ORIGINAL RESEARCH ARTICLE

Comparative Efficacy of Guanfacine Extended Release Versus Atomoxetine for the Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents: Applying Matching-Adjusted Indirect Comparison Methodology

Vanja Sikirica · Robert L. Findling · James Signorovitch · M. Haim Erder · Ryan Dammerman · Paul Hodgkins · Mei Lu · Jipan Xie · Eric Q. Wu

Abstract

Background No head-to-head clinical trials have been published comparing guanfacine extended release (GXR) and atomoxetine (ATX): two nonstimulants approved for the treatment of attention-deficit/hyperactivity disorder (ADHD). However, other study designs or methods could be used to indirectly compare these two medications. Matching-adjusted indirect comparison (MAIC) is a recent methodology that utilizes individual patient data (IPD) from clinical trials for one treatment and published aggregate data from another treatment to estimate the relative efficacy of both, providing rapid, reliable comparative efficacy results.

Objective The aim of this study was to compare the efficacy of GXR and ATX for the treatment of ADHD using MAIC.

Study Design A systematic literature search was conducted to identify ATX and GXR trials published through December 2012. Studies were selected for MAIC analyses on the basis of having comparable trial characteristics and study designs. Summary data from selected ATX trials and IPD from selected GXR trials were used. MAIC methodology ensured comparable populations: target doses for the 'base case' comparison were selected on the basis of maximum effective dosage ranges from the US FDA-approved product labels (GXR 0.09–0.12 mg/kg/day, ATX 1.2 mg/kg/day for children and adolescents weighing ≤ 70 kg). Individuals from GXR trials were selected if they matched inclusion/exclusion criteria from selected ATX trials; selected GXR IPD were then re-weighted to match the published ATX trial mean baseline characteristics and placebo outcomes. Sensitivity analyses were conducted, examining different dosage ranges and repeating the analysis in a larger number of trials, allowing for larger and more heterogeneous trial populations.

Main Outcome Measure The primary outcome measure was change in ADHD Rating Scale IV (ADHD-RS-IV) total score.

Results Using MAIC in the base case comparison, significantly greater reductions in mean (standard error; SE) ADHD-RS-IV total scores from baseline to end of study were observed in patients treated with GXR relative to ATX [-7.0 (2.2); $p < 0.01$]. Significantly greater reductions for GXR over ATX were also demonstrated for hyperactivity/impulsivity [-3.8 (1.2); $p < 0.01$] and inattention [-3.2 (1.3); $p < 0.05$] subscales of the ADHD-RS-IV. Similar results were observed in MAIC sensitivity analyses evaluating other dosage ranges and using more heterogeneous trial populations (e.g., larger randomized sample, broader subject weight range, additional trials). Mean (SE) decreases in ADHD-RS-IV total scores were greater for GXR relative to ATX when including IPD for those administered GXR at lower than target dosage (0.075–0.090 mg/kg/day) compared with ATX at target dosage (1.2 mg/kg/day), with a relative improvement of

V. Sikirica (✉) · M. H. Erder · R. Dammerman · P. Hodgkins
Global Health Economics and Outcomes Research, Shire Development LLC, 725 Chesterbrook Boulevard, Wayne, PA 19087, USA
e-mail: vsikirica@shire.com

R. L. Findling
The Johns Hopkins University and the Kennedy Krieger Institute, Baltimore, MD, USA

J. Signorovitch · M. Lu · J. Xie · E. Q. Wu
Analysis Group, Inc., Boston, MA, USA

−6.0 (2.7) ($p < 0.05$). Reductions in ADHD-RS-IV total scores were also greater for GXR in another MAIC examining GXR at target dosage (0.09–0.12 mg/kg/day) and a broader range of ATX dosages (including three additional trials evaluating ATX ≥1.2 mg/kg/day); relative improvement for GXR versus ATX administered at target dosage or higher was −7.6 (1.4) ($p < 0.01$).

Conclusion After adjusting for difference in baseline trial characteristics using MAIC, GXR appears to be more efficacious than ATX for the treatment of ADHD. Results were consistent in a variety of dosage range comparisons and within increasingly heterogeneous trial populations.

1 Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a common psychiatric disorder in childhood, affecting an estimated 3–7 % of school-age children in the USA [1]. While stimulants have proven to be safe and effective first-line therapy for most patients with ADHD [2], some patients may have a contraindication or may be intolerant, unresponsive, or only partially responsive to stimulant treatment. In addition, families or patients may prefer not to use stimulant therapy [2, 3]. For these individuals, nonstimulant treatments may be considered as treatment alternatives [2]. Atomoxetine (ATX; Strattera®, Eli Lilly and Company), the first nonstimulant approved in the USA for the treatment of ADHD, is a selective norepinephrine reuptake inhibitor indicated for use as monotherapy in children (aged 6 years and older) and adults [4]. ATX has also been approved for use in Canada, Mexico, and several other countries throughout the world, including in Europe, Asia, Australia, Africa, and Latin America. An extended-release formulation of the selective α_{2A}-adrenergic receptor agonist guanfacine (GXR; Intuniv®, Shire Pharmaceuticals Inc.) became the second nonstimulant approved by the Food and Drug Administration (FDA) for the treatment of ADHD. GXR is indicated for use both as monotherapy and as an adjunct to stimulant medications in children and adolescents aged 6–17 years [5]. Recently, a third non-stimulant, clonidine extended release (Kapvay™; Shionogi Inc.), has also been approved by the FDA for monotherapy and adjunctive therapy in the treatment of ADHD. At present, GXR and clonidine extended release formulations are only licensed for use in ADHD in the USA.

