Evidence-based Practice Center Systematic Review Protocol

Project Title: *Efficacy and Comparative Effectiveness of Off-Label Use of Atypical Antipsychotics – Update*

I. Background and Objectives for the Systematic Review

Atypical antipsychotics are approved by the United States Food and Drug Administration (FDA) for treatment of schizophrenia and bipolar disorder. These drugs have been studied for off-label use in several conditions. A 2007 study on Efficacy and Comparative Effectiveness of Off-label Use of Atypical Antipsychotics reviewed the scientific evidence on the safety and effectiveness for off-label uses of these drugs.[1] The study examined 84 published studies on atypical antipsychotics and found that the most common off-label uses of the drugs were treatment of depression, obsessive-compulsive disorder, posttraumatic stress disorder, personality disorders, Tourette's syndrome, autism, and agitation in dementia. It concluded that with few exceptions, there was insufficient high-quality evidence overall to reach conclusions about the efficacy of any off-label indications of these medications. It also found strong evidence that atypical antipsychotics can increase chances of adverse events such as significant weight gain, sedation, and gastrointestinal problems. Future research areas suggested by the study include safe treatment for agitation in dementia, association between the increased risk of death and antipsychotics drugs, and comparison of the development of adverse effects between patients taking atypical antipsychotics and those taking typical doses of conventional antipsychotics.

Since the publication of this report, some important changes have occurred and part of the conclusions of the original report might be out of date:

- New studies, especially new randomized controlled trials (RCTs), have provided more efficacy evidence on off-label use of atypicals (for example, new studies support efficacy of aripiprazole on agitation in dementia patients);
- New off-label uses of atypicals have been identified and examined (e.g., eating disorders, insomnia, attention deficit hyperactivity disorder (ADHD), anxiety, and substance abuse);
- Some previous off-label uses have been approved for on-label use (e.g., aripiprazole was approved as an adjunct for major depressive disorder in November 2007; olanzapine/fluoxetine combination and quetiapine were approved for bipolar depression; risperidone was approved for autism treatment in October 2006)
- Autism (included in the original systematic review[1]) and all on-label indications of atypicals will be reviewed in a study on the comparative effectiveness of typical and atypical antipsychotics conducted by another EPC. Analysis of data on children will be divided between us and this other EPC;
- Two new atypicals (asenapine and iloperidone) were recently approved by FDA for the treatment of schizophrenia and bipolar disorder; it is not clear if they have been prescribed for other conditions;
- New or increased adverse effects of off-label indications have been observed;
- The strength of evidence supporting some of the 2007 conclusions has increased.

Reflecting the above changes, many new studies have emerged. An update of the original

report [1] is needed to better understand the trend of off-label prescriptions and the risks and benefits associated with these off-label indications. Further, two key questions remain unclear due to insufficient information in the previous study: subpopulation (i.e., race / ethnicity, gender) that would benefit most from atypical antipsychotics, and appropriate dose and time limit. This update will try to address these two topics.

II. The Key Questions

Question 1: What are the leading off-label uses of atypical antipsychotics in the literature? How have trends in utilization changed in recent years, including inpatient versus outpatient use? What new uses are being studied in trials?

Population:

Adults, 18 years old and older, with the following disorders:
- Obsessive-compulsive disorder (OCD)
- Post-traumatic stress disorder (PTSD)
- Personality disorders (primarily borderline)
- Agitation in dementia (primarily in the elderly)
- Major depressive disorder (aripiprazole has already been approved)

Adults, 18 years old and older, as well as children (below 12 years old) and adolescents (12 – 17 years old) with the following disorders:
- Eating disorders, including anorexia nervosa and bulimia
- Attention deficit hyperactivity disorder (ADHD)
- Tourette's Syndrome
- Insomnia
- Anxiety disorders

Interventions:

Atypical antipsychotics approved by the U.S. FDA:
- Olanzapine
- Quetiapine
- Risperidone
- Ziprasidone
- Aripiprazole
- Paliperidone
- Asenapine
- Iloperidone

Comparators:

Four types of trials will be classified and examined:
- "Head-to-head" trials: trials that evaluate one atypical antipsychotic against another and provide direct evidence of comparative effectiveness;
- "Active" controlled trials: trials that compare an atypical antipsychotic with

another class of medication, often first generation antipsychotics
- "Placebo" controlled trials: trials that compare atypical antipsychotics with a placebo;
- "Augmentation" trials: trials that compare an antipsychotic taken with another medication with the other medication alone.

