research and Quality (AHRQ) comparative effectiveness review (CER) titled "First- and Second-Generation Antipsychotics for Children and Young Adults."[1] The purpose of the CER was to review and synthesize the evidence regarding the benefits and harms of first- and second-generation antipsychotics (FGAs and SGAs) (see Tables A and B) for the treatment of various psychiatric and behavioral conditions in individuals 24 years of age or younger. Table C shows the key questions from this CER.

Table A. Food and Drug Administration-approved first-generation antipsychotics

Generic Name	Indications	Age Group for Which Approved
Chlorpromazine	Schizophrenia	Adults and children (1–12 years)
	Bipolar disorder (mania)	
	Hyperactivity	
	Severe behavioral problems	
Droperidol	Agitation	Adults and children
Fluphenazine	Psychotic disorders	Adults
Haloperidol	Schizophrenia	Adults
	Tourette syndrome	
	Hyperactivity	
	Severe childhood behavioral problems	
Loxapine	Schizophrenia	Adults and children ≥12 years
Perphenazine	Schizophrenia	Adults and children ≥12 years
Pimozide	Tourette syndrome	Adults and children ≥12 years
Prochlorperazine	Schizophrenia	Adults and children >2 years and >20 pounds
	Generalized nonpsychotic anxiety	Adults
Thiothixene	Schizophrenia	Adults and children ≥12 years
Thioridazine	Schizophrenia	Adults and children
Trifluoperazine	Schizophrenia	Adults and children ≥6 years
	Generalized nonpsychotic anxiety	Adults

Table B. Food and Drug Administration-approved second-generation antipsychotics

Generic Name	Indications	Age Group for Which Approved
Aripiprazole	Schizophrenia	Adults and adolescents (13–17 years)
	Bipolar disorder (manic/mixed) monotherapy or adjunctive to lithium or valproate	Adults and children (10–17 years)
	Adjunctive treatment of major depressive disorder	Adults
	Irritability Associated with autistic disorder	Children (6–17 years)
	Acute treatment of agitation	Adults
Asenapine	Acute schizophrenia	Adults
	Bipolar disorder type 1 (manic/mixed)	
Clozapine	Treatment resistant schizophrenia	Adults
	Reduce the risk of suicidal behavior in younger patients with schizophrenia.	
Iloperidone	Acute schizophrenia	Adults
Olanzapine	Schizophrenia Bipolar disorder (manic/mixed)	Adults and adolescents (13–17 years)
	Bipolar disorder	Adults
	Treatment resistant depression	
	Agitation associated with schizophrenia and bipolar I mania	
Paliperidone	Schizophrenia Schizoaffective disorder	Adults
Quetiapine	Schizophrenia	Adults and adolescents (13–17 years)
	Bipolar disorder (acute manic)	Adults, children, and adolescents (10–17 years)
	Bipolar disorder (depression)	Adults
	Bipolar disorder (maintenance)	
	Adjunctive therapy for major depressive disorder	
Risperidone	Schizophrenia	Adults and adolescents (13–17 years)
	Bipolar disorder (manic/mixed)	Adults and adolescents (10–17 years)
	Irritability associated with autism	Children (5–16 years)
Ziprasidone	Schizophrenia	Adults
	Bipolar disorder (manic/mixed)	
	Bipolar disorder (maintenance)	
	Acute agitation in patients with schizophrenia	

Table C. Key Questions from the CER[a]

Number	Key Question
1	What is the comparative efficacy or effectiveness of first-generation antipsychotics (FGAs) and second generation antipsychotics (SGAs) for treating disorder-specific and nonspecific symptoms? Included disorders: pervasive developmental disorders, including autistic disorder, Rett syndrome, childhood disintegrative disorder, Asperger syndrome, and pervasive developmental disorder not otherwise specified; Attention deficit hyperactivity disorder (ADHD) and disruptive behavior disorders, including conduct disorder, oppositional defiant disorder, and disruptive behavior disorder not otherwise specified; bipolar disorder, including manic or depressive phases, rapid cycling, and mixed states; schizophrenia and schizophrenia-related psychoses, including schizoaffective disorder and drug-induced psychosis; Tourette syndrome; obsessive-compulsive disorder; post-traumatic stress disorder; anorexia nervosa; and behavioral symptoms, including aggression, agitation, anxiety, behavioral dyscontrol, irritability, mood lability, self-injurious behaviors, and sleep disorders.
2	Do FGAs and SGAs differ in the following medication-associated adverse events: overall adverse events; specific adverse events; withdrawals and time-to-withdrawal due to adverse events; and persistence and reversibility of adverse events?
3	Do FGAs and SGAs differ in the following other short- and long-term outcomes (short-term outcomes are defined as outcomes occurring within 6 months; long-term outcomes are defined as outcomes occurring after 6 months): response rate with corresponding dose, duration of response, remission, relapse, speed of response, and time to discontinuation of medication; growth and maturation; cognitive and emotional development; suicide-related behaviors (including ideation) or death by suicide; medication adherence and persistence; school performance and attendance; work-related functional capacity; patient insight into illness; patient-, parent-, or care provider–reported outcomes, including levels of physical activity or inactivity, and diet (e.g., caloric intake, food preferences); health-related quality of life; legal or justice system interaction (e.g., arrests, detention); health care system utilization (e.g., protective services, social services); and "outcomes that matter" to children, young adults, and their families?
4	Do the efficacy and risks of FGAs and SGAs vary in differing subpopulations including sex, age group, race, comorbidities (including substance abuse and ADHD), co-treatment (versus monotherapy), first episode psychosis versus treatment in context of history of prior episodes, duration of illness, and treatment history?

For Key Question (KQ) 1, with few exceptions, the Comparative Effectiveness Review (CER)[1] reported that the evidence comparing FGAs with SGAs was insufficient to allow for conclusions. Where an interclass difference was noted, the strength of evidence (SOE) grade was reported as low. The CER reported insufficient studies to allow for within-class efficacy/effectiveness comparisons with several exceptions. In these exceptions, either no difference was noted (supported by low SOE), or the difference found was supported with low SOE. The CER reported low to moderate SOE to support SGAs as a class over placebo for certain outcome-disorder pairs as summarized in Table D.

Figure A. Analytic framework depicting relationships between key questions, populations, interventions, outcomes, and components of evidence gaps

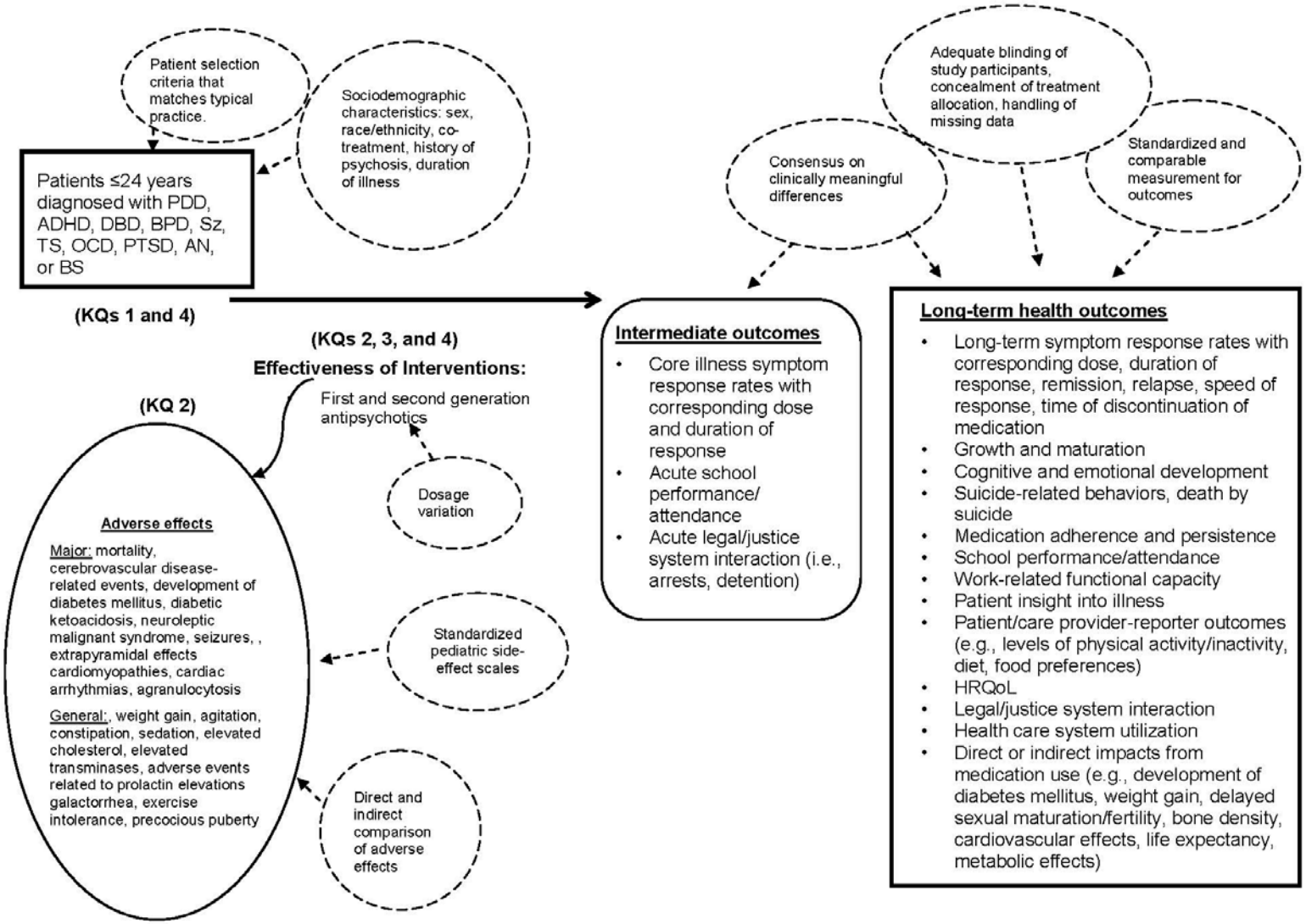


Table D. Summary SOE for SGAs vs. placebo

	Pervasive Developmental Disorders	ADHD/Disruptive Behavior Disorders	Bipolar Disorder	Schizophrenia	Tourette Syndrome
Outcome					
Behavioral symptoms	*Low SOE*[a]	*Moderate SOE*[b]	*NA*	*NA*	*NA*
Mania	*NA*	*NA*	*Low SOE*	*Insufficient Evidence*	*NA*
Clinical impressions	*No difference (low SOE)*	*Moderate SOE*	*Moderate SOE*	*Moderate SOE*	*Insufficient Evidence*
Positive symptoms	*NA*	*NA*	*NA*	*Moderate SOE*	*NA*
Tic severity	*NA*	*NA*	*NA*	*NA*	*Moderate SOE*

[a]Behavioral symptoms defined by the Aberrant Behavior Checklist (ABC) or the Childhood Autism Rating Scale (CARS)
[b]Behavioral symptoms defined by the ABC, Behavior Problem Inventory (BPI), or the Nisonger Child Behavior Rating Scale (NCBRF)

Regarding adverse events for FGAs compared with SGAs (KQ 2), SGAs were significantly favored over haloperidol for extrapyramidal symptoms (low SOE). Haloperidol was favored over olanzapine for body composition (low SOE). All other adverse events were not significant (low SOE) or had insufficient evidence. For all comparisons of different FGAs or FGA with placebo, there was insufficient evidence to draw a conclusion for adverse events.

For KQ 3, the evidence was rated as insufficient to draw conclusions for health-related quality of life; legal interactions; and other patient-, parent-, or care provider–reported outcomes for all conditions. Short- and long-term outcomes were reported in nine studies examining pervasive developmental disorders and in eight studies examining ADHD and disruptive behavior disorders.

Evidence and conclusions for KQ 4 are based on studies that compared outcomes across various patient subpopulations. Few studies identified differences in the results across subpopulations. Few associations between the patient or clinical variables and outcomes were supported by more than one study. Few studies reported on key health outcomes, and the duration of most studies was short, limiting conclusions about outcomes such as health-related quality of life, social and occupational functioning, and other long-term effectiveness outcomes, parent- or care provider–reported outcomes, and long-term effects of acute adverse events.

Methods

Identifying Evidence Gaps and Developing PICOTS

We developed a preliminary list of evidence gaps based on SOE and other information gleaned from the results and limitations sections of the CER. Our main focus was on capturing topics with insufficient information. We then applied the PICOTS from the CER[1] inclusion/exclusion criteria and developed an analytic framework (Figure A) to show the relationships between the evidence gaps, PICOTS, and key questions.

We identified a broad range of potential stakeholders, who represented one or more perspectives, including patient and family advocacy groups; health care providers, including diagnosticians and treatment experts; educators of preschool and school-age children; researchers, including those with experience in pharmacology, psychiatry, education, epidemiology, and screening tools; state policymakers and payers of services; professional provider and educator organizations; individuals with knowledge of health services delivery

systems; and research funders. The stakeholders contributed to this project via email, conference calls, and online prioritization activities. We scheduled two rounds of conference calls using GoToMeeting® and two rounds of an online prioritization with the stakeholder group.

The stakeholders received a preliminary list of evidence gaps and an analytic framework showing the relationships between the key questions, PICOTS elements, and components of the evidence gaps as part of their orientation materials. During the first call, we invited stakeholders to comment on and make contributions to the list of evidence gaps. We also reviewed a list of ongoing research studies, developed by the project team through searching online research registries, to help identify new data that might be pertinent to evidence gaps. After receiving stakeholder input, project investigators revised the list of evidence gaps and applied the PICOTS elements to the new and revised gaps.

Criteria for Prioritizing Evidence Gaps

The project team developed an online prioritization tool and invited the stakeholders to rank the revised list of evidence gaps in order of priority to produce an upper tier of evidence gaps. To complement the stakeholders' own perspectives during the prioritization process, we provided the stakeholders with a modified version of the Effective Health Care (EHC) Program Selection Criteria.

Engaging Stakeholders to Prioritize Evidence and Develop Research Needs

During the second call, we reviewed and discussed the results of the prioritization exercise, finalized the upper tier of evidence gaps, and asked stakeholders for feedback on the PICOTS and thoughts on potential research designs for these upper-tier gaps.

Following this discussion, we applied the updated PICOTS framework to the upper-tier evidence gaps and translated them into research questions. We then invited the stakeholders to reprioritize only the upper tier of the evidence gaps using the online prioritization tool to create a final list of prioritized research needs. The final list of prioritized research needs was not shared with the stakeholders until the public comment period of this report.

Developing Research Questions and Determining Potential Research Designs

We applied study design considerations including issues of validity; resources required; ability to recruit subjects or obtain data; and potential ethical, legal, or social issues, to the top-ranked research needs. We also performed sample power analyses to help identify pragmatic barriers of the potential designs. We did not ask stakeholders to rank study designs or provide input to the proposed study designs.

Results

From the original 16 evidence gaps, of which the stakeholders prioritized 14, the stakeholders deemed 6 as the highest-priority research needs after 2 rounds of prioritization. In this executive summary, we present each research need and the research team's initial views of the potential study designs that could be used to address the priority research need. A discussion of the potential study design considerations may be found in the full FRN report along with a table describing additional characteristics and study design considerations. The six upper-tier research needs in Table E are not ranked.

Table E. Six high-priority research needs identified by stakeholders

Research Need

What is the long-term comparative effectiveness between and within classes of antipsychotics as measured in outcomes related to the disorder of interest, its co-morbidities, associated behavioral features, social-occupational outcomes, and outcomes identified as important by patients and their families?

Design considerations

- Nonrandomized comparative designs
- Observational studies
    - Prospective cohort design
    - Retrospective Cohort Design

Research Need

What are the comparative long-term risks of medication exposure between and within antipsychotic classes?

Design considerations

- Nonrandomized comparative designs
- Observational studies
    - Prospective cohort design
    - Retrospective cohort design

Research Need

What is the efficacy and effectiveness of first or second generation antipsychotics for adolescents and young adults with schizophrenia, in the following outcome domains: core features of the disorder, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient- and parent-reported outcomes, those related to high risk behaviors, and suicide-related behavior?

Design considerations

- Nonrandomized comparative designs
- Observational studies
    - Prospective cohort design
    - Retrospective cohort design
    - Case control study

Research Need

Are there subgroups of patients, based on baseline demographic/clinical characteristics or physical and/or mental health co-morbidities, for which first and second generation antipsychotics differ in efficacy, effectiveness, or frequency of adverse events?

Design considerations

- Nonrandomized comparative design
- Observational studies
    - Prospective cohort design
    - Retrospective cohort design
    - Meta-analysis of individual patient data

Research Need

What is the efficacy and effectiveness of first or second generation antipsychotics for individuals with attention deficit hyperactivity disorder and disruptive behavior disorders in the following outcome domains: core ADHD symptoms, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient and parent-reported outcomes, outcomes related to high risk behaviors, and suicide-related behavior?