The objective of the current study was to compare the relative efficacy of GXR and ATX as nonstimulant monotherapies for the treatment of ADHD in children and adolescents. While a head-to-head randomized controlled study is generally considered the most robust method for comparing two competing interventions, this type of study data is rarely available given the time, expense, and dependency on manufacturers' incentive to conduct such a study [6, 7]. Indeed, no head-to-head randomized controlled study has been conducted to date comparing GXR and ATX. However, there is an ever-increasing demand from US healthcare payers for comparative evidence, or comparative effectiveness research (CER), for competing treatment interventions to inform treatment decisions and shape best practices [8]. In the absence of direct comparisons from head-to-head randomized controlled studies, other methods for CER can be utilized, such as an indirect treatment comparison (ITC), mixed treatment comparison, or meta-regression [7]. While CER typically refers to comparative 'effectiveness' based on real world data, this study examines comparative 'efficacy' as measured in a controlled environment. In the absence of comparative effectiveness or efficacy data for GXR and ATX, this comparative efficacy analysis was conducted as part of CER. For the purposes of this publication, we will use the term CER to include 'efficacy.'

Indirect treatment comparisons are made using available data from separate clinical trials of two different treatments in a number of ways. Traditional ITCs have well-recognized limitations. Comparisons based solely on published summary data (e.g., means), such as a naïve ITC, are subject to potential biases when only a small number of trials are available, and the summary data being compared may not be anchored to a common comparator or adjusted for placebo effects. Other ITC methods that may account for baseline differences may be subject to inconsistent conclusions depending upon the measure of effect used (e.g., odds ratio, relative risk) [9]. Meta-regression can be used to adjust for trial-level factors that differ between trials (e.g., mean age), but can be unreliable for small numbers of trials and may be subject to ecological bias (i.e., potential false inferences arising by using group rather than individual patient level data) [9, 10].

When only a few studies are available for an ITC, many of the limitations of ITC can be overcome if individual patient data (IPD) are available for just one of the treatments being compared. A recent methodology in CER, matching-adjusted indirect comparison (MAIC) [9, 11], leverages IPD from one treatment arm to model the observed baseline differences between two treatments. IPD from clinical trials for a given treatment can be re-weighted to exactly match the summary baseline characteristics and placebo outcomes from trials of another treatment for which only published data are available [9, 11]. After adjusting for the imbalance in baseline characteristics, study outcomes can be meaningfully evaluated across trials. These results can either stand on their own or be used to inform efficient prospective trial designs by informing power calculations. In the current analysis, available IPD from GXR studies were re-weighted to match summary

published data from ATX studies, allowing for a more rapid, comparable, and likely less biased comparison of efficacy between GXR and ATX.

2 Methods

2.1 Trial Selection and Outcomes

A systematic literature search was conducted to include controlled clinical trials enrolling children and adolescents (age ≤18 years) with ADHD treated with GXR or ATX in Europe, North America, or Australia. The search was conducted in MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE, PsycINFO, and the Cochrane Central Register of Controlled Trials. In addition, online documents (clinicaltrials.gov), published systematic reviews, meta-analyses, and post hoc analyses were reviewed for identification of additional studies. Primary publications written in English and published through December 2012 were included [keywords or Medical Subject Headings (MeSH) keywords: (ADHD or attention-deficit hyperactivity disorder) AND

(guanfacine extended release or atomoxetine)]. Studies for data comparison were selected if they had comparable study characteristics based upon consistent study design (multicenter, double-blind, randomized, placebo-controlled trials), length of treatment, and commonly reported efficacy [ADHD Rating Scale IV (ADHD-RS-IV) scores]. Trials were excluded on the basis of the following criteria: open-label studies; studies focusing on ADHD subgroups (e.g., requiring certain comorbidities; requiring the study drug to be used as a specific line of treatment; or requiring symptom severity to be measured by scales other than the ADHD-RS-IV [e.g., Clinical Global Impressions-Severity (CGI-S) ≥4 at baseline]); studies with combination/adjunctive therapy with GXR or ATX (including combination with behavioral or educational programs); ADHD symptoms reported by teachers; age younger than 6 years; no outcome of interest (e.g., ADHD-RS-IV score change from baseline); no baseline ADHD-RS-IV values; or no outcome of interest for individual study arms. A total of six trials (two GXR [12, 13] and four ATX trials [14–17]) met inclusion criteria (Fig. 1a, b). Study design characteristics of the selected trials are available in the Electronic Supplementary Material (Appendix 1).