Outcomes: For efficacy, we will report commonly used objective outcomes such as symptom scores, response rates, laboratory data, and time to disease recurrence; for effectiveness, we will report outcome measures such as general health outcomes (e.g., SF-36), quality of life, and mortality. Outcomes are listed in appendix A.

Timing: The duration of the majority of the controlled trials in this area last from 6 to 26 weeks. However, per our technical expert panel's suggestion, we will not set a limit on the minimum trial length for efficacy for each disorder/condition – single dose or short term trials less than 6 weeks will be included.

Settings: All settings in which trials take place, including inpatient hospitalization, outpatient treatment and long-term care facilities.

Question 2: What does the evidence show regarding the efficacy and comparative effectiveness of atypical antipsychotics for off-label indications, such as depression?

> **Sub-Question 2**: How do atypical antipsychotic medications compare with other drugs, including first generation antipsychotics, for treating off-label indications?

Population: Same as in Question 1

Intervention: Same as in Question 1

Comparator: Same as in Question 1

Outcomes: Same as in Question 1. For the efficacy, we will report commonly used objective outcomes such as symptom scores, response rates, laboratory data, and time to disease recurrence; for the effectiveness, we will report outcome measures such as general health outcomes (e.g., SF36), quality of life, and mortality. Outcomes are listed in appendix A.

Timing: Same as in Question 1

Settings: Same as in Question 1

Question 3: What subset of the population would potentially benefit from off-label uses? Do efficacy, effectiveness, and harms differ by race/ethnicity, gender, and age group? By severity of condition and clinical subtype?

Population: Demographic, clinical, and severity subsets of the populations described in KQ1. Demographic subsets include different racial/ethnic groups, different age groups, and different genders. For clinical subsets, it is expected that only a small number of trials investigate specific subtypes (for example, inattentive versus hyperactive-impulsive type ADHD) which makes a comparative study infeasible. When data are available, clinical subtypes of the conditions of interest will be examined (for instance, combat-related PTSD and non combat-related PTSD). Severity subsets of population are categorized as groups with mild, moderate, or severe condition.

Intervention: Same as in Question 1

Comparator: Same as in Question 1

Outcomes: Efficacy and effectiveness outcomes are the same as in KQ1 and KQ2. Outcome measures for harms, or adverse events, are the same as in KQ4 below.

Timing: Same as in Question 1

Settings: Same as in Question 1

Question 4: What are the potential adverse effects and/or complications involved with off-label prescribing of atypical antipsychotics? How do they compare within the class and with other drugs used for the conditions?

Population: Same as in Question 1

Intervention: Same as in Question 1

Comparator: Same as in Question 1

Outcomes: All reported side effects and adverse events will be abstracted from clinical trials and large observational studies, regardless of study duration. Adverse events will be analyzed for all controlled trials as well as observational studies with at least 1,000 cases. Adverse events associated with on-label use of the atypical antipsychotic medications will be summarized. The primary focus will be on the following.

Mortality

Cardiovascular
Myocardial infarction
Arrhythmia - tachycardia
Blood pressure increase / decrease

Neurological
Cerebrovascular accident (CVA)
Akathisia
Extrapyramidal Symptoms

Tardive Dyskinesia
Sedation
Dizziness

Blood Dyscrasias
Neutropenia
Agranulocytosis
Leukopenia

Metabolic syndrome
Weight gain/loss
Hyperglycemia/diabetes
Hyperlipidemia

Timing: Same as in Question 1

Settings: Same as in Question 1

Question 5: What is the effective dose and time limit for off-label indications?

Population: Same as in Question 1

Intervention: Same as in Question 1

Comparator: Same as in Question 1

Outcomes: We will examine and summarize dose and trial length data, make comparison when data are available and comparable, and document the effective dose and time limit for all off-label indications by individual drugs.

Timing: Same as in Question 1

Settings: Same as in Question 1

III. Analytic Framework

Figure 1 presents the analytic framework for the update of this Comparative Effectiveness Review, with the five key questions depicted within the context described in the previous sections. First, by reviewing utilization data, surveys on prescribing patterns, and general information about the leading off-label uses, any new off-label uses and trends in utilization in the target populations will be summarized. Next, by using data from clinical trials and large cohort studies, evidence of benefits and harms in treating the mental health conditions will be documented. The evidence of benefits – efficacy and comparative effectiveness (versus placebo, versus other atypicals, or versus conventional therapy) of the off-label indications – will be evaluated separately for each of the atypical antipsychotics within condition (dementia, OCD,

PTSD, depression, etc.) via the examination of selected outcome measures, mainly symptom response rates measured by recognized psychometric tools.