Design considerations

- Randomized comparative designs
- Nonrandomized comparative designs
- Observational studies
    - Prospective cohort design
    - Retrospective cohort design
    - Case control study
    - Meta-analysis of individual patient data

Table E. Six high-priority research needs identified by stakeholders (continued)

Research Need

What is the efficacy and effectiveness of first or second generation antipsychotics for adolescents and young adults with bipolar disorder, in the following outcome domains: core features of the disorder, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient- and parent-reported outcomes, outcomes related to high risk behaviors, and suicide-related behavior?
 	
Design considerations
• Randomized comparative designs
• Nonrandomized comparative designs
• Observational studies
  o Prospective cohort design
  o Retrospective cohort design
  o Case control study
  o Meta-analysis of individual patient data

Discussion

We worked with a group of stakeholders to ultimately identify six high priority research needs in the area of antipsychotic usage in youth. The stakeholders prioritized general medication safety and effectiveness issues across disorders over disorder-specific medication effectiveness gaps.

Although randomized controlled trials (RCTs) may be an ideal study design for many effectiveness questions, they are not viable for most of the high priority research needs because of sample size and length of followup demands. Prospective cohort designs that follow youth with a range of mental health disorders could be valuable for comparative effectiveness and safety research but are hampered by cost and logistical concerns. Secondary data analysis techniques are limited by population heterogeneity, short length of followup, and lack of appropriate measures in source trials. In some situations where raw data could be shared among investigators, meta-analysis of individual patient data could be considered. Lastly, for some questions, registries with linkages to clinical data sets could be a lower cost approach, allowing both prospective and retrospective evaluations, but such efforts are nascent in mental health and have confidentiality and methodological barriers.

Despite the aim of the CER[1] to evaluate both antipsychotic classes, stakeholders pointed out that future attempts to use observational designs to compare FGAs with SGAs would be hampered by the relatively low rate at which FGAs are used to treat youth. Even in the context of RCTs, feasibility and applicability issues may be strong barriers for FGA evaluation for most disorders.

There are limitations and challenges related to the future research needs process. Stakeholder input was essential, but scheduling challenges led to incomplete participation from some members. Further, conference call time constraints may have led to certain opinions not being expressed. To accommodate these challenges, we provided opportunities for stakeholders to provide feedback by email, but we did not speak with stakeholders by telephone individually. Inherent in the process is a challenging tension between the need to develop a list of digestible evidence gaps for a diverse group of stakeholders and the need to remain faithful to the purpose, findings, and intent of the original CER on antipsychotics in youth.

Conclusions

Overall, the stakeholders demonstrated engagement in our discussions of research challenges in the field and were able to perform the ranking process without difficulty. The six high priority

research needs included a broad range of issues cutting across disorders, key clinical outcomes, safety outcomes, and methodological concerns. PICOTS development aided our consideration of study design issues, and our sample power analyses demonstrated the pragmatic barriers that many of the potential designs will present. Although large long-term multisite clinical trials may be the gold standard to assess many of the questions of importance, issues of feasibility have greatly limited the number of such large pragmatic trials in mental health to date. Large prospective cohort studies of youth exposed to antipsychotics may be viable and offer considerable analytic flexibility, but they are also costly. Patient registries with linkages to clinical datasets may allow for more efficient evaluation of some questions with advanced analysis methods, but the infrastructure for this needs considerable investment, and its development may face considerable hurdles relating to information privacy. Meta-analysis of existing trials data and of individual patient data may prove helpful but will likely be limited to evaluation of specific shorter-term outcomes. Despite its limitations, the structured process used in this project may prove to be an effective way of reaching relative consensus on research priorities in this broad and complex topic area.

Reference

1. Seida JC, Schouten JR, Mousavi SS, et al. First- and Second-Generation Antipsychotics for Children and Young Adults. Comparative Effectiveness Review No. 39. (Prepared by the University of Alberta Evidence-based Practice Center under Contract No. 290-2007-10021.) AHRQ Publication No. 11(12)-EHC077-EF. Rockville, MD. Agency for Healthcare Research and Quality. February 2012.

Background

Objective

The purpose of this future research needs (FRN) report is to develop a list of stakeholders' research needs related to the comparative effectiveness of first- and second-generation antipsychotics (FGAs and SGAs) in pediatric and young adult populations.

Context

This FRN report is based on an Agency for Healthcare Research and Quality (AHRQ) comparative effectiveness review (CER) titled "First- and Second-Generation Antipsychotics for Children and Young Adults."[1] We reviewed the CER in draft form. The purpose of the CER was to review and synthesize the evidence regarding the benefits and harms of FGAs and SGAs for the treatment of various psychiatric and behavioral conditions in children and young adults 24 years of age or younger.

FGAs and SGAs are commonly categorized into two classes (Appendix A includes lists of FGAs and SGAs). FGAs, also known as typical antipsychotics, were developed in the 1950s (Appendix A, Table A-1). FGAs are used to treat psychotic symptoms such as auditory and visual hallucinations and delusions through several proposed mechanisms, including through the blockade of dopamine neuro-receptors. FGAs are associated with various adverse effects. These side effects include extra-pyramidal symptoms (EPS). EPS is a group of movement disorders, including acute dystonic reactions *(severe spasms of various muscle groups*), akathisia (*a feeling of motor restlessness*), pseudo-parkinsonism (*medication-induced motor slowness and rigidity*), and tardive dyskinesia (*repetitive low amplitude movements, most often of facial muscles, insidious and chronic in nature*). The most severe antipsychotic-associated potential side effect is neuroleptic malignant syndrome (NMS), characterized by hyperthermia, rigidity, rhabdomyolysis, renal failure, delirium, cardiovascular instability, and death. SGAs, also known as atypical antipsychotics, emerged in the 1980s. SGAs (Appendix A, Table A-2) are generally thought to have a lower risk of EPS.[2,3] The risk of NMS is rare for both medication classes, and researchers are uncertain about whether there is an intra-class risk difference for NMS.[4] However, SGAs are associated with a higher risk of a range of metabolic side effects, including weight gain; dyslipidemia; insulin resistance; the development of type 2 diabetes; and, rarely, hyperglycemic coma.[2]

The review[1] was prompted by the observation that the use of antipsychotics, particularly SGAs, for children and adolescents has increased markedly during the past 20 years.[5-9] Prescribing antipsychotics to the pediatric population is controversial because of a relative lack of high-quality and longitudinal studies on which to base clinical practice recommendations. For the majority of antipsychotic drugs, approved indications in the United States are restricted to the treatment of childhood schizophrenia and bipolar disorders. See Appendix A for a list of antipsychotics and FDA-approved indications. The U.S. Food and Drug Administration (FDA) approved risperidone in 2006 and aripiprazole in 2009 for the treatment of irritability associated with autism. Off-label prescriptions are given to younger children for a range of indications including behavioral symptoms (e.g., aggression) that are related to diagnosable conditions (e.g., attention deficit hyperactivity disorder [ADHD]). In general, much prescribing of SGAs for children and adolescents does not appear to be guided by evidence of clinical benefit or risk of harms.[10]

Scope of Comparative Effectiveness Review

The CER[1] focused on the following key questions to address the issues presented above.

Key Question 1. What is the comparative efficacy or effectiveness of first generation antipsychotics (FGAs) and second generation antipsychotics (SGAs) for treating disorder-specific and nonspecific symptoms? Included disorders:

- Pervasive developmental disorders, including autistic disorder, Rett syndrome, childhood disintegrative disorder, Asperger syndrome, and pervasive developmental disorder not otherwise specified;
- Attention deficit hyperactivity disorder (ADHD) and disruptive behavior disorders, including conduct disorder, oppositional defiant disorder, and disruptive behavior disorder not otherwise specified;
- Bipolar disorder, including manic or depressive phases, rapid cycling, and mixed states;
- Schizophrenia and schizophrenia-related psychoses, including schizoaffective disorder and drug-induced psychosis;
- Tourette syndrome;
- Obsessive-compulsive disorder;
- Post-traumatic stress disorder;
- Anorexia nervosa; and
- Behavioral symptoms, including aggression, agitation, anxiety, behavioral dyscontrol, irritability, mood lability, self-injurious behaviors, and sleep disorders.

Key Question 2. Do FGAs and SGAs differ in the following medication-associated adverse events:

- Overall adverse events?
- Specific adverse events?
- Withdrawals and time to withdrawal due to adverse events?
- Persistence and reversibility of adverse events?

Key Question 3. Do FGAs and SGAs differ in the following other short- and long-term outcomes (short-term outcomes are defined as outcomes occurring within 6 months; long-term outcomes are defined as outcomes occurring after 6 months):

- Response rate with corresponding dose, duration of response, remission, relapse, speed of response, and time to discontinuation of medication?
- Growth and maturation?
- Cognitive and emotional development?
- Suicide-related behaviors (including ideation) or death by suicide?
- Medication adherence and persistence?
- School performance and attendance?
- Work-related functional capacity?
- Patient insight into illness?
- Patient-, parent-, or care provider–reported outcomes, including levels of physical activity or inactivity, and diet (e.g., caloric intake, food preferences)?
- Health-related quality of life?

- Legal or justice system interaction (e.g., arrests, detention)?
- Health care system utilization (e.g., protective services, social services)?
- "Outcomes that matter" to children, young adults, and their families?

Key Question 4. Do the efficacy and risks of FGAs and SGAs vary in differing subpopulations including

- Sex?
- Age group (<6 years [preschool], 6–12 years [preadolescent], 13–18 years [adolescent], 19–24 years [young adult])?
- Race?
- Comorbidities, including substance abuse and ADHD?
- Cotreatment versus monotherapy?
- First-episode psychosis versus treatment in context of history of prior episodes (related to schizophrenia)?
- Duration of illness?
- Treatment naïve versus history of previous antipsychotics use?

Findings of the CER

Results and conclusions from the CER are based on 140 included articles, of which 81 were unique studies.[1] The studies included 62 randomized controlled trials (RCTs), 2 nonrandomized controlled trials (NRCTs), and 17 cohort studies (9 prospective and 8 retrospective). The number of participants in the studies ranged from 8 to 335 (median=42). The mean age of study participants ranged from 4.0 to 21.5 years (median=13.6). Few studies included young adults ages 19 to 24 years. None of the included studies examined obsessive-compulsive disorder (OCD), post-traumatic stress disorder, or anorexia nervosa. Overall, 38 studies provided head-to-head evidence on a total of 19 comparisons of different antipsychotics. In addition, 17 studies compared different doses of the same antipsychotic, and 26 studies compared a single antipsychotic with placebo. The CER authors used the EPC GRADE[11] method to grade the strength of evidence (SOE) for specific outcomes as insufficient, low, moderate, or high.

For KQ 1, the authors of the CER[1] graded evidence directly comparing FGAs with SGAs and antipsychotics within each class as insufficient to draw conclusions, with the exception of studies examining some outcomes for pervasive developmental disorders (PDDs) and schizophrenia, which were graded low.

For most conditions, direct comparative studies meeting study selection criteria were lacking; as a consequence, the majority of the findings in the CER[1] involved indirect evidence that compared SGAs with placebo. The CER[1] reported that there were not sufficient studies to evaluate or compare either medication class for OCD, post-traumatic stress disorder, or eating disorders. Findings for other disorders are described below.

- PDDs: For PDDs, no statistically significant difference was observed between FGAs and SGAs for behavioral symptoms associated with PDDs based on two RCTS. SGAs were favored over placebo for behavioral symptoms associated with PDDs with low SOE but did not separate from placebo for clinical global impressions. The SOE for these findings was low.
- ADHD and disruptive behavior disorders (DBDs): Among those with ADHD and DBDs, SGAs were superior to placebo for a number of behavior symptoms and clinical global

impressions with a moderate SOE rating. There was no difference between SGAs and placebo for aggression or anxiety (low SOE).

- Bipolar disorder: For bipolar disorder, the CER reported that SGAs were favored over placebo for clinical global impressions (moderate SOE) and for the treatment of mania (low SOE). However, SGAs did not differ from placebo for depression symptoms in this group (low SOE).
- Schizophrenia: For schizophrenia, SGAs were favored over placebo for clinical global impressions and for positive and negative symptoms (moderate SOE). SGAs were favored over FGAs for clinical global impressions (low SOE), but no difference was found between the classes for positive and negative symptoms. Several within-class SGA comparisons did not differ on positive and negative symptoms or clinical global impressions (low SOE).
- Tourette syndrome: SGAs were favored over placebo for tics (moderate SOE).

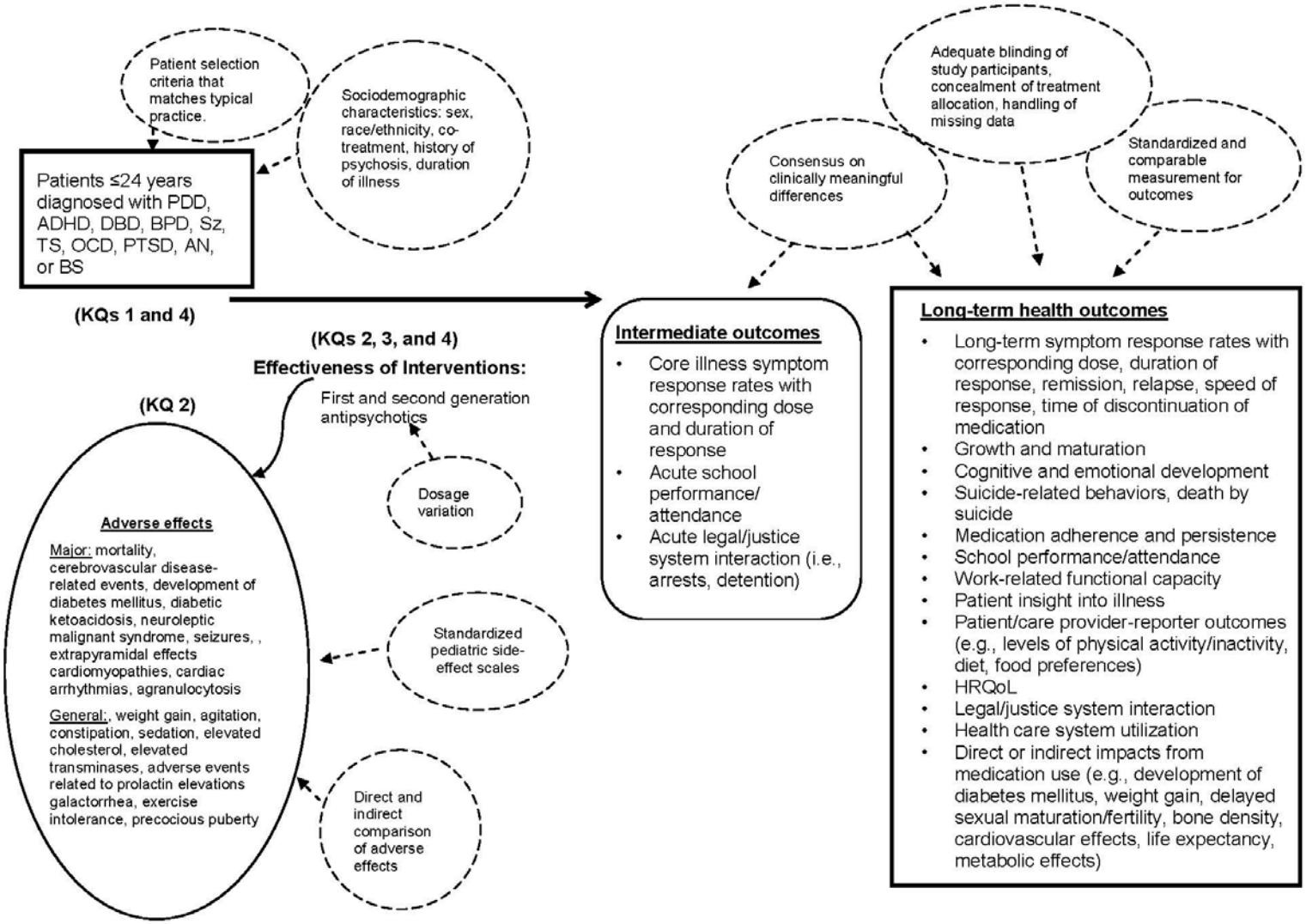
Regarding KQ 2, the CER evaluated the following adverse event outcomes: EPS, insulin resistance, prolactin-related and sexual adverse events, weight gain/body composition side effects, dyslipidemia, and sedation. SGAs were significantly favored over haloperidol for EPS (low SOE). Haloperidol, an FGA, was favored over olanzapine for body composition (low SOE). All other adverse events either lacked sufficient evidence for conclusions or were not significantly different for between-class (SGA vs. FGA) comparisons. This was also true for a number of within-class (SGA vs. SGA) comparisons (low SOE). For all direct comparisons of different FGAs with each other or for indirect comparisons of FGAs with placebo, evidence was insufficient to draw a conclusion for adverse events. For SGAs, adverse event profiles of various SGAs showed that risperidone was favored over olanzapine for dyslipidemia (moderate SOE). Olanzapine was favored over risperidone for prolactin-related events (moderate SOE). Both quetiapine and risperidone were favored over olanzapine for body composition (moderate SOE). For comparisons of SGAs with placebo for nearly all outcomes, the placebo group experienced significantly fewer adverse events than the group receiving SGAs. One exception was a significant effect in favor of aripiprazole over placebo for prolactin-related adverse events (moderate SOE).