Fig. 1 a Guanfacine extended-release trial selection. b Atomoxetine trial selection. *ADHD* attention-deficit/hyperactivity disorder, *ADHD-RS-IV* ADHD Rating Scale IV, *MAIC* matching-adjusted indirect comparison, *RCT* randomized controlled trial
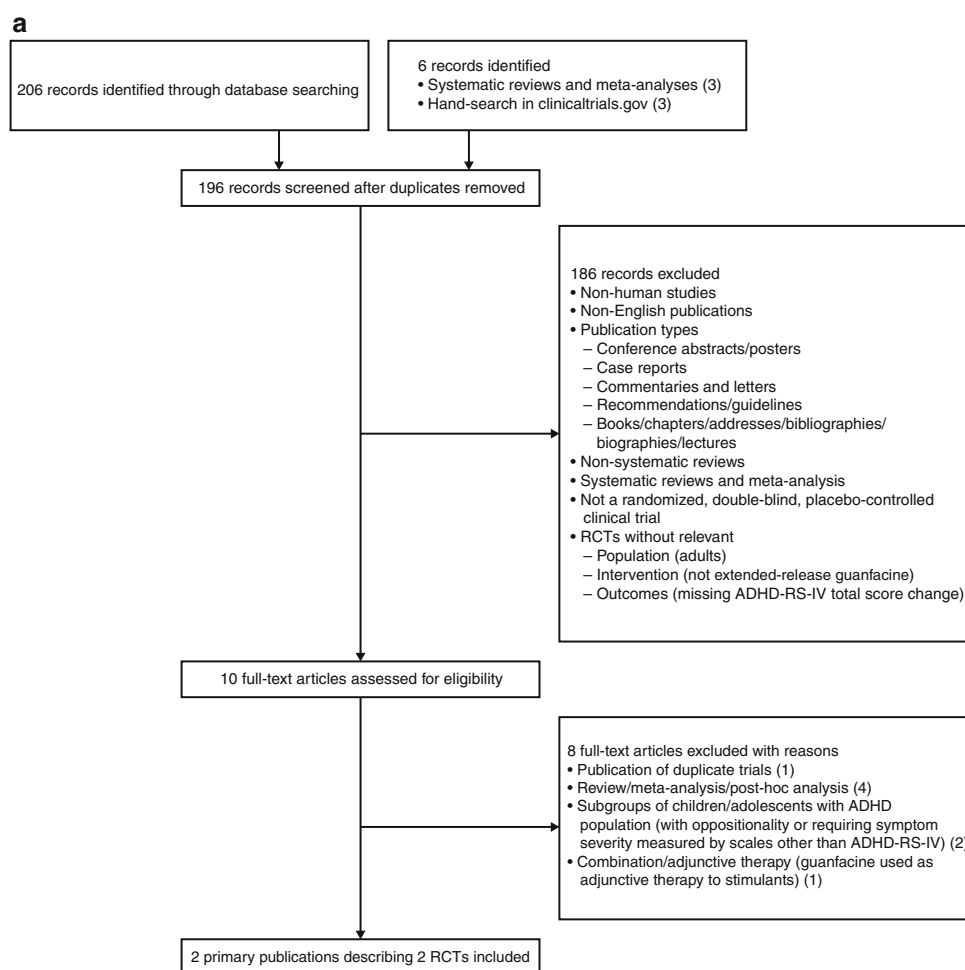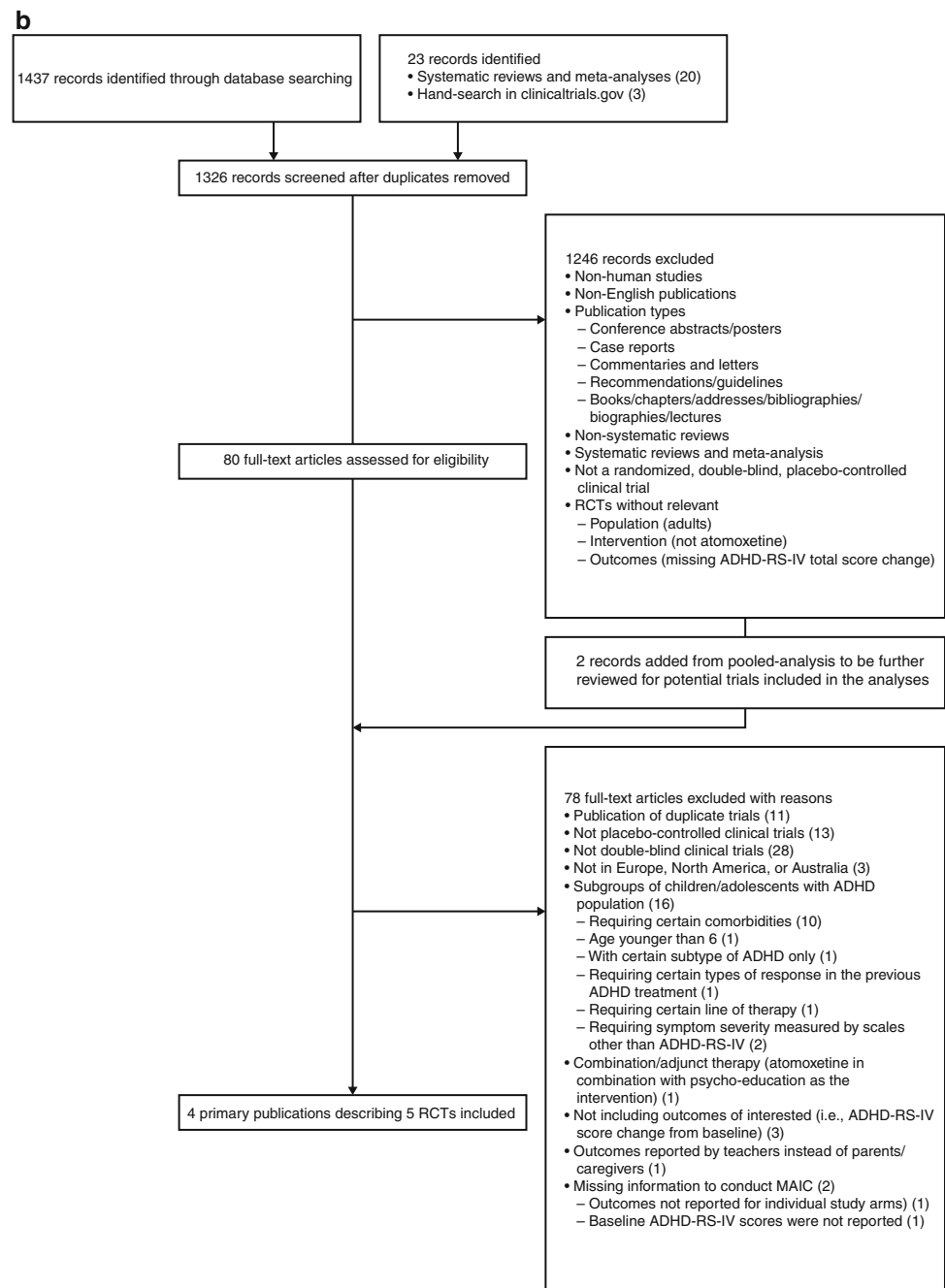
Fig. 1 continued

b

1437 records identified through database searching

23 records identified
• Systematic reviews and meta-analyses (20)
• Hand-search in clinicaltrials.gov (3)