Benefits and harms for specific subpopulations (by gender, age, and race/ethnicity) or related to other important factors (setting, severity of condition, length of use, and dosage) will be documented. Special attention will be given to identify the effective dose and time limit for off-label indications. The evidence of risks – adverse events associated with off-label indications – will be summarized, first within individual drugs across condition, and then compared within the class and with other drugs used for the conditions.

IV. Methods

A.   Criteria for Inclusion/Exclusion of Studies in the Review

The included populations, interventions, comparators, and outcomes are described above. Studies that did not report any outcomes of efficacy, effectiveness, safety/adverse events, or utilization patterns will be excluded.

B.   Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions.

A librarian will perform the initial literature search adopting a search strategy similar to that completed for the 2007 review, with a focus on updating original/full searches. As we conducted an update search on June 1, 2008, the search for aripiprazole, olanzapine, quetiapine, risperidone, and ziprasidone will cover from that date to present. The search for off-label use for the other listed new drugs will cover all years starting the date they were approved by FDA and brought on the market. Utilization data will be added to the literature searches, as will off-label use for new conditions (ADHD, eating disorders, insomnia, and anxiety). Reference mining will help identify additional articles to be included.

The following databases will be searched:

Databases
- DARE (Database of Abstracts of Reviews of Effects)
- Cochrane library of systematic reviews
- CENTRAL (Cochrane Central Register of Controlled Trials)
- PubMed (National Library of Medicine, includes MEDLINE)
- EMBASE (Biomedical and pharmacological bibliographic database)
- CINAHL (Cumulative Index to Nursing and Allied Health Literature)
- PsycINFO (Journals, books, reports, and dissertations on psychology and related fields)

Other sources
- Clinicaltrials.gov
- References of included studies
- References of relevant reviews
- Personal files from related topic projects

C.   Data Abstraction and Data Management

Data will be independently abstracted by a health services researcher and a psychiatrist trained in the critical assessment of evidence. The following data will be abstracted from included trials: trial name, setting, population characteristics (including sex, age, ethnicity, and diagnosis), eligibility and exclusion criteria, interventions (dose, frequency, and duration), any co-interventions, other allowed medication, comparisons, and results for each outcome. Intent-to-treat results will be recorded if available.

For efficacy/effectiveness outcomes, a statistician will extract data. The psychiatrist will choose which outcomes are most appropriate to pool. Poolability across studies is also important; the psychiatrist, the statistician, and the project team will jointly make the selection based on their professional knowledge and also considering the frequency of an outcome measure being reported by the trials. A minimum of three studies is required for meta-analysis. For each treatment or placebo arm within a trial, the sample size, mean outcome, and standard deviation will be extracted. If a study does not report a followup mean or if a followup mean cannot be calculated from the given data, the study will be excluded from analysis. For those trials that do not report a follow-up standard deviation, we will impute one by assigning the average standard deviation from other trials that report the standard deviation for the same outcome. If fewer than two trials are available with standard deviations, then we will impute the followup standard deviation by taking one-fourth the theoretical range of the scale.

D.   Assessment of Methodological Quality of Individual Studies

To assess internal validity, we will abstract data on the adequacy of the randomization method; the adequacy of allocation concealment; maintenance of blinding; similarity of compared groups at baseline and the author's explanation of the effect of any between-group differences in important confounders or prognostic characteristics; specification of eligibility criteria; maintenance of comparable groups (i.e., reporting of dropouts, attrition, crossover, adherence, and contamination); the overall proportion of subjects lost to follow-up and important differences between treatments; use of intent-to-treat analysis; post-randomization exclusions,

and source of funding. We will define loss to follow-up as the number of patients excluded from efficacy analyses, expressed as a proportion of the number of patients randomized.

To assess external validity, we will record the number screened, eligible, and enrolled; the use of run-in and washout periods or highly selective criteria; the use of standard care in the control group; and overall relevance. Funding source will be also abstracted.

To arrive at a quantitative measure, we will use the Jadad scale, which was developed for drug trials[2]. This method measures quality on a scale that ranges from 0-5, assigning points for randomization, blinding, and accounting for withdrawals and dropouts. (Across a broad array of meta-analyses, an evaluation found that trials scoring 0-2 report exaggerated results compared with trials scoring 3-5. The latter have been called "good" quality and the former called "poor" quality.)