For KQ 3, the evidence was rated as insufficient to draw conclusions about either short- or long-term outcomes for health-related quality of life; legal interactions; and other patient-, parent-, or care provider–reported outcomes for all conditions. Short- and long-term outcomes were reported in nine studies examining PDDs and in eight studies examining ADHD and DBDs. Medication adherence was not statistically or clinically different between SGAs and placebo for both conditions (low SOE). Eleven bipolar studies provided data for other outcomes. Medication adherence was statistically significantly better for placebo than for SGAs (RR, 2; 95% CI, 1.0 to 4.0) (low SOE). SGAs and placebo did not significantly differ for suicide-related behaviors (moderate SOE). A total of 22 studies provided data on a variety of short- and long-term outcomes for patients with schizophrenia and related psychosis. The studies found no significant difference in medication adherence between FGAs and SGAs, olanzapine and quetiapine, olanzapine and risperidone, or SGAs and placebo (low SOE). Similarly, SGAs and placebo did not differ in suicide-related behaviors (low SOE). Other outcomes were reported by four studies on Tourette syndrome and two studies on behavioral symptoms; the evidence was insufficient for all of the outcomes and comparisons examined in these studies.

For KQ 4, evidence and conclusions were based on 36 studies that compared outcomes across various patient subpopulations. Sex and age were examined most frequently. Overall, few studies identified differences in the results across subpopulations. Few associations between the patient or clinical variables and outcomes were supported by more than one study. Studies frequently had discordant conclusions in whether there was a significant association between subpopulations and outcomes and the direction of this association.

Because of several common methodological limitations in the studies that were included in the review, the authors concluded that much of the evidence is of low or insufficient strength and limits the ability to draw conclusions. The limitations noted by the CER authors include inadequate blinding of patients and outcome assessors, incomplete outcomes data because of loss to followup and inadequate handling of missing data (this includes cross-overs and the addition of other interventions), and lack of consistency and precision of results across studies because of the use of various scales and surrogate measures for outcomes and small sample sizes. Another limitation of the evidence is that approximately 80 percent of the trials were funded by pharmaceutical manufacturers. Further, few studies reported on key patient-centered health outcomes, and the duration of most studies was short (median of 8 weeks). These characteristics limit the ability to draw conclusions about outcomes such as health-related quality of life; social and occupational functioning; legal interactions; and patient-, parent-, or care provider–reported outcomes and long-term effects of more acute/subacute adverse events such as neuroleptic-induced weight gain on the development of diabetes, dyslipidemia, hypertension, and cardiovascular morbidity. The CER reported that the results of subgroup and regression analyses were often poorly described in the studies (e.g., few studies reported whether an association was significant), limiting the conclusions that could be drawn with respect to subpopulations. Figure 1 shows the relationships between the evidence gaps, PICOTS, and key questions.

Figure 1. Analytic framework depicting relationships between key questions, populations, interventions, outcomes, and components of evidence gaps



Methods

Evidence Gaps

The authors of the CER[1] identified several topics with little or no evidence on which to base conclusions. As part of the FRN identification and prioritization process, we developed a preliminary list of evidence gaps (Appendix B) based on SOE and other information gleaned from the results and limitations sections of the CER. Our main focus was on capturing topics with insufficient information. We then applied the PICOTS from the CER[1] inclusion/exclusion criteria and developed an analytic framework (Figure 1).

Identification of Stakeholders

We identified a broad range of potential stakeholders in consultation with our AHRQ Task Order Officer and during an internal planning meeting to which we invited representatives from the EPC that produced the draft review. During the meeting, we discussed potential stakeholders known to the experts on our team; we also asked the CER authors for names of stakeholders that they had used in generating the CER. Each potential stakeholder completed a statement of disclosure regarding conflicts of interest and had to be approved for participation by AHRQ prior to the first stakeholder call.

We sought a variety of individuals who represented one or more perspectives, including patient and family advocacy groups; health care providers, including diagnosticians and treatment experts; educators of preschool and school-age children; researchers, including those with experience in pharmacology, psychiatry, education, epidemiology, and screening tools; state policymakers and payers of services; professional provider and educator organizations; individuals with knowledge of health services delivery systems; and research funders. Some individuals could represent more than one stakeholder group. Our purpose in seeking these different perspectives was to produce a group that represented varied points of view on issues related to the use of FGAs and SGAs in pediatric and young adult populations.

We invited 17 individuals and organizations to participate in the stakeholder group. To ensure the inclusion of specific perspectives, we asked invited individuals who were not able to participate to refer us to someone else from their agency or organization. We provided potential stakeholders with a brief description of the project, including their role and the amount of time we expected them to contribute. Twelve stakeholders accepted our invitation.

Identification of Evidence Gaps

Throughout this report, the terms “evidence gap,” “research question,” and “research need” are defined as follows: Evidence gap is a gap in the CER that limited the ability to make conclusions on the questions asked. Research question is a statement of the purpose of a study. Research need is the top tier of prioritized evidence gaps identified through stakeholder engagement. Our process for identifying and prioritizing FRNs is shown in Appendix C.

After identifying the preliminary list of evidence gaps from the CER,[1] we presented the list (Appendix B) to the stakeholders as part of their orientation materials. During the first call, we invited stakeholders to comment on and make contributions to the list of evidence gaps. Project investigators revised the list of evidence gaps after receiving stakeholder input and applied the PICOTS elements to the new and revised gaps to ensure that each gap addressed one or more PICOTS elements and was within the scope of the CER.

Criteria for Prioritization

After the stakeholders had an opportunity to review, discuss, and revise the list of evidence gaps, we asked them to prioritize the gaps using specific criteria. The criteria included their own perspectives and the interests of their constituents along with the following criteria adapted from the Effective Health Care (EHC) Program Selection Criteria (Appendix D): importance, desirability of new research/duplication, and impact (Table 1). The proposed prioritization criteria emphasized the elements that were most applicable to stakeholders considering FRNs on a topic already under review by the EHC. We did not ask the stakeholders to consider the other two elements of the EHC Program Selection Criteria, appropriateness and feasibility, at this juncture.

Table 1. Suggested prioritization criteria, adapted from the AHRQ EHC Program Selection Criteria

Importance	Represents a significant disease burden; large proportion or priority population
	Is of high public interest; affects health care decision making, outcomes, or costs for a large proportion of the US population or for a priority population in particular
	Represents important uncertainty for decision makers
	Incorporates issues around both clinical benefits and potential clinical harms
	Represents important variation in clinical care, or controversy in what constitutes appropriate clinical care
	Represents high costs due to common use, to high unit costs, or to high associated costs to consumers, to patients, to health care systems, or to payers
Desirability of New Research/ Duplication	Would not be redundant (the proposed topic is not already covered by available or soon-to-be available high-quality systematic review by AHRQ or others)
Potential Impact	Potential for significant health impact, significant economic impact, potential change, potential risk from inaction, addressing inequities and vulnerable populations, and/or addressing a topic with clear implications for resolving important dilemmas in health and health care decisions made by one or more stakeholder groups.

Before the prioritization exercise, the project team prepared a list of ongoing studies for the stakeholders. The purpose of this list was to help the stakeholders identify any potential evidence gaps that might be addressed by current research and to use that information to consider whether new research would be duplicative. We developed the list by reviewing titles and short descriptions of research studies obtained through searching online research registries. We searched clinicaltrials.gov, HSRProj, NIH RePORTER, and the International Clinical Trials Registry Project (ICTRP) to find relevant ongoing or recently completed research. Search strategies and the list of ongoing studies presented to the stakeholders are shown in Appendix E. Two people from the research team independently reviewed each ongoing study title and abstract and applied the inclusion and exclusion criteria from the draft review to determine which studies might have met the inclusion criteria for the CER if they had been completed and published results before the CER cutoff date. During the first stakeholder call, the stakeholders reviewed the list of ongoing studies and provided feedback on its completeness and the relevance of the research studies to the existing evidence gaps.

Engagement of Stakeholders, Researchers, and Funders

The stakeholders contributed to this FRN project via email, conference calls, and Web-based prioritization activities. We planned two conference calls using GoToMeeting® and two rounds of a Web-based prioritization exercise (Appendix F) with the stakeholder group. Prior to the first call, we sent orientation materials to the stakeholders. These materials included the executive summary of the draft review and a description of the FRN project and its goals. Stakeholders also received a meeting packet that included the preliminary list of evidence gaps, the modified EHC Program Selection Criteria, and a list of ongoing studies that were reviewed to determine if they met the inclusion criteria for the review.

After the first and second conference calls, we accepted comments and edits via email from the stakeholders on the list of evidence gaps and PICOTS documents that were discussed during the respective calls. These comments were reflected in changes to the pertinent documents and reflected in the meeting summaries.

Between the first and second calls, we asked the stakeholders to prioritize the revised list of evidence gaps. This was the first round of prioritization, and it was based on the complete list of initial evidence gaps. The second round of prioritization occurred after the second call and consisted of only the top-tier evidence gaps (i.e. those that were ranked the highest in the first round). The Web-based prioritization exercises allowed each stakeholder to distribute a limited number of star-shaped indicators (referred to as stars in the remainder of this document) to those they viewed as the highest-priority gaps. In the first round, we gave each stakeholder a total of eight stars, which they could distribute among 14 evidence gaps. In the second round, the stakeholders received eight stars to distribute among eight evidence gaps. In both rounds, a single person could place up to three stars on any one gap.

Prior to the second call, we sent the stakeholders the results of the first online prioritization exercise. For the upper-tier, or highest-ranking, evidence gaps, we also shared the draft PICOTS. During the second call, we reviewed and discussed the results of the prioritization exercise and asked stakeholders for feedback on the PICOTS and for any thoughts on potential research designs for the upper-tier evidence gaps. Following this discussion, we applied the updated PICOTS framework to the upper-tier evidence gaps and translated them into research questions. We then invited the stakeholders to reprioritize the upper tier of the evidence gaps stated as research questions using the same Web-based prioritization tool used in the first round.

After the second round of prioritization, we identified the top-ranked research needs as determined by the stakeholders' prioritization. We present these in this report as the prioritized FRNs for the comparative effectiveness of FGAs and SGAs in the pediatric and young adult populations. The final list of prioritized research needs was not shared with the stakeholders until the public comment period of the draft report.

Research Question Development and Research Design Considerations

Using guidance from a draft AHRQ methods paper titled "Framework for Considering Study Designs for Future Research Needs," we considered several factors and identified potential study designs to address each of the highest-ranked research needs. We considered factors such as advantages of the study design to produce a valid result; resource use, size, and duration; potential social, legal, and ethical issues; and availability of data or ability to recruit participants. We received informal input from the stakeholders on the second conference call regarding study

designs, but we did not ask stakeholders to formally rank study designs because such an exercise would add considerably to the length and complexity of the process, and some stakeholders might not have the technical expertise to engage in this process. We present the study design considerations as suggestions to potential funders and researchers. These suggestions are based on a composite evaluation of the extant literature reviewed in the CER,[1] relevant ongoing studies, stakeholder input during the future research needs process, and our power and sample size analyses.

Results

Evidence Gaps After Stakeholder Input

During the first conference call, the stakeholders reviewed and commented on the list of preliminary evidence gaps (Appendix B) gleaned from the CER. Subsequently, the project team revised the list of evidence gaps based on comments from the stakeholders (Table 2). We asked the stakeholders to prioritize 14 of the 16 evidence gaps; the project team and the stakeholders identified the remaining two gaps as methodological shortcomings relating to bias or funding issues could not be translated into research questions. These were not prioritized along with the other gaps, but the project team considered them when developing potential study design considerations. Three methods-oriented gaps, however, were prioritized along with the other gaps in the first round of prioritization because the project team (with stakeholder input) determined that these particular gaps, although they represented limitations in research methods, did not require additional evidence; the issues related to implementation rather than requiring additional empirical evidence.

Table 2. Evidence gaps after stakeholder input

Evidence Gap[a]	KQ
NOTE: The following group of evidence gaps relate to issues thought to be applicable across all studied disorders and include issues relating to safety, long term efficacy, and subpopulations of interest.	N/A
For children, adolescents, and young adults, the extant literature is limited with regard to examining the long-term efficacy and effectiveness of 1st and 2nd generation antipsychotics in all disorders and behaviors of interest. Examples of longer term outcomes of interest may include outcomes important to parents and patients such as school performance, emotional development, or legal system interactions.	KQ 1, 2
For children, adolescents, and young adults, the extant literature is limited with regard to studies examining the long-term safety of 1st and 2nd generation antipsychotics in all disorders and behaviors of interest. Examples of long term adverse outcomes of interest may include obesity, diabetes, cardiovascular events, or tardive dyskinesia	KQ 2, 3
The extant literature is limited with regard to evidence that allows for comparisons between and within classes of 1st and 2nd generation antipsychotics for any shorter term adverse event outcome. These outcomes include sedation, EPS, weight gain/body composition, insulin resistance, sexual adverse events, and dyslipidemia	KQ 3
The extant literature is limited with regard to evidence to determine if there are differences in efficacy, effectiveness, or adverse events for population sub-groups. Sub-groups include sex, age, race/ethnicity, co-morbidities, co-treatment, history of psychosis, history of treatment failure, or duration of illness.	KQ 4
NOTE: The following 6 evidence gaps relate to specific disorders and focus not only on efficacy data from placebo controlled trials, but also on the state of the evidence as it relates to comparisons within and between the classes of 1st generation and 2nd generation antipsychotics. Further, these gaps discuss the state of the evidence for short and longer term outcomes.	N/A
The evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for the treatment of pervasive developmental disorders, including autistic disorder, Asperger's syndrome, and pervasive developmental disorder not otherwise specified.	KQ 1
For children, adolescents, and young adults with both attention deficit hyperactivity disorder and disruptive behavior disorders, who have failed or had inadequate response to other therapies, there is moderate strength of evidence to support 2nd generation antipsychotics when compared with placebo for improving some behavior symptoms, most notably disruptive behaviors. However, the evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for other clinically meaningful outcomes of interest. These other outcomes include: core ADHD symptoms, anxiety, social/occupational functioning, health related quality of life (HRQL), legal interactions, medication adherence, patient and parent-reported outcomes, and suicide related behavior.	KQ 1

Table 2. Evidence gaps after stakeholder input (continued)

Evidence Gap[a]	KQ
For older adolescents and young adults with bipolar disorder, there was moderate strength of evidence to support 2nd generation antipsychotics over placebo for clinical global impressions. However, the evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for other outcomes such as: aggression, depression, manic symptoms, social/occupational functioning, health related quality of life (HRQL), legal interactions, medication adherence, patient and parent-reported outcomes, and suicide-related behavior.	KQ 1
For adolescents and young adults with schizophrenia, there was moderate strength of evidence to support 2nd generation antipsychotics over placebo for several outcomes, including Children's Global Assessment Scale (CGAS), clinical global impressions, and positive components of the Positive and Negative Symptoms Scale (PANSS). However, the evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for other clinically meaningful outcomes such as: aggression, depression, social/occupational functioning, HRQL, legal interactions, medication adherence, patient and parent reported outcomes, and suicide related behavior.	KQ 1
There was moderate strength of evidence favoring 2nd generation antipsychotics over placebo for tic symptom improvement from studies of all patients with Tourette syndrome. However, evidence is lacking with regard to other clinically meaningful outcomes for this group, and specifically for those who have failed previous treatments. Other clinically meaningful outcomes include: clinical global impressions, obsessive-compulsive symptoms, social/occupational functioning, HRQL, medication adherence, patient and parent reported outcomes, and suicide related behavior.	KQ 1
For those with Obsessive Compulsive Disorder, Eating Disorders, and Post Traumatic Stress Disorder who have failed or had inadequate treatment with other therapies, there is a paucity of data regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics.	KQ 1
NOTE: The group of gaps below relates to issues more methodological in nature and are not specific to drug class or disorder.	N/A
The extant literature is limited with regard to the consistent use of standardized pediatric side-effect scales (e.g., the Safety Monitoring Uniform Report Form).	KQ 2
The extant literature is limited with regard to consistent and comparable outcomes and outcome measurements across the studied disorders and behaviors of concern.	KQ 3, Methods
The extant literature demonstrates a lack of consensus on minimal clinically important differences within many disorders.	KQ 3, Methods
The extant literature lacks large-scale effectiveness studies that are generalizable to the broader population seen in clinical practices.	Methods, All KQs
NOTE: The following gaps were not part of the list of evidence gaps that the stakeholders prioritized. The project team considered these methodological shortcomings of existing research and the issues were of implementation rather than requiring additional empiric evidence.	N/A
The extant literature is limited with regard to efficacy studies with adequate blinding of study participants and outcome assessors, the adequate concealing of allocation and the appropriate handling and reporting of missing data.	Methods, All KQs
The extant literature is limited with regard to independent/investigator driven research efforts which increases the potential for overestimated treatment effects associated with industry-funded research.	Methods, All KQs

[a]All of the evidence gaps refer to the population of children, adolescents, and young adults 24 years of age or younger.
Abbreviations: ADHD=attention deficit hyperactivity disorder; CGAS=Children's Global Assessment Scale; EPS=extra-pyramidal symptoms; HRQL=health related quality of life; KQ=key question; PANSS=Positive and Negative Symptoms Scale.