1326 records screened after duplicates removed

1246 records excluded
• Non-human studies
• Non-English publications
• Publication types
  – Conference abstracts/posters
  – Case reports
  – Commentaries and letters
  – Recommendations/guidelines
  – Books/chapters/addresses/bibliographies/
    biographies/lectures
• Non-systematic reviews
• Systematic reviews and meta-analysis
• Not a randomized, double-blind, placebo-controlled
  clinical trial
• RCTs without relevant
  – Population (adults)
  – Intervention (not atomoxetine)
  – Outcomes (missing ADHD-RS-IV total score change)

80 full-text articles assessed for eligibility

2 records added from pooled-analysis to be further
reviewed for potential trials included in the analyses

78 full-text articles excluded with reasons
• Publication of duplicate trials (11)
• Not placebo-controlled clinical trials (13)
• Not double-blind clinical trials (28)
• Not in Europe, North America, or Australia (3)
• Subgroups of children/adolescents with ADHD
  population (16)
  – Requiring certain comorbidities (10)
  – Age younger than 6 (1)
  – With certain subtype of ADHD only (1)
  – Requiring certain types of response in the previous
    ADHD treatment (1)
  – Requiring certain line of therapy (1)
  – Requiring symptom severity measured by scales
    other than ADHD-RS-IV (2)
• Combination/adjunct therapy (atomoxetine in
  combination with psycho-education as the
  intervention) (1)
• Not including outcomes of interested (i.e., ADHD-RS-IV
  score change from baseline) (3)
• Outcomes reported by teachers instead of parents/
  caregivers (1)
• Missing information to conduct MAIC (2)
  – Outcomes not reported for individual study arms) (1)
  – Baseline ADHD-RS-IV scores were not reported (1)

4 primary publications describing 5 RCTs included

The efficacy outcome analyzed was the mean change from baseline in the ADHD-RS-IV score at the final on-treatment assessment prior to down-titration (i.e., drug tapering). The ADHD-RS-IV total score was evaluated as the primary endpoint; secondary endpoints included the ADHD-RS-IV hyperactivity/impulsivity and inattention subscale scores.

2.2 Dose Selection

For the purposes of a primary base case comparison of GXR and ATX, target doses used for comparison were selected on the basis of the maximum recommended effective dosages from their respective FDA-approved labels [4, 5]. For ATX, a body weight-based daily dose of 1.2 mg/kg is considered the target for children up to 70 kg [4]—thus, 1.2 mg/kg was chosen for the base case analysis; although a maximum recommended daily dose is 1.4 mg/kg, no additional benefit was demonstrated at doses >1.2 mg/kg/day for children and adolescents weighing up to 70 kg. Among children and adolescents weighing more than 70 kg, the maximum recommended daily dose should not exceed 100 mg [4]. The FDA-approved dose range of

GXR is 1–4 mg once daily [5]. In order to match the published body weight-based ATX trial data, randomized fixed-dose data from individuals in the selected GXR trials were converted to body weight-based doses (also reported in the package insert and in the published study results) [5, 12, 13]. In GXR monotherapy clinical trials, clinically relevant improvements (e.g., in ADHD-RS-IV scores) were observed beginning in the 0.05–0.08 mg/kg/day dose range, with additional benefit observed at doses up to 0.12 mg/kg/day [5]. As the number of patients who received exactly GXR 0.12 mg/kg/day based on their weight in the selected GXR studies was too small to enable adequate comparisons, a dosage range for the base case analysis was expanded to 0.09–0.12 mg/kg/day range.

In both GXR monotherapy studies meeting the initial trial selection criteria [12, 13], the target dose range of GXR 0.09–0.12 mg/kg/day was used. Of the four ATX studies meeting the initial trial selection, only one study [15] included a treatment arm at the target dose of 1.2 mg/kg/day, and therefore only this study met the additional inclusion criteria of having a treatment dosage not exceeding the maximum dose at which efficacy was observed. Thus, the base case analysis was drawn from three trials: two GXR trials (GXR administered at 0.09–0.12 mg/kg/day) [12, 13] and one ATX trial (ATX administered at 1.2 mg/kg/day) [15].