E.   Data Synthesis

Our a priori analytic plan is to summarize the evidence for efficacy and effectiveness (versus placebo or versus conventional therapy) for each condition (dementia, depression, personality disorders, etc.) for each of the atypical antipsychotics, and across the class as a whole. The evidence of risks (adverse events) will be summarized within drug (each atypical antipsychotic separately) across condition. This strategy has ample support in the literature, with many examples of drugs that demonstrate similar efficacy across a class of drugs and are then distinguished on the basis of their adverse events profile.

For the efficacy and comparative effectiveness analyses, we will focus on controlled trials that report outcomes without a minimum trial length. Effect sizes will be calculated for each comparison. If all trials within a condition and subgroup use the same scale, then the effect size does not need to be standardized and a mean difference will be calculated. For subgroups where pooling is done across several scales, we will calculate an unbiased estimate using the Hedges' g effect size. Since most of the scales used as outcome measures in the pooled analyses are scored so that more severely symptomatic persons have higher scores, a negative effect size indicates that the atypical drug has a higher efficacy than does the comparison arm (active control or placebo arm).

For trials that are judged sufficiently clinically similar to warrant meta-analysis, we will estimate a pooled random-effects estimate of the overall mean difference in outcome measure. The individual trial mean differences are weighted by both within-study variation and between-study variation in this synthesis.

We will assess publication bias for each condition that is pooled. Tests will be conducted using the Begg adjusted rank correlation test and the Egger regression asymmetry test.

All meta-analyses will be conducted with Stata statistical software, version 8.2 (Stata Corp., College Station, Texas).

For groups of trials not judged sufficiently clinically similar to support meta-analysis, we will perform a narrative synthesis.

F.   Grading the Evidence for Each Key Question

A synopsis of the evidence will be provided for each of the key research questions. The overall quality of evidence for outcomes will be assessed using a method developed by the Grade Working Group,[3] which classified the grade of evidence across outcomes according to the following criteria:

- High: Further research is very unlikely to change our confidence on the estimate of effect.

- Moderate: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

- Low: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

- Very Low: Any estimate of effect is very uncertain.

The body of evidence will be evaluated by taking the risk of bias of individual studies, the consistency across studies, and, where available and appropriate, the directness and the precision of results.

V. References

1. Shekelle P, Maglione M, Bagley S, et al. Comparative Effectiveness of Off-label Uses of Atypical Antipsychotics. (Prepared by the Southern California/RAND Evidence-based Practice Center under Contract No. 290-02-0003.). Comparative Effectiveness Review #6.. Rockville, MD: Agency for Healthcare Research and Quality. May 2006. Available at: **www.effectivehealthcare.ahrq.gov/reports/final.cfm**
2. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; 17(1):1-12.
3. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004:328(7454):1490.

VI. Definition of Terms

1. Off-label: An off-label use is any use of the medication that is not specifically mentioned in the FDA approval labeling (including use for a different condition, different dose, or in a different population than the one that has been approved).

2. Atypical antipsychotic: second generation antipsychotic medication. They are used to treat psychiatric conditions and have varying mechanisms of action. They are grouped together due to their difference from the group of conventional, or typical, antipsychotic medications.

3. Efficacy: efficacy trials (explanatory trials) determine whether an intervention produces the expected result under ideal circumstances. Efficacy studies are frequently conducted in large tertiary care or referral settings.

4. Effectiveness: effectiveness trials (pragmatic trials) measure the degree of beneficial effect under "real world" clinical settings. Effectiveness studies are usually conducted in primary care settings and have less stringent eligibility criteria than efficacy studies.

VII. Summary of Protocol Amendments

The Key Questions were posted for public comment on the AHRQ Effective Health Care Program website. Some of the comments had already been addressed in our protocol (for instance, one person suggested including insomnia, which is already an included indication). We amended the protocol in the following areas in response to the comments, following a discussion with our technical expert panel:

- "Outcomes" section: we added general health outcomes (e.g., SF-36) and quality of life measures, if reported.

- "Adverse Events" section: adverse events will be analyzed for all controlled trials as well as observational studies with at least 1,000 cases. We will also summarize adverse events associated with on-label use of the atypical antipsychotic medications.

- "Timing" section: we will not limit the trial duration for efficacy for each disorder/condition – single dose or short term trials will be included.