Prioritization Results

We conducted two rounds of prioritization to ultimately identify six high-priority research needs. In the first round, the stakeholders determined that 9 of the 14 evidence gaps were of higher priority than the rest of the evidence gaps. Ten of the 12 stakeholders (83 percent) completed the first online prioritization exercise. The number of stars allocated to each evidence gap ranged from 1 to 18. Based on the distribution of the stars allotted to each gap, the upper-tier evidence gaps were the 9 gaps that received four or more stars. On reviewing the upper- and lower-tier gaps, and considering stakeholder comments, the project team removed one of the gaps from the next round of prioritization because the group considered it to be a methods gap

too broad to frame as a research question. Ultimately, 8 evidence gaps moved forward to the next round of prioritization.

Between the first and second rounds of prioritization, the project team applied a PICOTS framework to the upper-tier evidence gaps and developed related research questions for each of them. Stakeholders provided input for the PICOTS framework during the second conference call. Table 3 shows the PICOTS used by the project team to develop research questions from the upper-tier evidence gaps.

Table 4 shows the list of 8 evidence gaps and related research questions that the stakeholders prioritized in the second round. Ten out of 12 stakeholders (83 percent) completed the second round of prioritization. The results are shown in Figure 2. The number of stars allotted to each research need ranged from 3 to 17. Out of eight high-priority evidence gaps, six were deemed by the stakeholders as the highest-priority research needs.

Table 3. PICOTS for the upper-tier evidence gaps and related research questions prioritized in round 2

The following descriptions of PICOTS define the scope of the evidence gaps and research questions identified through the Future Research Needs for the Comparative Effectiveness of First and Second Generation Antipsychotics stakeholder engagement process. Specific outcomes may apply to one or more research questions. For example, some questions focus on health and behavioral outcomes and others focus on adverse events; the outcomes described below apply to each as appropriate.
Population
All Questions: Children, youth, and young adults (24 years of age or younger) with one or more of the following disorders: PDD, ADHD, DBD, bipolar disorder, schizophrenia/schizophrenia-related psychosis, OCD, post-traumatic stress disorder, eating disorders (anorexia nervosa, bulimia nervosa, eating disorder not otherwise specified), tic disorders (Tourette syndrome), or nondisorder-specific severe behavioral issues (e.g., aggression). For Subgroups: Demographic and clinical subgroups of children, youth, and young adults 24 years of age or younger with one or more of the disorders listed above including: age, racial groups, gender, genetically defined subgroups, socioeconomic groups, level of education, or physical and mental health comorbidities (other psychiatric disorders, treatment history, substance abuse, learning disabilities, developmental disorders, language impairments, cognitive abilities).
Interventions
All Questions: Any FDA-approved FGA or SGA
Comparators
All Questions: Any other FDA-approved FGA or SGA, placebo, or another dose of the same antipsychotic
Outcomes (Efficacy or Effectiveness)
Disorder/illness-specific symptom-related outcomes: The following symptom domains are *examples* of domains of interest for each disorder or illness and not an exhaustive list of all potential outcomes that might be measured: • PDDs: repetitive behaviors, social/communication, agitation/aggression • ADHD and DBDs: impulsivity, defiant behavior, aggressive behavior • Bipolar disorder: mania, depression • Schizophrenia and related psychoses: positive and negative symptoms, cognitive function • Obsessive compulsive disorder: obsession severity, compulsion severity • Post-traumatic stress disorder: re-experiencing, avoidance behaviors • Anorexia nervosa/bulimia nervosa/eating disorder NOS; metabolic/nutritional, purging frequency/intensity • Tourette syndrome; tic severity Nondisorder-specific behaviors of interest: The following are *examples* of key common nondisorder-specific behaviors of interest that could be considered when evaluating efficacy/effectiveness: aggression, agitation, anxiety, behavioral dyscontrol, irritability, mood lability, self-injurious behaviors, and sleep disorders.

Table 3. PICOTS for the upper-tier evidence gaps and related research questions prioritized in round 2 (continued)

Outcomes (Related to Key Psychiatric Comorbidities)
The following are *examples* of key co-morbidities for the specific disorders of interest listed above that could also be considered when evaluating efficacy/effectiveness. • PDDs: ADHD, anxiety disorders • ADHD and DBDs: Substance abuse, depression, anxiety disorders • Bipolar disorder: ADHD, substance abuse/dependence • Schizophrenia and related psychoses: Depression, substance abuse/dependence, • OCD: Tourette syndrome/tic severity, ADHD • PTSD: Depression, substance abuse • Anorexia nervosa/bulimia nervosa/eating disorder NOS: Depression • Tourette syndrome: ADHD, OCD
Outcomes (Harms or Adverse Events)
Medication-associated adverse events, withdrawal due to adverse events, persistence and reversibility of adverse events. Examples of major adverse events include the following: • Mortality • Cerebrovascular disease-related events • Development of diabetes mellitus • Diabetic ketoacidosis • Neuroleptic malignant syndrome • Seizures • Tardive dyskinesia • Cardiomyopathies • Cardiac arrhythmias • Agranulocytosis • Extrapyramidal effects General adverse event examples include the following: • Weight gain (e.g., using body mass index growth charts) • Agitation • Constipation • Sedation • Elevated cholesterol • Elevated transaminases • Adverse events related to prolactin elevations • Galactorrhea/bloody galactorrhea • Exercise intolerance • Precocious puberty

Table 3. PICOTS for the upper-tier evidence gaps and related research questions prioritized in round 2 (continued)

Outcomes (Other Short-Term and Long-Term Outcomes)
• Response rates with corresponding dose, duration of response, remission, relapse, speed of response, time to discontinuation of medication • Growth and maturation • Cognitive and emotional development • Suicide-related behaviors or death by suicide • Medication adherence and persistence • School performance/attendance • Work-related functional capacity • Patient insight into illness • High-risk behavior outcomes (i.e., unintended pregnancy, accidental injury/death, sexually transmitted disease, substance abuse) • Patient or parent/care provider reported outcomes, including levels of physical activity/inactivity, diet (i.e., caloric intake, food preferences) • Health-related quality of life • Relationship functioning • Legal/justice system interaction (i.e., arrests, detention) • Health care system utilization (e.g., hospitalization rates, medication costs, outpatient expenditures) • "Outcomes that matter" to children, youth and young adults, and their families. These functional outcomes may reflect a developmental perspective.
Timing
• Short term: ≤ 6 months • Long term: > 6 months, generally years
Settings
All settings, including inpatient hospitalization and outpatient treatment, and practice-based research networks

Abbreviations: ADHD=attention deficit hyperactivity disorder; DBD=disruptive behavior disorder; FDA=U.S. Food and Drug Administration; FGA=first-generation antipsychotics; NOS=not otherwise specified; OCD=obsessive-compulsive disorder; PDD=pervasive developmental disorder; PICOTS=populations, interventions, comparators, outcomes, time frames, and settings; PTSD=post traumatic stress syndrome; SGA=second-generation antipsychotics

Table 4. Upper-tier evidence gaps and related research needs prioritized in round 2

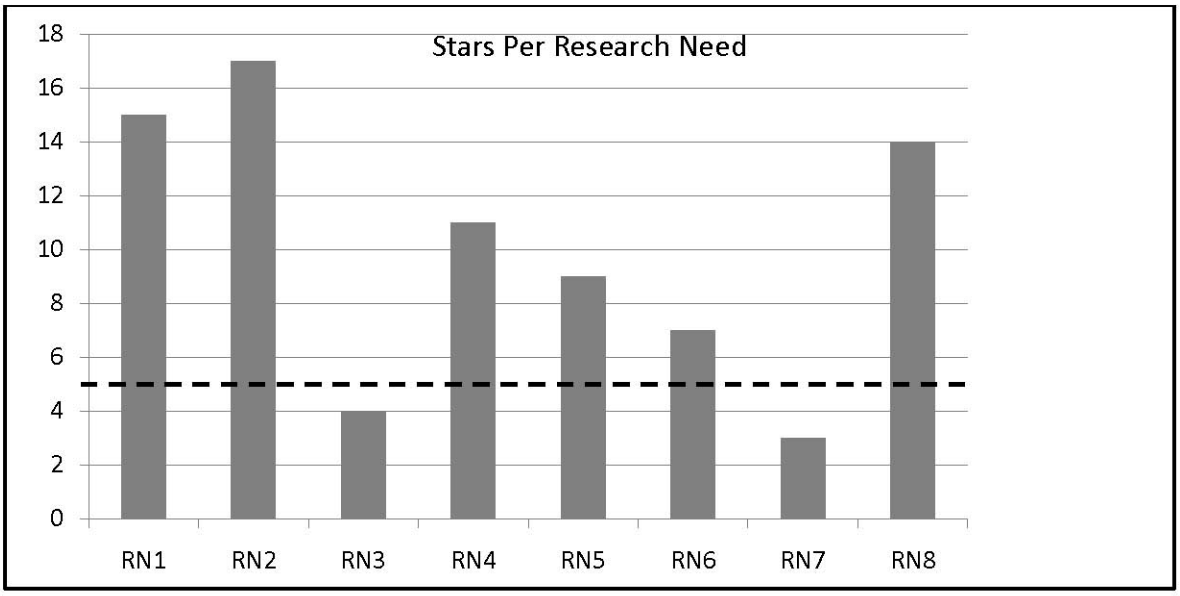
Evidence Gap (EG) 1 For children, adolescents, and young adults, the extant literature is limited with regard to examining the long-term efficacy and effectiveness of 1st and 2nd generation antipsychotics in all disorders and behaviors of interest. Examples of longer term outcomes of interest may include outcomes important to parents and patients such as school performance, emotional development, or legal system interactions. Key Question (KQ) 1, KQ2
Research Need (RN) 1 What is the long-term comparative effectiveness between and within classes of antipsychotics as measured in outcomes related to the disorder of interest, its co-morbidities, associated behavioral features, social-occupational outcomes, and outcomes identified as important by patients and their families?
EG2 For children, adolescents, and young adults, the extant literature is limited with regard to studies examining the long-term safety of 1st and 2nd generation antipsychotics in all disorders and behaviors of interest. Examples of long term adverse outcomes of interest may include obesity, diabetes, cardiovascular events, or tardive dyskinesia. KQ2, KQ3
RN2 What are the comparative long-term risks of medication exposure between and within antipsychotic classes?
EG3 The extant literature is limited with regard to evidence that allows for comparisons between and within classes of 1st and 2nd generation antipsychotics for any shorter term adverse event outcome. These outcomes include sedation, EPS, weight gain/body composition, insulin resistance, sexual adverse events, and dyslipidemia. KQ3
RN3 What are the comparative short-term risks of medication exposure between and within antipsychotic classes?
EG4 The extant literature is limited with regard to evidence to determine if there are differences in efficacy, effectiveness, or adverse events for population sub-groups. Sub-groups include sex, age, race/ethnicity, co-morbidities, co-treatment, history of psychosis, history of treatment failure, or duration of illness. KQ4
RN4 Are there subgroups of patients based on baseline demographic/clinical characteristics or physical and/or mental health co-morbidities for which first and second generation antipsychotics differ in efficacy, effectiveness, or frequency of adverse events?

Table 4. Upper-tier evidence gaps and related research questions prioritized in round 2 (continued)

EG6[a] For children, adolescents, and young adults with both attention deficit hyperactivity disorder and disruptive behavior disorders, who have failed or had inadequate response to other therapies, there is moderate strength of evidence to support 2nd generation antipsychotics when compared with placebo for improving some behavior symptoms, most notably disruptive behaviors. However, the evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for other clinically meaningful outcomes of interest**.** These other outcomes include: core ADHD symptoms, anxiety, social/occupational functioning, health related quality of life (HRQL), legal interactions, medication adherence, patient and parent-reported outcomes**,** and suicide related behavior. KQ1
RN5 What is the efficacy and effectiveness of first or second generation antipsychotics for individuals with attention deficit hyperactivity disorder and disruptive behavior disorders in the following outcome domains: core ADHD symptoms, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient and parent-reported outcomes, those related to high risk behaviors, and suicide related behavior?
EG7 For older adolescents and young adults with bipolar disorder, there was moderate strength of evidence to support 2nd generation antipsychotics over placebo for clinical global impressions. However, the evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for other outcomes such as: aggression, depression, manic symptoms, social/occupational functioning, health related quality of life (HRQL), legal interactions, medication adherence, patient and parent-reported outcomes, and suicide-related behavior. KQ1
RN6 What is the efficacy and effectiveness of first or second generation antipsychotics for adolescents and young adults with bipolar disorder in the following outcome domains: core features of the disorder, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient and parent-reported outcomes, those related to high risk behaviors, and suicide related behavior?
EG8 For adolescents and young adults with schizophrenia, there was moderate strength of evidence to support 2nd generation antipsychotics over placebo for several outcomes, including Children's Global Assessment Scale (CGAS), clinical global impressions, and positive components of the Positive and Negative Symptoms Scale (PANSS). However, the evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for other clinically meaningful outcomes such as: aggression, depression, social/occupational functioning, HRQL, legal interactions, medication adherence, patient and parent reported outcomes, and suicide related behavior. KQ1
RN7 What is the efficacy and effectiveness of first or second generation antipsychotics for adolescents and young adults with schizophrenia in the following outcome domains: core features of the disorder, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient and parent reported outcomes, those related to high risk behaviors, and suicide related behavior?
EG11[a] The extant literature is limited with regard to the consistent use of standardized pediatric side-effect scales (e.g., the Safety Monitoring Uniform Report Form). KQ2
RN8 Which new or existing tools or methods should be consistently applied in community care and/or pragmatic research settings to measure the relevant adverse effects (including behavioral side effects) related to antipsychotic usage?

[a]The numbers for the Evidence Gaps (EGs) correspond to the numbers assigned in the list that was prioritized, only those that emerged as the upper tier EGs are listed in this table, therefore some numbers appear out of sequence.

Figure 2. Distribution of stars among eight research needs



The vertical axis represents the number of stars assigned. The bars represent the research needs (RNs). The dashed line represents the cutoff point between the upper-tier and lower-tier research needs. All research needs with five or more stars represent the stakeholders' highest priority research needs.

Highest Priority Research Needs: PICOTS Information and Considerations of Potential Research Designs

For each of the six highest-priority research needs below, we state the need in the form of a research question, highlight key points of the stated need, and describe considerations for research designs. The PICOTS that define the populations, interventions, comparators, outcomes, time frames, and settings for the research needs identified by the stakeholders are listed above in Table 3. This set of top-tier research needs should be considered as a whole, without any particular ranking. Design considerations are focused on maximum reduction of bias and feasibility to maximize probability of completion. A general description of potential study designs for these research needs is included in Appendix G.

Research Need

What is the long-term comparative effectiveness between and within classes of antipsychotics as measured in outcomes related to the disorder of interest, its co-morbidities, associated behavioral features, social-occupational outcomes, and outcomes identified as important by patients and their families?

Considerations for Potential Research Designs

A well-designed RCT is an ideal design but not likely to have sufficient power to detect differences between and within classes of antipsychotics for all long-term outcomes of interest. Although random assignment is the best way to avoid biased estimates of treatment effects, a number of factors may make randomized trials difficult in this research area. Child research participants (or their parents) may often be unwilling to accept random assignment to treatment. Further, where there is a perceived or real lack of equipoise (e.g., in comparing FGAs versus

SGAs), random assignment may be neither ethical nor acceptable to participants. Finally, randomized trials are resource intensive, and many of the research questions (e.g., between-drug effectiveness comparisons) may require large samples because of small treatment effects.

Therefore, the following designs could be considered:

- Nonrandomized comparative designs
- Observational studies
  - Prospective cohort design
  - Retrospective cohort design

Nonrandomized Comparative Design: For this population, a nonrandomized type of design would allow researchers to attend to challenges of examining long-term outcomes in this population. Treatment may be assigned on the basis of provider or practice. Remaining challenges for this specific research need include handling high levels of psychiatric comorbidity for some disorders of interest, lack of consensus on disease phenomenology and nosology within the field for some disorders of interest, and high levels of concern regarding selection bias (for example, children who receive antipsychotics may have more behavioral severity and are, therefore, less able to participate in clinical research). Some of the outcomes of interest for the research needs discussed here may exist in some studies, but the heterogeneity of the study populations and the outcomes (and measurement tools used) has not allowed for meta-analysis on most topics of interest.

Observational Studies, such as prospective and retrospective cohort designs, are also useful for addressing the complex comparative effectiveness questions raised in this research need. Although similar to nonrandomized comparative designs (trials in which treatment is not assigned at random), cohort design participants (or their parents) are allowed to select interventions. Registries of antipsychotic-exposed individuals with linkages to clinical datasets such as outpatient and pharmacy claims data and/or electronic medical record (EMR) systems will be important but are only beginning to develop in the US and Canada. Practice based research networks in child mental health may also be vehicles for evaluating the needs below but such platforms are nascent in this field.