2.3 Patient Selection

While only summary published data were available from the ATX trials, IPD were available from both GXR trials. In order to be included in the base case analyses, individual patients in the GXR trials [12, 13] were required to meet the published inclusion/exclusion criteria from the Michelson et al. 2001 ATX trial [15]: patients had to have a baseline symptom severity score of $\geq$1.5 standard deviations (SDs) above age and gender normative values on the ADHD-RS-IV total score or the hyperactivity/impulsivity or inattention subscale scores (Fig. 2). Patients from the GXR trials were further selected into the base case analysis cohorts on the basis of their expected body weight-based dosing. In the original GXR study designs, patients were randomized into fixed dose cohorts of GXR (1–4 mg) or placebo [12, 13]. For patients randomized to GXR, expected body weight-based dosing was calculated on the basis of patients' assigned GXR dosage and baseline body weight. Similarly, the expected body weight-based dosing was evaluated for patients randomized to placebo to determine the probability of allocation into each dosing arm (1–4 mg/day); patients were not randomized to any 'strength' of placebo, but in order to match them to the active arm with different treatment strengths, the probability to receive one of the active treatments that the placebo patients would have potentially been assigned to was
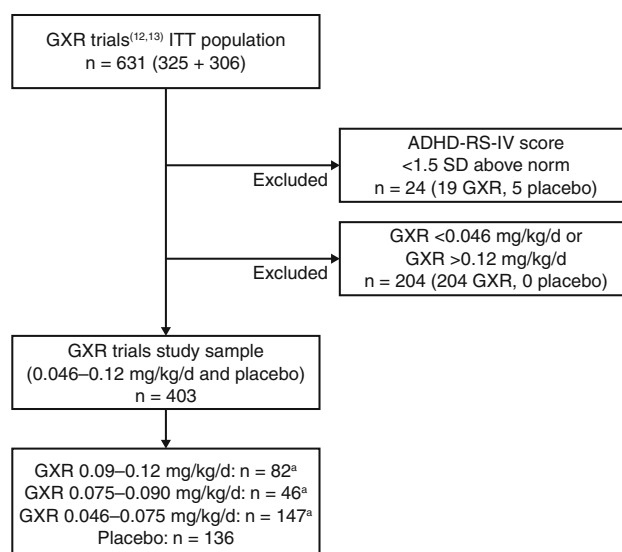
Fig. 2 Patient selection: matching inclusion/exclusion criteria. [a]Individual patients could have been included in more than one dose cohort where the dose ranges overlap. *ADHD* attention-deficit/hyperactivity disorder, *ADHD-RS-IV* ADHD Rating Scale IV, *GXR* guanfacine extended release, *ITT* intention-to-treat, *SD* standard deviation

calculated. Patients randomized to placebo were excluded if, on the basis of their body weight, they would not fit into one of the body weight-adjusted dosing arms. Placebo groups across GXR trials were then collapsed into one group. For the base case analysis, this patient selection process allowed the comparative patient groups to be as homogenous as possible within the specified dosage ranges.

2.4 Statistical Analysis

Comparative efficacy was analyzed before and after matching with IPD. In the unadjusted comparison (i.e., before matching with IPD), selected IPD that met all study selection inclusion and exclusion criteria were pooled from both GXR trials and compared with the summary results from the ATX trial [15] included in the base case analysis. Baseline characteristics and efficacy were compared for the GXR- and ATX-treated patients: continuous baseline variables and efficacy results were compared using Student's *t* tests, and categorical baseline variables were compared using Chi-squared tests.

For the MAIC analyses (after matching with IPD), patients across GXR and ATX trials were matched for age, percentage of female patients, baseline ADHD-RS-IV hyperactivity/impulsivity and inattention subscale scores, percentage of inattentive and hyperactive/impulsive subtypes, and corresponding placebo response. Individual patients in the GXR trials were assigned weightings such that their baseline characteristics and average placebo outcomes (either ADHD-RS-IV total score or subscale scores) from GXR trials exactly matched those reported for the

ATX trial. Weights were modeled as a linear combination of all reported baseline characteristics (age, percentage of female patients, race, weight, height, ADHD-RS-IV baseline inattentive and hyperactivity/impulsivity subscale scores, ADHD subtype, and disease duration), as well as pairwise interactions, and quadratic and cubic terms based on these measures in the GXR trials. The weights were then estimated using linear programming to minimize the sum of the squared weights (i.e., their deviation from uniform weighting) under the constraints that all weights were non-negative and that all weighted mean baseline characteristics and placebo arm outcomes were exactly balanced between the GXR and ATX trials. After matching, efficacy outcomes for the re-weighted GXR-treated patients were compared with those from the ATX trial in comparable trial populations. To assess statistical significance, a bootstrap procedure was applied [18]. Patients in the GXR trials were randomly sampled with replacement to generate 1,000 bootstrap replicates. Estimation of the weights and weighted-comparisons between GXR and ATX were repeated for each replicate. Statistical testing was then based on two-sample t tests with unequal variance, incorporating the bootstrap standard errors (SEs) for GXR outcomes and reported SEs for ATX outcomes. Two-sided statistical significance was assessed at the $\alpha = 0.05$ level.

2.5 Sensitivity Analysis

In order to test the robustness of the base case results, sensitivity analyses were conducted. From the two GXR studies used in the base case comparison, different cohorts of patients receiving GXR at lower doses than the target dose (0.075–0.090 and 0.046–0.075 mg/kg/day) were selected and were compared with the same base case target dose of ATX (1.2 mg/kg/day) in the three-trial sensitivity analysis [12, 13, 15].

To determine if the base case results were generalizable to a broader heterogeneous trial population, a second MAIC analysis was conducted where ATX doses ≥ 1.2 mg/kg/day were compared against three GXR dose cohorts (0.09–0.12, 0.075–0.090, and 0.046–0.075 mg/kg/day). For this analysis, three additional ATX trials met the selection criteria in addition to the three trials included in the base case ('six-trial' analysis). The three additional ATX trials included in the six-trial sensitivity analysis did not report results with a fixed 1.2 mg/kg/day ATX dosage, but rather varied dosages of ATX (titrated up to 2 mg/kg/day) [14, 16, 17].

Statistical evaluations for the sensitivity analyses were conducted in a manner similar to that described for the base case comparison.