NOTE: The following protocol elements are standard procedures for all protocols.

VIII. Review of Key Questions

For Comparative Effectiveness reviews (CERs) the key questions were posted for public comment and finalized after review of the comments. For other systematic reviews, key questions submitted by partners are reviewed and refined as needed by the EPC and the Technical Expert Panel (TEP) to assure that the questions are specific and explicit about what information is being reviewed.
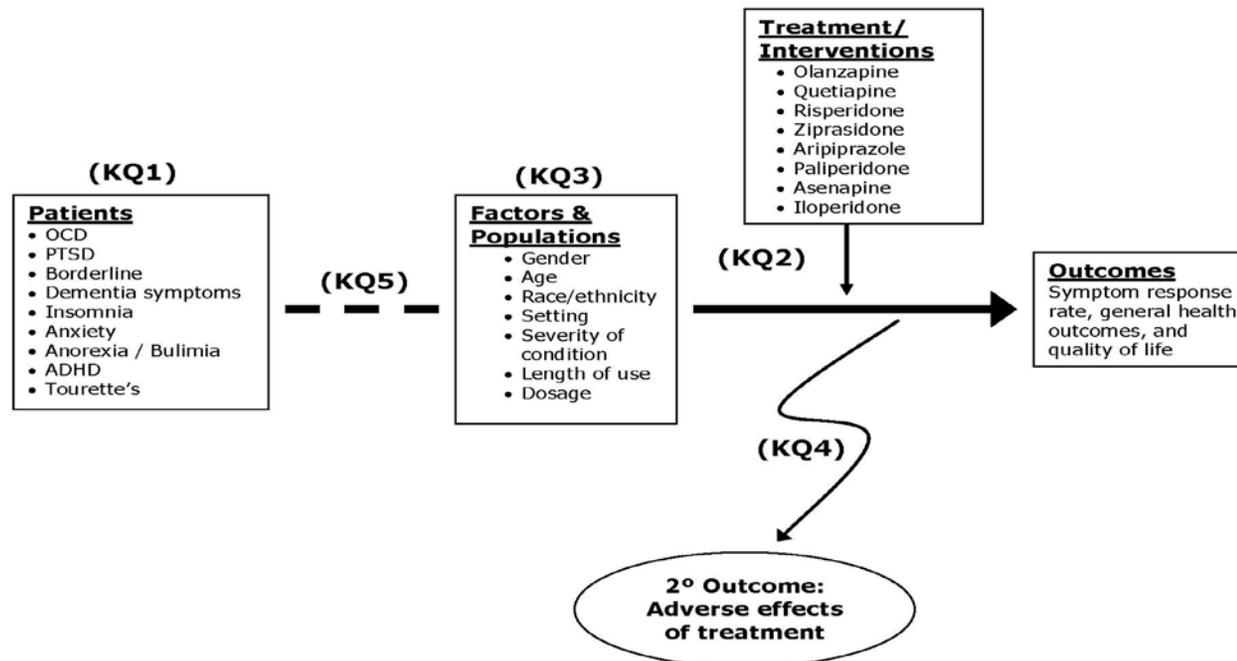
IX. Technical Expert Panel (TEP)

A TEP panel is selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design and/or methodological approaches do not necessarily represent the views of individual technical and content experts. The TEP provides information to the EPC to identify literature search strategies, review the draft report and recommend approaches to specific issues as requested by the EPC. The TEP does not do analysis of any kind nor contribute to the writing of the report.

X. Peer Review

Approximately five experts in the field will be asked to peer review the draft report and provide comments. The peer reviewer may represent stakeholder groups such as professional or advocacy organizations with knowledge of the topic. On some specific reports such as reports requested by the Office of Medical Applications of Research, National Institutes of Health there may be other rules that apply regarding participation in the peer review process. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical briefs, be published three months after the publication of the Evidence report.

It is our policy not to release the names of the Peer reviewers or TEP panel members until the report is published so that they can maintain their objectivity during the review process.