Research Need

What are the comparative long-term risks of medication exposure between and within antipsychotic classes?

Considerations for Potential Research Designs

Similar to the previous research need, a well-designed RCT is an ideal design but not likely to have sufficient power to detect differences between and within classes of antipsychotics in terms of change in harms for all long-term outcomes of interest. Therefore, the following designs could be considered:

- Nonrandomized comparative designs
- Observational studies
  - Prospective cohort design
  - Retrospective cohort design

Research Need

What is the efficacy and effectiveness of first or second generation antipsychotics for adolescents and young adults with schizophrenia in the following outcome domains: core features of the disorder, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient and parent reported outcomes, those related to high risk behaviors, and suicide related behavior? Outcomes include aggression, depression, social/occupational functioning, health related quality of life (HRQL), legal interactions, medication adherence, patient and parent reported outcomes, and suicide related behavior.

Considerations for Potential Research Designs

Similar to the previous research needs, a well-designed RCT is an ideal design but not likely to have sufficient power to detect differences along the outcomes of interest given the focus on adolescents and young adults with schizophrenia. Thus, for this research need, the study designs listed below could be considered:

- Nonrandomized comparative designs
- Observational studies
  - Prospective cohort design
  - Retrospective cohort design
  - Case control study

Research Need

Are there subgroups of patients based on baseline demographic/clinical characteristics or physical and/or mental health co-morbidities for which first and second generation antipsychotics differ in efficacy, effectiveness, or frequency of adverse events? Sub-groups include sex, age, race/ethnicity, co-morbidities, co-treatment, history of psychosis, history of treatment failure, or duration of illness.

Considerations for Potential Research Designs

For this research need, the study designs listed below could be considered:

- Nonrandomized comparative designs
- Observational studies
  - Prospective cohort design
  - Retrospective cohort design
  - Meta-analysis of individual patient data. This type of evaluation is appropriate for gathering information on the comparative long-term risks of antipsychotic medication exposure between and within antipsychotic classes. Another appropriate design would involve pooling data from existing trials for simultaneous data analysis. Advantages of this study design for producing a valid result include increased statistical power, increased ability to account for sample heterogeneity, and increased frequency of low-base rate behaviors/harms. Similar to NRCTs, careful consideration of selection bias and unmeasured confounders need to be controlled for results to be definitive and to get closer to drawing causal inferences. Integration of multiple-armed trials would allow testing of several hypotheses regarding agents. However, use of this study design requires the existence of underlying databases with common data definitions and willingness of investigators to pool data. Appendix G lists additional relevant characteristics and considerations of this design.

Research Need

What is the efficacy and effectiveness of first or second generation antipsychotics for individuals with attention deficit hyperactivity disorder and disruptive behavior disorders in the following outcome domains: core ADHD symptoms, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient and parent-reported outcomes, those related to high risk behaviors, and suicide related behavior?

Considerations for Potential Research Designs

For this research need, the study designs listed below could be considered:

- Randomized comparative designs
- Nonrandomized comparative designs
- Observational studies
  - Prospective cohort design
  - Retrospective cohort design
  - Case control study
  - Meta-analysis of individual patient data

Research Need

What is the efficacy and effectiveness of first or second generation antipsychotics for adolescents and young adults with bipolar disorder in the following outcome domains: core features of the disorder, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient and parent-reported outcomes, those related to high risk behaviors, and suicide related behavior?

Considerations for Potential Research Designs

For this research need, the study designs listed below could be considered:

- Randomized comparative designs
- Nonrandomized comparative designs
- Observational studies
  - Prospective cohort design
  - Retrospective cohort design
  - Case control study
  - Meta-analysis of individual patient data

Similar to other disorder-specific needs, data from multiple studies can be pooled, and propensity matching can be used to compare outcomes of interventions. Also, registry studies, possibly using paired electronic health records or claims data, may allow for evaluation of some questions. Key considerations are whether the available data from such databases have the outcomes of interest noted.

Sample Size Requirements

This section provides a simple framework for thinking about the sample size requirements associated with common study designs that could be used to address one or more of the highest-priority research needs. The examples focus on relatively simple comparisons of benefits or harms between two drugs or between a drug and placebo. Studies to establish equivalence between two drugs would require larger sample sizes and different analysis techniques.

Comparisons of more than two drugs would require somewhat larger total sample sizes (for example, see Cohen[12]) as well as consideration of the increase in Type I error (the probability of finding an effect where none exists) associated with multiple comparisons.

The sample size estimates assume statistical power of 0.80 and a two-tailed significance level of 0.05. These estimates are approximate, and they do not account for design improvements that could increase power, such as the use of repeated measures or the inclusion of covariates to increase the precision of estimates.

Sample Size Estimate Example 1: Comparison of Means

This example applies to randomized and nonrandomized trials and to prospective and retrospective cohort studies. The example involves a continuous measure of an adverse effect but is also applicable to beneficial outcomes (i.e., efficacy or effectiveness studies); this example is relevant to a number of Research Needs.

Correll et al.[13] studied cardiometabolic risk in 205 children and adolescents with psychotic, mood, or aggressive behavior disorders who were untreated or newly treated with quetiapine, risperidone, aripiprazole, or olanzapine. For 12-week weight gain, the authors reported within-drug effect sizes (ESs) of 0.04, 0.77, 0.77, 0.84, and 0.97, respectively. (The average weight gain for the four antipsychotics ranged from 8 percent to 15 percent of baseline weight, above the usual clinical significance threshold of ≥7 percent). Table 5 shows that in a cohort study finding a within-drug adverse effect similar in magnitude to the minimum 12-week weight gain above (ES, 0.77) would require a sample of 14 (the sample size listed for an effect size of 0.80 and one sample). Finding a drug-placebo difference equivalent to the minimum difference between drug and no treatment (ES, 0.73, assuming that the effect sizes reported for weight gain were based on a common denominator) would require about 29 people per group (the sample size listed for an effect size of 0.75 and two samples).

Table 5. Approximate sample size required per group for comparisons of means[14] (two-tailed test, α=.05, power=.80)

Effect Size	One Sample (Paired)	Two Samples
0.10	787	1,571
0.15	351	699
0.20	198	393
0.25	128	252
0.30	89	175
0.35	66	129
0.40	51	99
0.45	41	78
0.50	33	64
0.55	28	53
0.60	24	45
0.65	21	38
0.70	18	33
0.75	16	29
0.80	14	26

Table 6 highlights the substantially greater number of participants necessary to test directly small effects such as differences in weight gain among SGAs. Even if a between-drug difference were equivalent to the maximum between-drug difference in the study by Correll et al.[13] (ES, 0.20, or an average gain of 15 percent versus 8 percent relative to baseline weight), detecting that difference would require a much larger sample size, approximately 393 per group (ES, 0.20, two samples).

Table 6. Approximate sample size required per group for comparisons of two proportions[19]

	Proportion (Drug A)									
Proportion (Drug B or Placebo)	**.20**	**.30**	**.35**	**.40**	**.45**	**.50**	**.55**	**.60**	**.65**	**.70**
.10	199	62	43	32	25	20	16	15[a]	14[a]	13[a]
.20		294	138	82	54	39	29	23	18	15
.30			1,377	356	163	93	61	42	31	24
.35				1,471	376	170	96	62	43	31
.40					1,534	388	173	97	62	42
.45						1,565	392	173	96	61
.50							1,565	388	170	93
.55								1,534	376	163
.60									1,471	356
.65										1,377

[a]To maintain expected cell counts of at least 5 for valid statistical inference, the minimum sample size per group is increased slightly.

Sample Size Estimate Example 2: Comparison of Proportions

This example applies to randomized and nonrandomized trials, prospective and retrospective cohort studies, and case-control studies. The example involves dichotomous measures of efficacy or effectiveness but is also applicable to studies of adverse effects.

In an industry-sponsored clinical trial, Tohen et al.[15] compared olanzapine with placebo in the treatment of 161 adolescents with bipolar mania. They measured response rates of 44.8 percent and 18.5 percent, respectively, and remission rates of 35.2 percent and 11.1 percent, respectively, defining response as a ≥50 percent reduction in the Young Mania Rating Scale score and remission as having an endpoint score of ≤12. Haas et al.[16] compared two risperidone dose regimens with placebo in 169 youths from the same population. Based on the same definition of treatment response, they found response rates of 59 percent and 63 percent in the treatment group (depending on dose regimen) and a response rate of 26 percent in the placebo group. In this pair of trials, the minimum difference between drug and placebo (i.e., the difference in remission rates between olanzapine and placebo in the study by Tohen et al.[15]) was approximately 35 percent versus 10 percent (OR, 4.85; equivalent ES, 0.87[17]). All other things being equal, this effect would be considered clinically significant because the number needed to treat is only four.[18] (Number needed to treat, or NNT, is the number of people who would need to be treated for one additional person to experience remission.) In a new study, if one hypothesized a difference in remission rates of 35 percent (Drug A) versus 10 percent (placebo), Table 6 shows that finding an effect of this size would require a sample of 43 people per group. Generally, compared with the efficacy trials cited here, effectiveness studies are likely to involve even smaller effects because of greater heterogeneity in the patient populations and less control over experimental conditions. Therefore, effectiveness studies with similar outcomes measures are likely to require larger sample sizes.

In contrast to the drug-placebo example in the preceding paragraph, the difference between response rates for low-dose risperidone (in the study by Haas et al.[16]) and olanzapine (in the study by Tohen et al.[15]) was smaller, 59 percent versus 44.8 percent (OR, 1.77; equivalent ES, 0.32; NNT, 7). In a new study, if one hypothesized a difference in response rates of 60 percent (Drug A) versus 45 percent (Drug B), Table 6 shows that detecting a difference of that size would require a larger sample, about 173 people per group.

If one hypothesized response rates that were slightly lower and slightly closer together, say, 50 percent for Drug A and 40 percent for Drug B (OR, 1.5; equivalent ES, 0.22; NNT, 10), Table 6 shows that the required sample size would be about 388 subjects per group, as large as many of the largest trials in this field. We would consider such a difference clinically important because only 10 people would need to be treated with Drug A instead of Drug B for one additional person to experience treatment response, an important outcome.

Discussion

This project used a structured approach to further refine and prioritize specific areas of future research addressing antipsychotic usage in youth that were necessary to sufficiently address key questions in the CER "First- and Second-Generation Antipsychotics for Children and Young Adults."[1] This project clearly identified evidence gaps described in the report, worked with stakeholders to prioritize this list of gaps and transform them into research questions, further prioritized these needs, and considered PICOTs and potential study designs for the highest-priority needs.

We assembled a diverse team of stakeholders who provided knowledge in the broad and complex area of antipsychotic usage in pediatric mental health. The stakeholder group included clinicians, advocates, researchers, education specialists, and funders. Our conference calls before and after the first round of prioritization revealed several key themes that helped inform our development of the evidence gap list. These themes included a desire for improved understanding of long-term safety risks, improved understanding of how antipsychotics may affect real-world outcomes, and the need for evaluation of generalizable study populations that emphasize comorbidity over diagnostic purity. The stakeholders emphasized the importance of longer-term study periods (years versus months) and de-emphasized the utility of more short-term RCTs. Stakeholders made special note of the lack of patient registries with linkages to clinical datasets as well as the lack of long-term cohort studies that could facilitate comparative effectiveness research to assess benefits and harms over the last several years. This process of refinement helped create a list of 16 evidence gaps that were based on those derived from the CER but also reflected the perspectives of the stakeholders. Our initial prioritization exercise included 14 of the 16 evidence gaps and revealed 8 higher-priority evidence gaps. From this, 6 high-priority research needs emerged.

Although significantly lower than the original number of 14 evidence gaps prioritized by the stakeholders, the final list of 6 highest-ranking research needs incorporates a broad spectrum of issues. The needs identified ranged from methodological considerations to needs emphasizing medication safety issues to clinically oriented effectiveness research needs. Through their rankings, the stakeholders clearly emphasized general medication safety and effectiveness issues over disorder-specific medication effectiveness gaps. This emphasis was notable given the general lack of efficacy data in many of the disorders of interest. In our conference calls, some stakeholders indicated concern that antipsychotics are used with increasing frequency among youth for behavioral symptoms such as irritability and aggression that are common across many disorders rather than for symptoms specific to disorders (e.g., positive symptoms of schizophrenia) or for the disorders themselves. In addition, several stakeholders noted on both conference calls that the established efficacy data for antipsychotics is of low to moderate effect size and more disorder-specific efficacy trials of existing medications should, therefore, not be the focus of research investment. In conference calls, the stakeholders acknowledged that effectiveness studies in this area are hampered by a lack of appropriate, validated, and practical measurement tools to allow for surveillance of effectiveness and safety outcomes in more practice-based and or pragmatic research settings. Of note, this gap related to assessment tools was highly ranked in the first round of gap prioritization but was not among the six highest-ranked items in the second round.

Three evidence gaps gleaned from the CER and identified by our stakeholders were deemed by the FRN team as important but not eligible for ranking. These concerns were either too broad in scope, related to very specific design flaws about which there is broad agreement, or related to

sources of bias and were not translatable into scientific research questions. These concerns were left as general recommendations for consideration and are listed below.

- The extant literature lacks large-scale effectiveness studies that are generalizable to the broader population seen in clinical practices.
- The extant literature is limited with regard to efficacy studies with adequate blinding of study participants and outcomes assessors, adequate concealing of allocation, and appropriate handling and reporting of missing data.
- The extant literature is limited with regard to independent/investigator-driven research efforts, which increases the potential for overestimated treatment effects associated with industry-funded research.

In our consideration of potential study designs, we proposed the notion of randomized trials for only some research need areas, reflecting the input of stakeholders who noted the clear pragmatic barriers to answering long-term effectiveness and safety questions with this design. Our consideration of sample size was informative to illustrate the wide range of complexity and resources that would be required to undertake some of the potential trial designs. For example, evaluating between drug differences for weight gain in a prospective cohort design may require up to 400 per group when comparing two drugs and becomes notably more resource intensive for the comparison of three or more drugs. In such cases, sample size and cost will need to be weighed carefully against the importance of the research question. Large prospective cohort designs that follow youth with a range of mental health disorders and treatments would be valuable for comparative effectiveness and safety research but are hampered by cost and difficulties with loss to followup.

It is possible that some of these effectiveness and safety questions could be evaluated with complex secondary data analysis techniques. Unfortunately, datasets from trials are often not useful for these purposes because of short length of followup or lack of appropriate measures. When evaluating shorter-term safety concerns, combining data from trials may be more feasible because some safety outcomes tend to be more homogenous across studies (e.g., weight gain). Meta-analysis is a commonly employed tool in comparative effectiveness research but may not be a feasible tool for many questions because of a lack of studies in an area, study length, or the heterogeneity of populations and measurement tools across existing studies. In some situations where original raw data from previously completed studies could be shared among investigators, advanced techniques such as meta-analysis of individual patient data could potentially be used to deal with the heterogeneity problem.

Patient registries with linkages of patient-reported outcomes to clinical datasets (such as claims or EMR data) are a likely lower-cost approach for some questions about effectiveness and safety. However, such data-linked registries are nascent in mental health and may be hampered by confidentiality issues and methodological limitations. Finally, stakeholders pointed out that any attempt to perform comparative effectiveness evaluations involving FGAs in youth outside of randomized trials would be hampered by the low prevalence of use of these agents in the community. Further, stakeholders discussed the notion that even large comparative trials between FGAs and SGAs for youth and or young adults may be hampered by applicability and feasibility because few clinicians would be inclined to use FGAs, except in the case of schizophrenia.

Limitations and Challenges

Stakeholder participation. The nature of the stakeholder process lends itself to certain important limitations. Although stakeholder input was essential, scheduling challenges led to incomplete participation from some members. To accommodate this challenge, we provided opportunities for stakeholders to provide feedback by email, but we did not speak with stakeholders by telephone individually. To account for the possibility that stakeholders who attended conference calls were unable to express an opinion during a call in which they participated, we provided the opportunity to offer further comment by email.

Key challenges related to gap development and presentation. The process required that we strike a balance between the need to develop a list of digestible evidence gaps with a diverse group of stakeholders and the need to remain faithful to the purpose, findings, and intent of the original CER on antipsychotics in youth. An inherent strength of the process is that neither the findings of the systematic review nor the views of the stakeholder panel fully dominate the final product.

In developing the gaps, we arranged them by key question from the CER.[1] Because of this organizational scheme, gaps associated with different key questions may have appeared quite similar and/or overlapping to stakeholders. For example, a stakeholder may have had trouble distinguishing between or prioritizing a gap about lack of long-term effectiveness data for a particular disorder versus a gap about long-term effectiveness data for antipsychotic use in youth in general. Some stakeholders may have been more accustomed to considering evidence gaps and research needs from their particular field of work and might have found it difficult to interpret and prioritize research questions written and organized by other authors. The project team and the stakeholders raised the question of whether gaps might be made more digestible if organized by theme or domain. For example, one proposed schema broke the gaps into methodological, treatment intervention-oriented, and outcome-oriented gaps.