3 Results

Baseline pooled patient characteristics for the base case and three-trial sensitivity analyses are shown in Table 1. Before matching, mean age between patients in the GXR trials and ATX trials was statistically significantly different. After matching, all means and SDs of continuous

Table 1 Baseline characteristics before and after matching in the three-trial base case and sensitivity analyses

Baseline characteristics	GXR base case analysis 0.09–0.12 mg/kg/day (n = 82)		GXR sensitivity analyses 0.046–0.075 mg/kg/day (n = 147)		0.075–0.090 mg/kg/day (n = 46)		ATX[a] (n = 84)
	Pre-match	Post-match	Pre-match	Post-match	Pre-match	Post-match	
Age, mean (SD), years	9.0 (1.8)*	11.5 (2.4)	10.9 (2.5)	11.5 (2.4)	10.1 (2.0)*	11.5 (2.4)	11.5 (2.4)
Females, % (SD)	24.4 (42.9)	28.6 (45.2)	26.5 (44.1)	28.6 (45.2)	19.6 (39.7)	28.6 (45.2)	28.6 (45.2)
ADHD subtype, % (SD)							
Inattentive	18.3 (38.7)	27.4 (44.6)	26.5 (44.1)	27.4 (44.6)	23.9 (42.7)	27.4 (44.6)	27.4 (44.6)
Hyperactive-impulsive	2.4 (15.4)	1.2 (10.8)	3.4 (18.1)	1.2 (10.8)	2.2 (14.6)	1.2 (10.8)	1.2 (10.8)
Combined	79.3 (40.5)	71.4 (45.2)	70.1 (45.8)	71.4 (45.2)	73.9 (43.9)	71.4 (45.2)	71.4 (45.2)
Baseline ADHD-RS-IV outcomes, mean (SD)							
Total score	40.1 (7.7)	39.2 (9.2)	38.9 (8.7)	39.2 (9.2)	39.5 (8.6)	39.2 (9.2)	39.2 (9.2)
Inattention subscale score	21.9 (3.8)	22.2 (4.0)	22.1 (4.0)	22.2 (4.0)	21.7 (4.4)	22.2 (4.0)	22.2 (4.0)
Hyperactivity/impulsivity subscale score	18.2 (5.7)	16.9 (7.1)	16.8 (6.7)	16.9 (7.1)	17.8 (6.1)	16.9 (7.1)	16.9 (7.1)

ADHD attention-deficit/hyperactivity disorder, *ADHD-RS-IV* ADHD Rating Scale IV, *ATX* atomoxetine, *GXR* guanfacine extended-release, *SD* standard deviation

* $p < 0.05$ compared with ATX

[a] GXR group was matched with the ATX group; therefore, the baseline characteristics of the ATX group remained the same before and after matching

Table 2 Unadjusted comparison of change from baseline at final on-treatment assessment in ADHD-RS-IV scores: three-trial base case and dosage range sensitivity analyses

Change in ADHD-RS-IV score, mean (SE)	GXR trials	ATX trial (1.2 mg/kg/day)	Comparison (GXR vs. ATX)
Placebo, n	136	83	
Total score	−10.6 (1.1)	−5.8 (1.2)	−4.8 (1.6)**
Hyperactivity/impulsivity subscale score	−5.0 (0.6)	−3.2 (0.6)	−1.8 (0.8)*
Inattention subscale score	−5.6 (0.6)	−2.5 (0.7)	−3.1 (1.0)**
GXR 0.09–0.12 mg/kg/day, n	82	84	
Total score	−22.8 (1.3)	−13.6 (1.5)	−9.2 (2.0)**
Hyperactivity/impulsivity subscale score	−11.5 (0.7)	−6.6 (0.8)	−4.9 (1.0)**
Inattention subscale score	−11.3 (0.7)	−7.0 (0.9)	−4.3 (1.1)**
GXR 0.075–0.090 mg/kg/day, n	46	84	
Total score	−18.5 (2.1)	−13.6 (1.5)	−4.9 (2.6)
Hyperactivity/impulsivity subscale score	−9.4 (1.1)	−6.6 (0.8)	−2.8 (1.3)*
Inattention subscale score	−9.1 (1.2)	−7.0 (0.9)	−2.1 (1.5)
GXR 0.046–0.075 mg/kg/day, n	147	84	
Total score	−17.9 (1.1)	−13.6 (1.5)	−4.3 (1.9)*
Hyperactivity/impulsivity subscale score	−8.4 (0.6)	−6.6 (0.8)	−1.8 (1.0)
Inattention subscale score	−9.4 (0.6)	−7.0 (0.9)	−2.4 (1.1)*

ADHD attention-deficit/hyperactivity disorder, *ADHD-RS-IV* ADHD Rating Scale IV, *ATX* atomoxetine, *GXR* guanfacine extended release, *SE* standard error

* $p < 0.05$ for comparisons between GXR vs. ATX trials

** $p < 0.01$ for comparisons between GXR vs. ATX trials

baseline variables and percentages of categorical variables were identical between all cohorts from the two pooled GXR trials and ATX trial; in addition, placebo group efficacy results were matched exactly between the GXR and ATX trials.

