**Supplement 3:** Study outcomes

The following outcomes scales were used:
1. Global state:

1.1. Clinical Global Impression Scale - CGI Scale ([Guy et al. 1976a](#REF-Guy-1976a))
Trials used this to assess both severity of illness and clinical improvement. The CGI is a seven-point scoring system usually used with low scores showing decreased severity and/or overall improvement.

1.2. Children’s Global Assessment Scale ([Shaffer et al. 1983](#REF-Shaffer-1983))
The Children’s Global Assessment Scale (CGAS) is used to provide a global measure of level of functioning in children and adolescents. The measure provides a single global rating only, on scale of 0-100. In making their rating, the clinician makes use of the glossary details to determine the meaning of the points on the scale.

2. Mental State/Behaviour:

2.1 Positive and Negative Syndrome Scale - PANSS ([Kay et al. 1987](#REF-Kay-1987))
This schizophrenia scale has 30 items, each of which can be defined on a seven-point scoring system varying from 1 - absent to 7 - extreme. This scale can be divided into three sub-scales for measuring the severity of general psychopathology, positive symptoms (PANSS-P), and negative symptoms (PANSS-N). A low score indicates lesser severity.

2.2. Young Mania Rating Scale—YMRS ([Young et al. 1978](#REF-Young-1978))

The YMRS includes 11 items and is widely used to assess mental state in adolescents with bipolar disorder. It has good sensitivity and specificity. Clinical studies have demonstrated effectiveness of the parent version of the scale.

2.3. Aberrant Behaviour Checklist - ABC ([Aman et al. 1985](#REF-Aman-1985); Brown et al. 2002)

Consists of 58 items subdivided amongst 5 scales: irritability, lethargy and social withdrawal, stereotypic behaviour, hyperactivity/noncompliance, and inappropriate speech. A score for each item ranged from 0 indicating "no problem" to 3 indicating "severe problem".

2.4. Antisocial Behaviour Scale - ABS ([Brown et al. 1996](#REF-Brown-1996))

Comprising the Proactive Aggression and Reactive Aggression subscales.

2.5. Nisonger Child Behavior Rating Form - NCBF ([Aman et al. 1996](#REF-Aman-1996), [Tasse et al. 1996](#REF-Tasse-1996), [Aman et al. 2008](#REF-Aman-2008))

The assessment contains 76 items 10 Positive/Social items and 66 Problem Behaviour items). The NCBRF is made up of three sections: 1. Where raters can identify unusual circumstances that may have affected the youth's behaviour; 2. where positive behaviours are rated, and 3. a listing of problem behaviours.

2.6. Verbal Learning Test-Children’s Version - VMLT-C ([Delis et al. 1987](#REF-Delis-1987))

The MVLT-C involves crepitated presentation word lists with requests of freely recall and discriminatively recognize presented words. This scale consists of two parts: (1) short-delay free recall and (2) long-delay free recall and recognition. A list of 15 words was presented (oral or by pictures). For the short- and long- delay free recall studies, the subjects are asked to enumerate the words they recalled.

2.8 Continuous Performance Test - CPT

CPT is an attention test, assessing vigilance thought computer testing with both an easy test and a hard test. In the easy test, stimuli were a witch and a princess appearing on a blank screen. Stimuli were presented for 2 seconds at regular intervals and were presented randomly at approximately 1:5 ration, with no two targets presented consecutively. Subjects were required to respond whenever the princess appeared. In the hard test, stimuli were identical to the easy condition, but the interval varied (from 1 to 4 seconds) and stimulus duration was 3/10 of a second. Targets were also presented randomly at an approximately 1:5 ratio, with no two targets presented consecutively.

2.7. Attention Deficit Hyperactivity Rating Scale IV – ADHD-RS-IV ([Reid et al. 1998](#REF-Reid-1998))

The ADHD Rating Scale-IV is completed independently by the parent and scored by a clinician. The scale consists of two subscales: inattention (9 items) and hyperactivity-impulsivity (9 items).

2.8. General Behaviour Inventory - GBI ([Findling et al. 2002](#REF-Findling-2002);[Youngstrom et al. 2001](#REF-Youngstrom-2001))

GBI is a two 10-item subscales, which assess symptoms of mania/hypomania and depression completed by both parents/guardians and subjects.

2.9. Children’s Depression Rating Scale – Revised, CDRS-R

CDRS-R is an abbreviated version of the GBI. Negative changes signify improvement.

3. Quality of life

3.1. Paediatric Quality of Life Enjoyment and Satisfaction Questionnaire— PQ‐LES‐Q ([Endicott et al. 1993](#REF-Endicott-1993))
The PQ‐LES‐Q is a 15‐item scale with a total score ranging from 14 to 70. The total score is calculated by adding scores for the first 14 items. Each item on the scale is scored by the young person from 1 to 5. Higher scores indicate greater enjoyment and satisfaction.

3.2. Paediatric Quality of Life Inventory - PedQL ([Varni et al. 2001](#REF-Varni-2001))

PedQL is a health related quality-of-life instrument developed and validated for use with children and adolescents, and items on the PedQL are reverse scored and linearly transformed to a 0 - 100 scale so that higher scores indicate better health-related quality of life.

3.3. Caregiver Strain Questionnaire - ([Brannan et al. 1997](#REF-Brannan-1997))

A 21-item self-reported instrument assessing the impact that caring for children and adolescents with serious emotional, mental, and behavioural disturbances has on families.

4. Adverse effects:

4.1. Simpson Angus Scale - SAS ([Simpson et al. 1970](#REF-Simpson-1970))
SAS is a 10-item rating scale that has been used widely for assessment of neuroleptic medication induced movement disorders in research settings. It consists of one item measuring gait (hypokinesia), six items measuring rigidity and three items measuring glabella tap, tremor and salivation, respectively.

4.2. Abnormal Involuntary Movements Scale ([Guy et al. 1976b](#REF-Guy-1976b))
This is a 12 item scale each item scored on a 0 to 4. It is used to monitor antipsychotic medication induced movement disorders. It has separate items for oral and facial movements, movement of extremities, trunk and global judgements.

4.3. Barnes Akathisia Rating Scale - BARS ([Barnes et al. 1989](#REF-Barnes-1989)).

This scale comprises items rating the observable, restless movements that characterise akathisia, a subjective awareness of restlessness and any distress associated with the condition. Items are rated from 0 - normal to 3 - severe. In addition, an item for rating global severity (from 0 - absent to 5 - severe) is included. A low score indicates low levels of akathisia.

4.4. Neurological Rating Scale (NRS) for Extrapyramidal Side Effects.