We worked to minimize the risk of arbitrary exclusion of gaps or needs by erring on the side of inclusion when cut points were not clear. Further, we provided the opportunity for discussion after the first round of prioritization on the subject of re-elevating low ranking gaps for inclusion for the second round of prioritization. However, the stakeholder group did not reach consensus on re-elevating any particular low ranking gap after the first prioritization exercise. Of note, the stakeholder process is not intended to delineate a clear rank order of research needs. Further, all of the generated comparative effectiveness evidence gaps have potential public health value and are listed in Table 2 in the results section of this report.

Conclusions

Overall, the stakeholders demonstrated engagement in our discussions of research challenges in the field and were able to perform the ranking process without difficulty. The six high priority research needs (Table 7) included a broad range of issues cutting across disorders, key clinical outcomes, safety outcomes, and methodological concerns. The prioritization revealed emphasis on safety and effectiveness needs across youth mental conditions rather than within disorder subgroups with the exception of bipolar disorder and schizophrenia. PICOTS development aided our consideration of study design issues and our sample power analyses demonstrated the clear pragmatic barriers that many of the potential designs will present. While large long-term multi-site clinical trials may be the gold standard to assess many of the questions of importance, feasibility has greatly limited the number of such large pragmatic trials in mental health to date. Large prospective cohort studies of youth exposed to antipsychotics may be viable and offer considerable analytic flexibility, but are also costly. Patient registries with linkages to clinical data sets may allow for more efficient evaluation of some questions with advanced analysis methods, but the infrastructure for this needs considerable investment and its development may face considerable hurdles relating to information privacy. Meta-analysis of existing trials data and meta-analysis of individual patient data may prove helpful, but will likely be limited to evaluation of specific shorter term outcomes. Despite its limitations, the structured process used in this project may prove to be an effective way of reaching relative consensus on research priorities in this broad and complex topic area.

Table 7. Six high priority research needs identified by stakeholders

What is the long-term comparative effectiveness between and within classes of antipsychotics as measured in outcomes related to the disorder of interest, its co-morbidities, associated behavioral features, social-occupational outcomes, and outcomes identified as important by patients and their families?
What are the comparative long-term risks of medication exposure between and within antipsychotic classes?
What is the efficacy and effectiveness of first or second generation antipsychotics for adolescents and young adults with schizophrenia in the following outcome domains: core features of the disorder, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient and parent reported outcomes, those related to high risk behaviors, and suicide related behavior?
Are there subgroups of patients based on baseline demographic/clinical characteristics or physical and/or mental health co-morbidities for which first and second generation antipsychotics differ in efficacy, effectiveness, or frequency of adverse events?
What is the efficacy and effectiveness of first or second generation antipsychotics for individuals with attention deficit hyperactivity disorder and disruptive behavior disorders in the following outcome domains: core ADHD symptoms, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient and parent-reported outcomes, those related to high risk behaviors, and suicide related behavior?
What is the efficacy and effectiveness of first or second generation antipsychotics for adolescents and young adults with bipolar disorder in the following outcome domains: core features of the disorder, its commonly associated co-morbidities and behavioral features, social/occupational functioning, patient and parent-reported outcomes, those related to high risk behaviors, and suicide related behavior?

References

1. Seida JC, Schouten JR, Mousavi SS, et al. First- and Second-Generation Antipsychotics for Children and Young Adults. Comparative Effectiveness Review No. 39. (Prepared by the University of Alberta Evidence-based Practice Center under Contract No. 290-2007-10021.) AHRQ Publication No. 11(12)-EHC077-EF. Rockville, MD. Agency for Healthcare Research and Quality. February 2012.
2. Leucht S, Corves C, Arbter D, et al. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. Lancet 2009 Jan 3;373(9657):31-41. PMID: 19058842.
3. Haddad PM, Das A, Keyhani S, et al. Antipsychotic drugs and extrapyramidal side effects in first episode psychosis: a systematic review of head-head comparisons. J Psychopharmacol 2011 Nov 5. PMID: 22057019.
4. Margetic B, Aukst-Margetic B. Neuroleptic malignant syndrome and its controversies. Pharmacoepidemiol Drug Saf 2010 May;19(5):429-35. PMID: 20306454.
5. Zito JM, Safer DJ, dosReis S, et al. Trends in the prescribing of psychotropic medications to preschoolers. JAMA 2000;238(8):1025-30. PMID: 10697062.
6. Zito JM, Safer DJ, DosReis S, et al. Psychotropic practice patterns for youth: a 10 year perspective. Arch Pediatric Adolesc Med 2003;157(1):17-25. PMID: 12517190.
7. Zito JM, Safer DJ, de Jong-van den Berg LT, et al. A three-country comparison of psychotropic medication prevalence in youth. Child Adolesc Psychiatry Ment Health 2008;121(1):26. PMID: 18817536.
8. Zito JM, Safer DJ, Sai D, et al. Psychotropic medication patterns among youth in foster care. Pediatrics 2008;121(1):e157-e63. PMID: 18166534.
9. Pathak S AS, Danielyan A, et al. Psychotropic utilization and psychiatric presentation of hospitalized very young children. J Child Adolesc Psychopharmacol 2004;14(3):433-42.
10. Kuehn BM. Studies shed light on risks and trends in pediatric antipsychotic prescribing. JAMA 2010 May 19;303(19):1901-3. PMID: 20483961.
11. Owens DK, Lohr KN, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions. Methods guide for effectiveness and comparative effectiveness reviews. 2008. PMID: 21433399.
12. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-9. PMID: 19565683.
13. Correll CU, Manu P, Olshanskiy V, et al. Cardiometabolic risk of second-generation antipsychotic medications during first-time use in children and adolescents. JAMA 2009 Oct 28;302(16):1765-73. PMID: 19861668.
14. Lenth RV. Java Applets for power and sample size [Computer software]. 2006-9. Available at: http://www.stat.uiowa.edu/~rlenth/Power. Accessed December 5, 2011.
15. Tohen M, Kryzhanovskaya L, Carlson G, et al. Olanzapine versus placebo in the treatment of adolescents with bipolar mania. Am J Psychiatry 2007 Oct;164(10):1547-56. PMID: 17898346.
16. Haas M, Delbello MP, Pandina G, et al. Risperidone for the treatment of acute mania in children and adolescents with bipolar disorder: a randomized, double-blind, placebo-controlled study. Bipolar Disord 2009 Nov;11(7):687-700. PMID: 19839994.
17. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med. 2000 Nov 30;19(22):3127-31. PMID: 11113947.
18. Citrome L. Compelling or irrelevant? Using number needed to treat can help decide. Acta Psychiatr Scand 2008 Jun;117(6):412-9. PMID: 18479317.
19. Brant R. Inference for proportions: comparing two independent samples [Computer software]. Available at: http://www.stat.ubc.ca/~rollin/stats/ssize/b2.html. Accessed December 6, 2011.

Abbreviations

ADHD	attention deficit hyperactivity disorder
AHRQ	Agency for Healthcare Research and Quality
CER	comparative effectiveness review
CGAS	Children's Global Assessment Scale
CI	confidence interval
DBDs	disruptive behavior disorders
EG	evidence gap
EHC	Effective Health Care
EMR	electronic medical record
EPS	extra-pyramidal symptoms
ES	effect size
FDA	U.S. Food and Drug Administration
FGA	first-generation antipsychotics
FRN	future research needs
HIPAA	Health Insurance Portability and Accountability Act of 1996
HRQL	health-related quality of life
ICTRP	International Clinical Trials Registry Project
KQ	key question
NIH	National Institutes of Health
NMS	neuroleptic malignant syndrome
NNT	number needed to treat
NRCTs	nonrandomized controlled trials
OCD	obsessive-compulsive disorder
OR	odds ratio
PANSS	Positive and Negative Symptoms Scale
PDDs	pervasive developmental disorders
PICOTS	populations, interventions, comparators, outcomes, time frames, and settings
RCTs	randomized controlled trials
RN	research need
RR	relative risk
SGA	second-generation antipsychotics
SOE	strength of evidence

Appendix A. Tables of FDA-Approved Indications for First- and Second-Generation Antipsychotics

Table A-1. Food and Drug Administration-approved first-generation antipsychotics

Generic Name	Indications	Age Group for Which Approved
Chlorpromazine	Schizophrenia	Adults and children (1–12 years)
	Bipolar disorder (mania)	
	Hyperactivity	
	Severe behavioral problems	
Droperidol	Agitation	Adults and children
Fluphenazine	Psychotic disorders	Adults
Haloperidol	Schizophrenia	Adults
	Tourette syndrome	
	Hyperactivity	
	Severe childhood behavioral problems	
Loxapine	Schizophrenia	Adults and children ≥12 years
Perphenazine	Schizophrenia	Adults and children ≥12 years
Pimozide	Tourette syndrome	Adults and children ≥12 years
Prochlorperazine	Schizophrenia	Adults and children >2 years and >20 pounds
	Generalized nonpsychotic anxiety	Adults
Thiothixene	Schizophrenia	Adults and children ≥12 years
Thioridazine	Schizophrenia	Adults and children
Trifluoperazine	Schizophrenia	Adults and children ≥6 years
	Generalized nonpsychotic anxiety	Adults

Table A-2. Food and Drug Administration-approved second-generation antipsychotics

Generic Name	Indications	Age Group for Which Approved
Aripiprazole	Schizophrenia	Adults and adolescents (13–17 years)
	Bipolar disorder (manic/mixed) monotherapy or adjunctive to lithium or valproate	Adults and children (10–17 years)
	Adjunctive treatment of major depressive disorder	Adults
	Irritability Associated with autistic disorder	Children (6–17 years)
	Acute treatment of agitation	Adults
Asenapine	Acute schizophrenia	Adults
	Bipolar disorder type 1 (manic/mixed)	
Clozapine	Treatment resistant schizophrenia	Adults
	Reduce the risk of suicidal behavior in younger patients with schizophrenia.	
Iloperidone	Acute schizophrenia	Adults
Olanzapine	Schizophrenia Bipolar disorder (manic/mixed)	Adults and adolescents (13–17 years)
	Bipolar disorder	Adults
	Treatment resistant depression	
	Agitation associated with schizophrenia and bipolar I mania	
Paliperidone	Schizophrenia Schizoaffective disorder	Adults
Quetiapine	Schizophrenia	Adults and adolescents (13–17 years)
	Bipolar disorder (acute manic)	Adults, children, and adolescents (10–17 years)
	Bipolar disorder (depression)	Adults
	Bipolar disorder (maintenance)	
	Adjunctive therapy for major depressive disorder	
Risperidone	Schizophrenia	Adults and adolescents (13–17 years)
	Bipolar disorder (manic/mixed)	Adults and adolescents (10–17 years)
	Irritability associated with autism	Children (5–16 years)
Ziprasidone	Schizophrenia	Adults
	Bipolar disorder (manic/mixed)	
	Bipolar disorder (maintenance)	
	Acute agitation in patients with schizophrenia	

Appendix B. Evidence Gaps Identified in the Draft Comparative Effectiveness Review

"First- and Second-Generation Antipsychotics for Children and Young Adults"

Reviewed on August 8, 2011

Note: All of the evidence gaps listed below refer to the population of children, adolescents, and young adults (24 years of age or younger).

I. Evidence gaps for specific disorders

Pervasive Developmental Disorders including Autistic Disorder, Asperger's Syndrome, and Pervasive Developmental Disorder not otherwise specified

1. The evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for the treatment of pervasive developmental disorders including autistic disorder, Asperger's syndrome, and pervasive developmental disorder not otherwise specified. Key Question (KQ) 1

Attention Deficit Hyperactivity Disorder and Disruptive Behavior Disorders

2. For those with attention deficit hyperactivity disorder and disruptive behavior disorders, there is moderate strength of evidence to support 2nd generation antipsychotics when compared with placebo for improving behavior symptoms and clinical global impressions. However, the evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for other clinically meaningful outcomes of interest**.** These other outcomes include: aggression, anxiety, social/occupational functioning, health related quality of life (HRQL), legal interactions, medication adherence, patient and parent-reported outcomes**,** and suicide related behavior. KQ1

Bipolar Disorder

3. For those with bipolar disorder, there was moderate strength of evidence to support 2nd generation antipsychotics over placebo for clinical global impressions. However, the evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for other outcomes such as: aggression, depression, manic symptoms, social/occupational functioning, health related quality of life (HRQL), legal interactions, medication adherence, patient and parent-reported outcomes, and suicide related behavior. KQ1

Schizophrenia

4. For those with schizophrenia, there was moderate strength of evidence to support 2nd generation antipsychotics over placebo for several outcomes including Children's Global

Assessment Scale (CGAS), clinical global impressions, and the Positive and Negative Symptoms Scale (PANSS). However, the evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is low or insufficient for other clinically meaningful outcomes such as: aggression, depressions, manic symptoms, social/occupational functioning, HRQL, legal interactions, medication adherence, patient and parent-reported outcomes, and suicide related behavior. KQ1

Tourette syndrome

5. For those with Tourette syndrome, there was moderate strength of evidence favoring 2nd generation antipsychotics over placebo for tic symptom improvement. However, the evidence regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics is otherwise low or insufficient for other clinically meaningful outcomes. These other outcomes include: clinical global impressions, obsessive-compulsive symptoms, social/occupational functioning, HRQL, medication adherence, patient and parent reported outcomes, and suicide related behavior. KQ1

Obsessive Compulsive Disorder, Eating Disorders, and Post Traumatic Stress Disorder

6. There is a paucity of data regarding the efficacy and effectiveness of 1st or 2nd generation antipsychotics for the treatment of Obsessive Compulsive Disorder, Eating Disorders, and Post Traumatic Stress Disorder. KQ1

II. Evidence gaps for all included disorders

7. The extant literature lacks sufficient evidence regarding comparisons between and within classes of 1st and 2nd generation antipsychotics for any of the included disorders, behaviors, or outcomes of concern for children and adolescents. KQ1, 2

8. For children and adolescents, the extant literature lacks studies examining the long-term (≥6 months follow up) efficacy and effectiveness of 1st and 2nd generation antipsychotics in all disorders and behaviors of interest. The median study duration of 8 weeks may be insufficient for evaluation of long term outcomes including outcomes important to parents and patients such as school performance, emotional development, or legal system interactions. KQ1, KQ2

9. In children and adolescents, the extant literature lacks studies examining the long-term (≥ 6 months follow up) safety of 1st and 2nd generation antipsychotics in all disorders and behaviors of interest. The median study duration of 8 weeks may be insufficient for evaluating long term adverse outcomes such as obesity, diabetes, or cardiovascular events. KQ2, KQ3

III. Evidence gaps for adverse events and subpopulations

10. In children and adolescents, while there is moderate strength of evidence to favor placebo over several 2nd generation antipsychotics with respect to several adverse event outcomes of interest (e.g. sedation, extra-pyramidal symptoms (EPS), weight gain, dyslipidemia) and

there is moderate strength of evidence favoring some 2nd generation antipsychotics over others (e.g. risperidone compared with olanzapine) in regards to weight gain, overall there is insufficient evidence to allow for comparisons between and within classes of 1st and 2nd generation antipsychotics for any adverse event outcome. These outcomes include sedation, EPS, weight gain/body composition, insulin resistance, sexual adverse events, and dyslipidemia. KQ3

11. The extant literature lacks evidence to determine if there are differences in efficacy, effectiveness, or adverse events for any population sub-group. Sub-groups include sex, age, race, co-morbidities, co-treatment, history of psychosis, or duration of illness. KQ4

IV. Evidence gaps for methodological issues

12. The extant literature lacks the use of standardized pediatric side-effect scales (e.g., the Safety Monitoring Uniform Report Form). KQ2

13. The extant literature lacks consistent and comparable outcomes and outcome measurements across the studied disorders and behaviors of concern. KQ3, Methods

14. The extant literature demonstrates a lack of consensus on minimal clinically important differences within many outcomes of interest across disorders. KQ3, Methods

15. The extant literature lacks large-scale effectiveness studies that are generalizable to the broader population seen in clinical practices. Methods/All KQs