Results from the unadjusted comparisons of the base case and associated three-trial sensitivity analyses are shown in Table 2. Unadjusted comparisons reveal that the placebo outcomes were different between the GXR and ATX groups. Compared with patients on placebo in the ATX trial, patients on placebo in the pooled GXR trials demonstrated significantly greater reductions in mean (SE) ADHD-RS-IV total [−4.8 (1.6); $p < 0.01$] and subscale scores [hyperactivity/impulsivity: −1.8 (0.8); $p < 0.05$; inattention: −3.1 (1.0); $p < 0.01$]. This suggests that unobserved variables (e.g., conduct of the trials, baseline characteristics, etc.) were different between the ATX and GXR studies and cohorts. Therefore, differences in placebo effect from cross-trial differences in unobserved variables were adjusted for in the MAIC model. In the base-case [target doses of GXR (0.09–0.12 mg/kg/day) and ATX (1.2 mg/kg/day)] unadjusted comparison before matching, GXR-treated patients demonstrated better efficacy (as shown by statistically significantly greater reductions in mean change from baseline for ADHD-RS-IV total score and subscale scores for hyperactivity/impulsivity and inattention) compared with ATX-treated patients.

After matching for baseline characteristics and differences in the placebo effect across treatments, the MAIC base case analysis confirmed significantly greater reductions for patients receiving GXR 0.09–0.12 mg/kg/day compared with ATX-treated patients in mean (SE) ADHD-RS-IV total score [−7.0 (2.2); $p < 0.01$] and subscale scores for hyperactivity/impulsivity [−3.8 (1.2); $p < 0.01$] and inattention [−3.2 (1.3); $p < 0.05$; Fig. 3]. Patients receiving a lower dose range of GXR (0.075–0.090 mg/kg/day) demonstrated significantly greater reductions in ADHD total score and in the hyperactivity/impulsivity subscale score compared with ATX-treated patients; patients receiving 0.046–0.075 mg/kg/day GXR demonstrated significantly greater reductions in the ADHD-RS-IV inattentive subscale score compared with ATX-treated patients (Fig. 3).

In the six-trial sensitivity analysis, before matching, patients receiving the target GXR dose (0.09–0.12 mg/kg/day) demonstrated significantly greater reductions versus ATX in mean ADHD-RS-IV total score, as well as subscale scores for hyperactivity/impulsivity and inattention, respectively (Table 3). As noted in the base case analysis, patients receiving placebo in the pooled GXR trials also demonstrated significantly greater reductions in ADHD-RS-IV total and subscale scores compared with patients receiving placebo in the pooled ATX trials. After matching, GXR patients in the target dose range (0.09–0.12 mg/kg/day) demonstrated significantly greater reductions over ATX in

a Total scores

b Hyperactivity/impulsivity subscale scores

c Inattention subscale scores
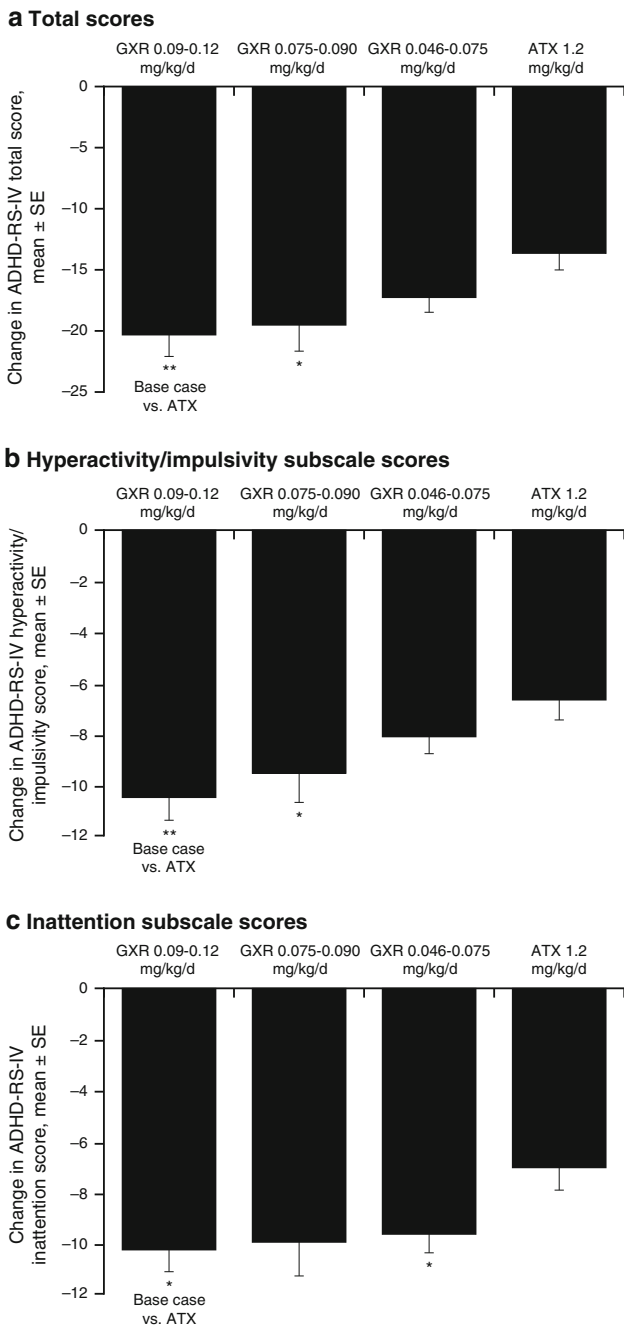
Fig. 3 Matching-adjusted indirect comparison of change from baseline at final on-treatment assessment in ADHD-RS-IV scores: three-trial base case and sensitivity analyses. *ADHD* attention-deficit/hyperactivity disorder, *ADHD-RS-IV* ADHD Rating Scale IV, *ATX* atomoxetine, *GXR* guanfacine extended release, *SE* standard error. *$p < 0.05$ compared with ATX. **$p < 0.01$ compared with ATX

mean (SE) ADHD-RS-IV total score [-7.6 (1.4); $p < 0.01$] and subscale scores for hyperactivity/impulsivity [-4.0 (0.8); $p < 0.01$] and inattention [-3.7 (0.8); $p < 0.01$], as did patients in the lowest GXR dose group (0.046–0.075 mg/kg/day) versus ATX. The differences between GXR 0.075–0.090 mg/kg/day and ATX in mean ADHD-RS-IV

total and subscale scores were not statistically significant (Fig. 4).