Figure 1. Analytic Framework for Comparative Effectiveness Review: Off-label Uses of Atypical Antipsychotics

Appendix A. Outcome Measures: Off-label Uses of Atypical Antipsychotics

Anxiety Disorder
  Beck Anxiety Inventory
  Hamilton Anxiety Rating Scale
  State-Trait Anxiety Inventory
  Generalized Anxiety Disorder (GAD) Questionnaire IV
  Daily Assessment of Symptoms - Anxiety (DAS-A)

Attention Deficit Hyperactivity Disorder (ADHD)
  Achenbach System for Empirically Based Assessment (ASEBA)
  ADD-H Comprehensive Teachers Rating Scale (ACTeRS)
  ADDES-Secondary Age
  ADHD Rating Scale-IV
  ADHD Symptom Checklist – 4 (ADHD-SC4)
  Attention-Deficit Disorders Evaluation Scale: Secondary-Age Student (ADDES-S)
  Beck Anxiety Inventory (BAI)
  Behavior Assessment System for Children-2 (BASC-2)
  Behavior Rating Inventory of Executive Functioning (child or adult version)
  Brown Attention-Deficit Disorders Scale
  Conners' Parent Rating Scale (age 3-17 years)
  Conners' Teacher Rating Scale (age 3 -17 years)
  Conners' Rating Scales-3 (Conners 3)
  Conners' Adult ADHD Rating Scales (CAARS)
  Conners' Comprehensive Behavior Rating Scales (Conners CBRS)
  Copeland Symptom Checklist for Adult Attention-Deficit Disorders (CSCAADD)
  Wender Utah Rating Scale (WURS) and Parent's Rating Scale (PRS)

Dementia
Dementia-agitation
  Agitation-Calmness Evaluation Scale - ACES
  Behavioral Pathology in Alzheimer's Disease Rating Scale - BEHAVE-AD (subscale: aggressiveness)
  Cohen-Mansfield Agitation Inventory - CMAI
  Neuropsychiatric Inventory, Nursing Home - NPI-NH (subscale: agitation)
  Neuropsychiatric Inventory - NPI (subscale: agitation)
  Neurobehavioral Scale
  Positive and Negative Symptom Scale - PANSS (subscale: excitement)

Dementia-cognition
  Mini Mental Status Exam - MMSE
  Alzheimer's Disease Assessment Scale - ADAS (cognition scale)

Dementia-global
Neuropsychiatric Inventory, Nursing Home - NPI-NH (total)
Neuropsychiatric Inventory - NPI (total)
Clinician's Interview-Based Impression of Change - CIBIC
Empirical Behavioral Pathology in Alzheimer's Disease Rating Scale - E-BEHAVE-AD (total)
Behavioral Pathology in Alzheimer's Disease Rating Scale - BEHAVE-AD (total)

Dementia-improvement
Clinical Global Impression Scale - CGI:I (improvement subscale)

Dementia-psychosis
Neuropsychiatric Inventory, Nursing Home - NPI-NH (subscale: psychosis)
Positive and Negative Symptom Scale - PANSS (subscale: psychosis)
Behavioral Pathology in Alzheimer's Disease Rating Scale BEHAVE-AD (sum of paranoid and delusional ideation and hallucinations items)
Brief Psychiatric Rating Scale - BPRS (subscale: psychosis) - it is the sum of unusual thought content, paranoia(or suspiciousness), hallucinations (or hallucinatory behavior), disorganized thinking (or conceptual disorganization)

Dementia-severity
Clinical Global Impression Scale - CGI:S (severity subscale)

Depression
Hamilton Depression Scale - HAM_D (HDRS)
Montgomery - Asberg Depression Rating Scale - MADRS
Bech-Rafaelson Melancholia Scale – BRMES
Beck Depression Inventory - BDI
Depression cluster - PDC
Center for Epidemiologic Studies Depression Scale - CES-D
Brief Symptom Inventory – BSI

Eating Disorders- Anorexia nervosa and Bulimia
Eating Disorders Examination Questionnaire Version (EDEQ)
Questionnaire on Eating and Weight Patterns- Revised (QEWP-R)
Structured Interview for Anorexia and Bulimia (SIAB)

Insomnia
Insomnia Severity Index (ISI)
Insomnia Symptom Questionnaire (ISQ)
Sleep Disorders Questionnaire
Pittsburgh Sleep Quality Index
SleepMed Insomnia Index
Medical Outcomes Study (MOS) Sleep Problem Index (SPI)

Obsessive Compulsive Disorder (OCD)
   Yale - Brown Obsessive Compulsive Scale – YBOCS

Post Traumatic Stress Disorder (PTSD)
   Clinician Administered PTSD Scale – CAPS

Tourette's Syndrome
   Tic Symptom Self Report – TSSR
   Yale Global Tic Severity Scale - YGTSS