V. Overarching methods related evidence gaps

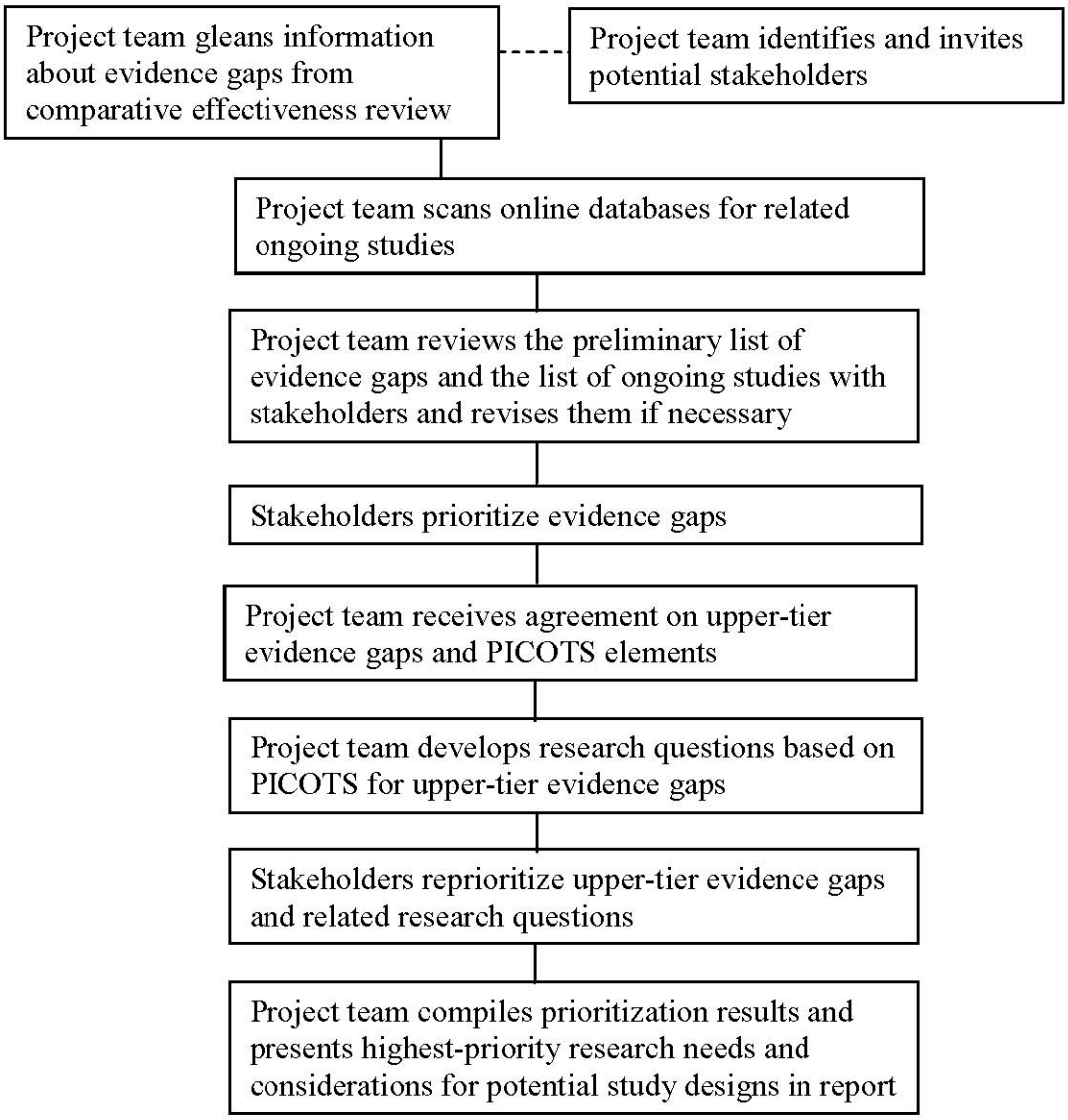
Note: We invite stakeholder input on these gaps, but these will not be part of the list of evidence gaps that will be prioritized. These methodological shortcomings of existing research will be considered when we develop proposed research designs for the highest ranked research needs.

16. The extant literature lacks efficacy studies with adequate blinding of study participants and outcome assessors, the adequate concealing of allocation and the appropriate handling and reporting of missing data. Methods/All KQs

17. The extant literature lacks independent/investigator driven research efforts which increases the potential for overestimated treatment effects associated with industry-funded research. Methods/All KQs

Appendix C. Flowchart of Future Research Needs (FRN) Process

Figure C-1. Flowchart of Future Research Needs (FRN) process



Appendix D. Effective Health Care Program Selection Criteria

Appropriateness:

- Represents a health care drug, intervention, device, technology or health care system/setting available (or soon to be available) in the United States.
- Relevant to 1013 enrollees (Medicare, Medicaid, S-CHIP, other federal health care programs).
- Represents one of the priority conditions designated by the U.S. Department of Health and Human Services (HHS).

Importance:

- Represents a significant disease burden, large proportion or priority population.
- Is of high public interest; affects health care decision-making, outcomes, or costs for a large proportion of the U.S. population or for a priority population in particular.
- Was nominated/strongly supported by one or more stakeholder groups.
- Represents important uncertainty for decisionmakers.
- Incorporates issues around both clinical benefits and potential clinical harms.
- Represents important variation in clinical care, or controversy in what constitutes appropriate clinical care.
- Represent high costs to consumers, patients, health care systems or payers; due to common use, high unit costs, or high associated costs.

Desirability of New Research/Duplication:

- Would not be redundant (i.e., the proposed topic is not already covered by available or soon-to-be available high quality systematic review by AHRQ or others).

Feasibility:

- Effectively uses existing research and knowledge by considering adequacy of research for conducting a systematic review, and newly available evidence

Potential Impact:

- Potential for significant health impact, significant economic impact, potential change, potential risk from inaction, addressing inequities and vulnerable populations, and/or addressing a topic with clear implications for resolving important dilemmas in health and health care decisions made by one or more stakeholder groups.

Appendix E. Searches for Ongoing Research May 23, 2011/June 6, 2011; and List of Relevant Ongoing Studies

I. NIH Reporter – limited to Active Projects and Award Notice Date greater than February 23, 2010

antipsychotics, antipsychotic

Using the OR radio button

= **264 results**

II. HSRProj – Advanced Search screen. Used the same search as the CER draft report appendix; limited to projects with a final year in the range of 2010–2018 (2018 is as high as the search limits go)

((asperger) OR (autistic disorder) OR autism OR schizophrenia OR (depression) OR (bipolar disorder) OR (obsessive-compulsive) OR (post-traumatic) OR (anorexia nervosa) OR anorexia) AND (antipsychotics)

= **8 results**

III. WHO ICTRP (International Clinical Trials Registry Platform)

Advanced Search screen:
http://apps.who.int/trialsearch/AdvSearch.aspx

Search in Condition box:
((asperger) OR (autistic disorder) OR autism OR schizophrenia OR (depression) OR (bipolar disorder) OR (obsessive-compulsive) OR (post-traumatic) OR (anorexia nervosa) OR anorexia) AND (antipsychotics) AND (child OR adolescent OR pediatric OR infant) = **0 results**

(asperger OR autistic disorder OR autism OR schizophrenia OR depression OR bipolar disorder OR obsessive-compulsive OR post-traumatic OR anorexia nervosa OR anorexia) AND (antipsychotics) AND (child OR adolescent OR pediatric OR infant) = **0 results**

Date of registration between 02/23/2010 and 05/23/2011

Status = Recruiting (other option is “all”; when I try the second strategy with all, still 0. When I try the **search** in the **title field instead of condition** and limit to recruiting studies only, **658 records for 651 trials found.**

IV. **ClinicalTrials.gov**—same strategy as with ICTRP, using “Search Terms” box. In advanced search screen, limited to child+adult (because original search sought infants–24, and child only covers infants–17. This only added one more trial.)

First Received: From 02/23/2010 to 05/23/2011.

((asperger) OR (autistic disorder) OR autism OR schizophrenia OR (depression) OR (bipolar disorder) OR (obsessive-compulsive) OR (post-traumatic) OR (anorexia nervosa) OR anorexia) AND (antipsychotics) AND (child OR adolescent OR pediatric OR infant)

= 27 results

List of Ongoing Studies of First- and Second-Generation Antipsychotics in Children and Young Adults as of May 2011

We searched NIH RePORTER, HSRProj, ClinicalTrials.gov, and International Clinical Trials Registry Platform (ICTRP) and obtained 951 records. We dually reviewed information from each record and identified 60 ongoing studies (after removing duplicates) that potentially met the inclusion criteria for the First- and Second-Generation Antipsychotics for Children and Young Adults CER.

Table. E-1. List of ongoing studies of first- and second-generation antipsychotics in children and young adults as of May 2011

Reg #	Title	PI	Funder
NCT01098110	6-Week Trial of the Efficacy and Safety of Asenapine Compared to Placebo in Subjects With an Acute Exacerbation of Schizophrenia (Study P06124)	Multi-center	Schering-Plough
NCT01136772	A Comparison of Long-Acting Injectable Medications for Schizophrenia	Scott Stroup, MD, MPH, Columbia University and Joseph P McEvoy, MD, Duke University	National Institute of Mental Health (NIMH)
NCT01149655	A Long-Term Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy, Safety, and Tolerability of Aripiprazole (OPC 14597) as Maintenance Treatment in Adolescent Patients	Eva Kohegyi, MD Otsuka Pharmaceutical Development and Commercialization, Inc.	Otsuka Pharmaceutical Development & Commercialization, Inc.
JPRN-JapicCTI-101147	A Long-Term, Extended Treatment Study of Aripiprazole in Pediatric Patients with Schizophrenia	Multi-Center (Japan)	Otsuka Pharmaceutical Co., Ltd.
7R34MH080791-04	A Novel Mulitmodal Intervention for Children with ADHD and Impaired Mood	James Waxmonsky, Florida International University	National Institute of Mental Health (NIMH)
JPRN-MIN000005355	A randomized open label study of the effects of aripiprazole in overweight and obstructive sleep apnea subjects with schizophrenia or schizoaffective disorder switched from other antipsychotic drugs.	Junichi Murakami, Biwako Hospital Department of Psychiatry	Biwako Hospital
NCT01212575	A Retrospective NIS to Evaluate the Use of Seroquel XR and IR in the Clinical Practice of Outpatients With Schizophrenia	Charlotte Emborg, Skovager 2, Risskov, Denmark	AstraZeneca
JPRN-JapicCTI-101146	A Short Treatment Study of Aripiprazole in Pediatric Patients with Schizophrenia	Multi-Center	Otsuka Pharmaceutical Co., Ltd.
NCT01299389	A Study of Paliperidone Palmitate in Patients With Schizophrenia	None named	Janssen Pharmaceutical K.K.
NCT01206517	A Study to Evaluate the Safety and Tolerability of Sublingual Asenapine in a Pediatric Population With Schizophrenia or Bipolar I Disorder (Study P06522)	None named	Schering-Plough
NCT01338298	Adjunctive Aripiprazole	Deanna L. Kelly Pharm.D., BCPP, Maryland Psychiatric Research Center	University of Maryland

Table. E-1. List of ongoing studies of first- and second-generation antipsychotics in children and young adults as of May 2011 (continued)

Reg #	Title	PI	Funder
NCT01197404	Affect Management for Early Adolescents	Christopher D. Houck, Rhode Island Hospital	Rhode Island Hospital
5R01MH075921-05	Antimanic Use During Pregnancy	Katherine Wisner, University Of Pittsburgh	National Institute of Mental Health (NIMH)
NCT01082848	ARIpiprazole in Anorexia NErvosa	Jaime Moyá, Hospital Clinic of Barcelona	Hospital Clinic of Barcelona
5R01MH078576-05	Behavioral Intervention to Reduce Novel Antipsychotic Medication Health Risks	Jeffrey Kelly, Medical College of Wisconsin	National Institute of Mental Health (NIMH)
NCT01333072	Biomarkers in Autism of Aripiprazole and Risperidone Treatment (BAART)	C. Lindsay DeVane, PharmD, Medical University of South Carolina	Medical University of South Carolina
NCT01164059	Clinical Effectiveness of Newer Antipsychotics in Comparison With Conventional Antipsychotics in Schizophrenia	Dr. Jürgen Timm	University of Bremen
5R01MH082839-03	Collaboration to Advance Negative Symptom Assessment in Schizophrenia	Jack Blanchard, University of Maryland	National Institute of Mental Health (NIMH)
(From HSR Proj)	Comparative Safety and Effectiveness of Antipsychotics	Stephen Crystal, Institute for Health, Health Care Policy and Aging Research, Rutgers, The State University of New Jersey	Agency for Healthcare Research and Quality
1R01HS018550-01A1	Comparative Safety of Atypical Antipsychotics in High-Risk U.S. Children with ADHD	David Rubin, Children's Hospital of Philadelphia	Agency for Healthcare Research and Quality
NCT01213836	Compare the Effect on Cognitive Functioning of Two Formulations of Seroquel, Seroquel XR and IR in Patients With Stable Schizophrenia	Eva Dencker Vansvik	AstraZeneca
(From HSR Proj)	Comparison of Long-Acting Injectable Medications for Schizophrenia--ACLAIMS	TS Stroup, NY Psychiatric Institute	National Institute of Mental Health (NIMH)
5K23MH085005-03	Counteracting Risperidone-Induced Hyperprolactinemia in Youths	Chadi Calarge, University of Iowa	National Institute of Mental Health (NIMH)
(From HSR Proj)	DNA Diagnostics for Minimizing Metabolic Side-Effects of Antipsychotics	Gualberto Ruano, Genomas Inc.	National Institute of Mental Health (NIMH)
5R01HS018577-02	Drug Cost Containment Changes and Quality of Care for Mentally Ill Dual Enrollees	Steven Soumerai, Harvard Pilgrim Health Care	Agency for Healthcare Research and Quality
NCT01160679	Effectiveness of Atypical Antipsychotics on Anhedonic Features in Patients With Schizophrenia	Sang-Woo Han, M.D., Ph.D., Department of Psychiatry	Astra-Zeneca
NCT01222793	Effects of Antipsychotics on Eating and Food Craving in People With Schizophrenia	Kimberly Warren, Ph.D., University of Maryland, Baltimore	National Institute on Drug Abuse (NIDA)
NCT01244815	Efficacy and Safety of Asenapine Treatment for Pediatric Bipolar Disorder (P06107)	None named	Schering-Plough

Table. E-1. List of ongoing studies of first- and second-generation antipsychotics in children and young adults as of May 2011 (continued)

Reg #	Title	PI	Funder
NCT01157559	Efficacy and Safety With Ziprasidone in First-Episode Psychosis	Young Chul Chung, Professor of Psychiatry, Chonbuk National University Hospital, Jeonju, Korea	Chonbuk National University Hospital
NCT01291511	Efficacy in Prevention of Relapse of Schizophrenia in Subjects Taking Either Placebo or Iloperidone.	None named	Novartis Pharmaceuticals
NCT01256177	Evaluate the Efficacy and Safety of Quetiapine Fumarate (SEROQUEL) Extended Release as Monotherapy in the Treatment of Patients With Bipolar Depression	None named	AstraZeneca
NCT01184443	Evaluation of the Efficacy and Safety of Olanzapine for Anorexia Nervosa in Children and Adolescents	Wendy J Spettigue, MD, FRCPC, Children's Hospital of Eastern Ontario	Children's Hospital of Eastern Ontario
NCT01349907	Extension Study of Asenapine for Pediatric Bipolar Disorder (P05898 AM1)	None named	Schering-Plough
NCT01190254	Fixed Dose Efficacy and Safety Study of Asenapine for the Treatment of Schizophrenia in Adolescents (Study P05896)	None named	Schering-Plough
NCT01190267	Flexible Dose, Long-Term Safety Study of Asenapine for the Treatment of Schizophrenia in Adolescents (Study P05897)	None named	Schering-Plough
1R03HS019024-01A1	Identifying Treatment-Resistant Depression in Automated Databases	Darren Toh, Harvard Pilgrim Health Care Inc.	Agency for Healthcare Research and Quality
5R01MH080325-03	Improving Metabolic Parameters of Antipsychotics Child Treatment	Linmarie Sikich, University North Carolina–Chapel Hill	NIMH
NCT01155544	Improving Outcomes in Psychosis Associated With Substance Use Using Aripiprazole	Serge Sevy, M.D., M.B.A., Feinstein Institute for Medical Research	NIMH
1R15MH094955-01	Long-Term Effects of Early-Life Antipsychotic Drug Treatment	Mark E. Bardgett, Northern	National Institute of Mental Health (NIMH)
NCT01142596	Long-Term Extension Trial of Asenapine in Subjects With Schizophrenia (Study P06125)	None named	Schering-Plough
3R21MH080968-02S1	Long-Term Safety and Genetic Risk Factors of Risperidone Treatment in Youth	Chadi Calarge, University of Iowa	National Institute of Mental Health (NIMH)
5R01MH045404-19	Maximizing Treatment Outcome in OCD	Edna Foa, University of Pennsylvania	National Institute of Mental Health (NIMH)
NCT01184235	Multidimensional Measurement of Psychopharmacological Treatment Response	Bill J Duke, M.A., Ph.D., Child Psychopharmacology Institute	Child Psychopharmacology Institute
5K23MH079498-02	Neuroprotective/Neurodevelopmental Effects-Antipsychotics in Adolescent Psychoses	Karin Borgmann-Winter, Children's Hospital of Philadelphia	National Institute of Mental Health (NIMH)
NCT01170117	Olanzapine Versus Placebo for Outpatients With Anorexia Nervosa	Evelyn Attia, MD, Columbia University Medical Center	National Institute of Mental Health (NIMH)

Table. E-1. List of ongoing studies of first- and second-generation antipsychotics in children and young adults as of May 2011 (continued)

Reg #	Title	PI	Funder
NCT01227668	OPT - Phase IV Long Term Maintenance Study of Aripiprazole for the Treatment of Irritability Associated With Autistic Disorder	None named	Bristol-Myers Squibb
NCT01154829	Pan European Collaboration on Antipsychotic Naive Schizophrenia (PECANS)	Birte Y Glenthoj, professor, University of Copenhagen, Psychiatric Center Glostrup	University of Copenhagen
NCT01075295	Prevention of Weight Gain in Early Psychoses	Rohan Ganguli, MD, Centre for Addiction and Mental Health	Centre for Addiction and Mental Health
NCT01282281	Prospective Metabolic Monitoring of Youth and Adults With Bipolar Disorder	Ayal Schaffer, MD and Benjamin Goldstein, MD, Sunnybrook Health Sciences Centre	Sunnybrook Health Sciences Centre
NCT01171937	Risperidone Treatment in Children With Autism Spectrum Disorder and High Levels of Repetitive Behavior	James T. McCracken, M.D., UCLA	University of California, Los Angeles; National Institutes of Health
NCT01117220	Safety and Efficacy of Ziprasidone In Children and Adolescents With Bipolar I Disorder (Manic Or Mixed) (Protocol A1281196)	None named	Pfizer
NCT01122927	Safety and Tolerability of Aripiprazole in Adolescents With Schizophrenia or Children and Adolescents With Bipolar I Disorder, Manic or Mixed Episode With or Without Psychotic Features.	None named	Otsuka Pharmaceutical Development & Commercialization, Inc.
NCT01124877	Safety and Tolerability of Flexible Doses of Oral Ziprasidone In Children and Adolescents With Bipolar I Disorder (Manic or Mixed)	None named	Pfizer
NCT01269710	Second-Generation Antipsychotic Treatment Indication Effectiveness and Tolerability in Youth (Satiety) Study	Linmarie Sikich, MD, UNC-Chapel Hill	University of North Carolina, Chapel Hill; Foundation of Hope, North Carolina
5R01MH080050-04	Stepped Pharmacotherapy for Aggressive Youth with ADHD	Joseph Blader, State University New York Stonybrook	National Institute of Mental Health (NIMH)
5R01MH077750-03	Stimulant and Risperidone for Youth with Severe Physical Aggression	Robert Findling, Case Western Reserve University	National Institute of Mental Health (NIMH)
2U01MH062565-06A2	Sustaining Remission of Psychotic Depression	Ellen White, University of Pittsburg	National Institute of Mental Health (NIMH)
NCT01119014	Tolerance and Effect of Antipsychotics in Children and Adolescents With Psychosis	Katrine Pagsberg, Bispebjerg Centre for Child and Adolescent Psychiatry, University of Copenhagen	University of Copenhagen
5R01MH081235-04	Ziprasidone Augmentation of SSRIs for Treatment-Resistant Depression (TRD)	Richard C. Shelton, Vanderbilt University	National Institute of Mental Health (NIMH)

Table. E-1. List of ongoing studies of first- and second-generation antipsychotics in children and young adults as of May 2011 (continued)

Reg #	Title	PI	Funder
NCT01172652	Ziprasidone in Bipolar Disorder With Comorbid Lifetime Panic or Generalized Anxiety Disorder(GAD)	Trisha Suppes, MD, PhD, VA Palo Alto Health Care System & Stanford School of Medicine	VA Palo Alto Health Care System

Appendix F. Online Prioritization Exercises

Figure F-1. Prioritization Exercise #1

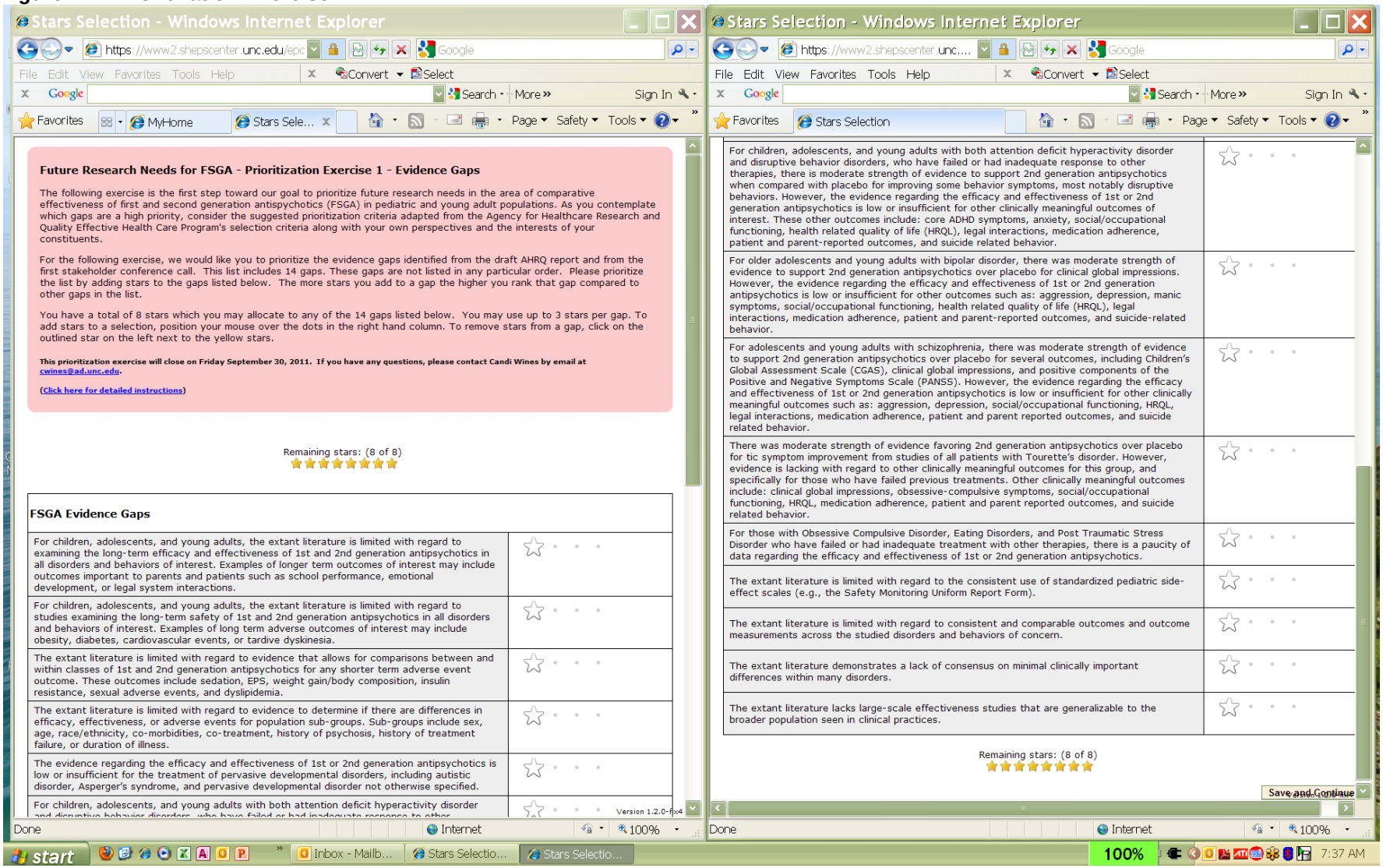
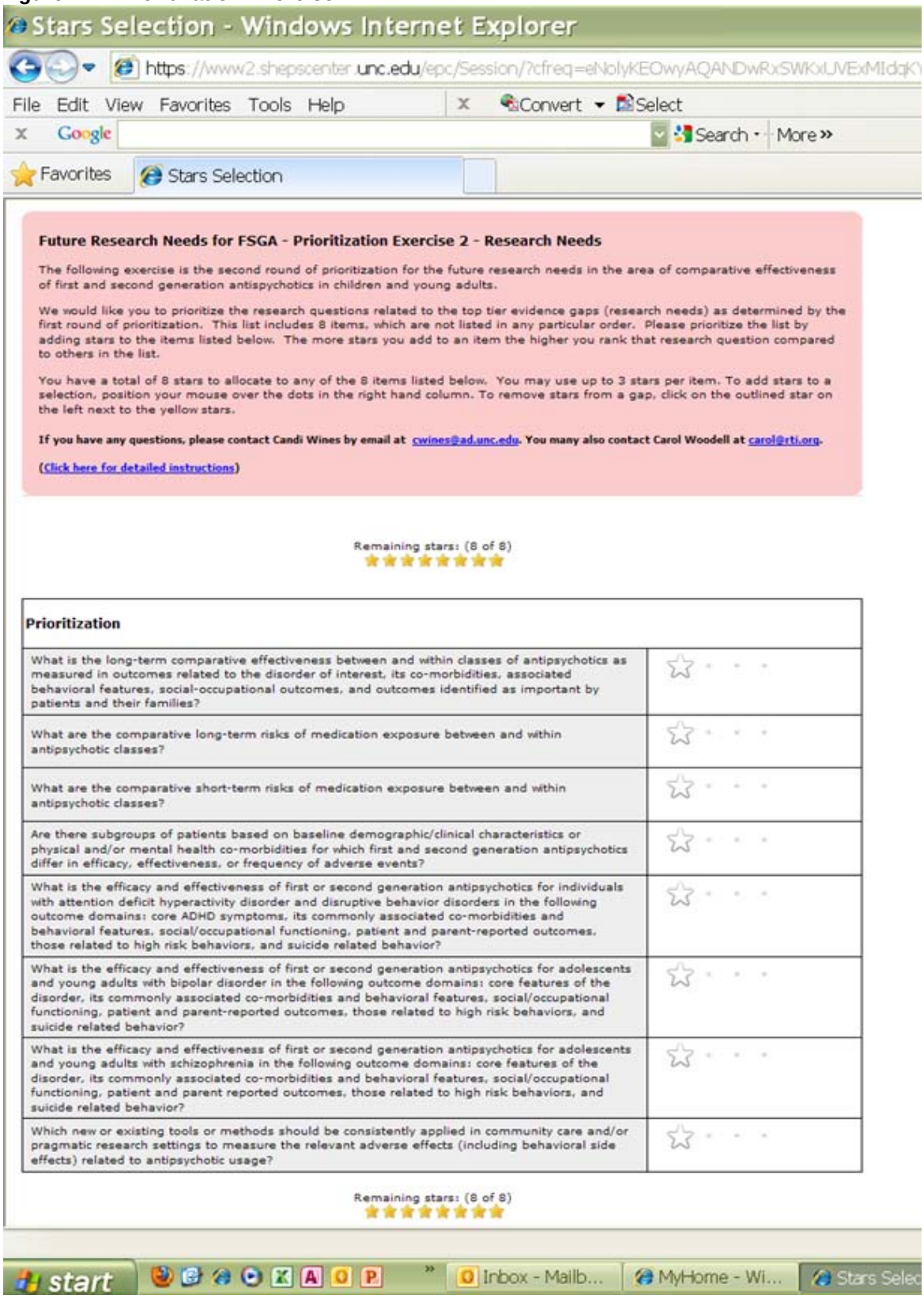


Figure F-2. Prioritization Exercise #2



Appendix G. Additional Relevant Characteristics and Study Design Considerations

Table G-1. Additional relevant characteristics and study design considerations

		Producing Valid Results and Bias	Ability to Recruit	Resource Use, Size and Duration	Ethical, Legal, and Social Issues
Clinical Trials	**RCT**	Best method to control for selection bias and both measured and unmeasured differences between groups at baseline, but this may come at the cost of generalizability. If feasible, this design is likely to produce the most valid results.	Challenge that patients tend to be reluctant to accept randomization.	An RCT has to be large, because the question compares active treatments, and the effect size may be small and easily swamped by other causes of morbidity and mortality in this population. Likely to require substantial resources to recruit large enough sample to evaluate short- and long-term outcomes.	Concerns may occur when treatment is assigned through random allocation. Patients and treating clinicians must perceive equipoise across interventions when invited to participate in a research study. Careful stopping and reporting rules will be important if evidence of significant benefit or harm is found. RCTs are typically performed when there is equipoise on the optimal treatment.
	NRCT	Ideal design consideration when randomization at the level of the individual is not feasible or ethical. Treatment assignment is not randomized. Thus, consideration of selection bias and unmeasured confounders need to be controlled. The results of assignment of practices will increase the generalizability of the results greatly but will sharply reduce the validity of the results. May be the optimal study design when treatment or exposure cannot ethically or practically be assigned. Baseline characteristics can be measured but may not be balanced between the two groups. If sample size is large, this design may be the best method to assess subgroup effects or incidence of harms. Statistical techniques to adjust for baseline differences may not completely control for potential bias. Crossover may occur, especially in longer duration observation.	Acceptability to potential subjects is better than randomized trials, in which a group is assigned at random. Some disadvantage in that subjects may be less willing to accept any assignment as opposed to patient choice.	Resources needed are generally high but can be less expensive than randomized trials. Study duration will depend on the underlying condition being assessed. For example, duration of acute infectious diseases may require only a brief followup. Because these are prospective studies, duration is longer than retrospective studies. Size will depend on the effect size being sought between intervention and comparator groups.	Concerns may occur when treatment is assigned by investigators. Patients and treating clinicians must perceive equipoise across interventions when invited to participate in a research study. Careful stopping and reporting rules will be important if evidence of significant benefit or harm is found.

Table G-1. Additional relevant characteristics and design considerations (continued)

		Producing Valid Results and Bias	Ability to Recruit	Resource Use, Size and Duration	Ethical, Legal, and Social Issues
Cohort Studies	**Prospective Study**	A study in which individuals in the group without the outcome(s) of interest (e.g., disease) are classified according to exposure status at baseline (exposed or unexposed) and then are followed over time to determine if the development of the outcome of interest is different in the exposed and unexposed groups. Although high potential for confounding, appropriate repeated measures may increase ability to control for confounding by indication.	More acceptable than designs in which treatment is assigned. Ability to recruit may depend on the respondent burden.	Resources needed are moderate to high; similar to a prospective trial, although less than a study design in which treatment is assigned. Size and duration will depend on the natural history of the condition under study and the effect size or incidence of harm thought to be clinically important.	Because patients and/or providers select the treatment, few ethical issues, although careful stopping and reporting rules will be important if evidence of significant benefit or harm is found.
	Retrospective Study	A study in which a group of individuals is identified by predetermined common features. The group is usually assembled using available data sources (e.g., administrative data). Individuals are classified according to exposure status (exposed or unexposed) at the time the group existed and are followed up to a prespecified endpoint to determine if the development of the outcome of interest is different in the exposed or unexposed group. Subjects are followed over time. The duration of the study is short because it is retrospective. Significant risk of selection bias, as subjects either choose their treatment or it is assigned by a health care provider. The generalizability of the study result will depend on the population sampled.	Very feasible, main concern is that of selection bias, depending on the source of the secondary data and missing variables. Negotiations with the holders of the secondary data may take significant time.	Less expensive because data used are already collected. The duration of the study is short relative to studies involving primary data collection. However, such studies often take somewhat longer than envisioned. Sample size may be very large given the increasing availability of large administrative claims and EMR databases.	Confidentiality and Health Insurance Portability and Accountability Act issues may arise when diverse databases are linked without specific patient consent.

Table G-1. Additional relevant characteristics and design considerations (continued)

		Producing Valid Results and Bias	Ability to Recruit	Resource Use, Size and Duration	Ethical, Legal, and Social Issues
	Meta-Analysis of Individual Patient Data	Appropriate for pooling data from existing trials for simultaneous data analysis. In a meta-analysis of individual patient data, individual data from existing studies are brought together using harmonized definitions and reanalyzed according to a prespecified protocol. When individual participant data are available, there is opportunity to standardize outcomes and exposure definitions, and use much more powerful analyses. A major advantage over a meta-analysis of group data is the ability to examine patient-level modifiers of the treatment effect (i.e., patient-level factors such as age, sex, and disease severity indices).	Very feasible, main concern is that of selection bias, depending on the source of the secondary data and missing variables. Negotiations with the holders of the secondary data may take significant time.	Less expensive because data used are already collected. A meta-analysis of individual patient data can take longer to complete than an analysis of a readily available dataset from a single study. Logistical complications include but are not limited to identification of data sources, convincing investigators to participate, standardizing definitions of interventions and outcomes, complying with Health Insurance Portability and Accountability Act and harmonizing definitions of exposures and outcomes across datasets.	Confidentiality and Health Insurance Portability and Accountability Act issues may arise when diverse databases are linked without specific patient consent.
Other	**Case Control Study**	A study in which participants are identified based on the known outcomes of interest (e.g., medication exposure). Exposure status is then collected based on the participants' past experiences. Exposure status is compared between the two (or more) groups: those who have the outcome of interest and those who do not have the outcome of interest (controls). This is a retrospective study that collects data on events that have already occurred.	Very feasible, main concern is that of selection bias, depending on the source of the secondary data and missing variables. Negotiations with the holders of the secondary data may take significant time.	Less expensive because data used are already collected. The duration of the study is short relative to studies involving primary data collection. However, such studies often take somewhat longer than envisioned. Sample size may be very large given the increasing availability of large administrative claims and EMR databases.	Confidentiality and Health Insurance Portability and Accountability Act issues may arise when diverse databases are linked without specific patient consent. Otherwise, issues are minimal, given study is observational.

Abbreviations: EMR = electronic medical record; NRCE = nonrandomized controlled trial; RCT = randomized controlled trial