4 Discussion

The purpose of CER is "to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances" [19]. Such evidence-based information is important to clinicians, patients, payers and policymakers in determining treatment decisions and best practices. In 2009, the *American Recovery and Reinvestment Act* (ARRA) set aside more than US$1 billion to fund CER projects [19], highlighting the overall need for comparative evidence. It appears that CER will continue in the future, as the 2010 *Patient Protection and Affordable Care Act* was passed, building on the funding initiated by the ARRA by providing new and increased funding for CER until 2019 that will reach an annual total of US$600 million during this time period [20]. With this in mind, a cost-effective approach for generating CER data would be to leverage available clinical trial data at the individual patient level.

MAIC overcomes limitations of unadjusted ITC methods by utilizing IPD, to compare treatment outcomes across balanced populations. The utility of MAIC has been demonstrated in a number of other therapeutic areas, such as psoriasis, newly diagnosed chronic myelogenous leukemia, and type 2 diabetes [9].

With IPD available from GXR clinical trials, it was possible to select a specific group of patients given equipotent (based on mg/kg/day) doses and compare them with patients in the target (maximum effective) dose ATX trials. In the present analysis, the unadjusted comparisons revealed significantly larger decreases in ADHD-RS-IV scores with GXR at the target dose compared with ATX. However, patients in the placebo arm of the GXR trials also revealed significantly larger decreases in ADHD symptom scores compared with ATX. The apparent decreases in ADHD-RS-IV total and subscale scores in this unadjusted analysis may be impacted by baseline differences between the GXR and ATX cohorts as well as differences in placebo effect. Re-weighting of IPD using MAIC methodology allowed for exact matching of average baseline characteristics of age, sex, and baseline ADHD-RS-IV scores across cohorts. Although the cohorts were not significantly different in all baseline variables to begin with, it should be noted that even non-statistically significant baseline differences can lead to significant confounding effects. In the base case analysis, after adjusting

Table 3 Unadjusted comparison of change from baseline at final on-treatment assessment in ADHD-RS-IV scores: six-trial sensitivity analysis

Change in ADHD-RS-IV score, mean (SE)	GXR trials	ATX trials (≥1.2 mg/kg/day)	Comparison (GXR vs. ATX)
Placebo, n[a]	136	348	
Total score	−10.6 (1.1)	−5.8 (0.6)	−4.8 (1.3)**
Hyperactivity/impulsivity subscale score	−5.0 (0.6)	−2.7 (0.3)	−2.3 (0.7)**
Inattention subscale score	−5.6 (0.6)	−3.0 (0.3)	−2.6 (0.7)**
GXR 0.09–0.12 mg/kg/day, n	82	506	
Total score	−22.8 (1.3)	−14.6 (0.6)	−8.2 (1.4)**
Hyperactivity/impulsivity subscale score	−11.5 (0.7)	−7.2 (0.3)	−4.3 (0.8)**
Inattention subscale score	−11.3 (0.7)	−7.4 (0.3)	−3.9 (0.8)**
GXR 0.075–0.090 mg/kg/day, n	46	506	
Total score	−18.5 (2.1)	−14.6 (0.6)	−3.9 (2.2)
Hyperactivity/impulsivity subscale score	−9.4 (1.1)	−7.2 (0.3)	−2.2 (1.1)*
Inattention subscale score	−9.1 (1.2)	−7.4 (0.3)	−1.6 (1.3)
GXR 0.046–0.075 mg/kg/day, n	147	506	
Total score	−17.9 (1.1)	−14.6 (0.6)	−3.3 (1.3)**
Hyperactivity/impulsivity subscale score	−8.4 (0.6)	−7.2 (0.3)	−1.3 (0.7)
Inattention subscale score	−9.4 (0.6)	−7.4 (0.3)	−2.0 (0.7)**

ADHD attention-deficit/hyperactivity disorder, *ADHD-RS-IV* ADHD Rating Scale IV, *ATX* atomoxetine, *GXR* guanfacine extended release, *SE* standard error

* $p < 0.05$ for comparisons between GXR vs. ATX trials

** $p < 0.01$ for comparisons between GXR vs. ATX trials

[a] Sample sizes for ATX trials were based on patients with reported efficacy

for baseline differences and placebo arm response rates, patients on the target dose of GXR demonstrated significantly better efficacy compared with patients on the target dose of ATX in ADHD symptom reduction. Results of sensitivity analyses were consistent across a range of GXR doses below the target dose when compared with ATX doses at the target dose or higher, confirming the generalizability of the base case comparison to a broader and more heterogeneous trial population which may be clinically relevant.

Our results demonstrate that, on average, GXR is more efficacious than ATX for the treatment of ADHD in children and adolescents; these results demonstrated consistent significant reductions in symptoms in GXR over ATX in the base case analysis and in several of the sensitivity analyses. In the MAIC base case analysis, we observed an average reduction in ADHD-RS-IV total score of 7.0 for GXR compared with ATX. Similarly, in the six-trial sensitivity analysis, a significant reduction in ADHD-RS-IV total score of 7.6 points more for GXR compared with ATX was observed. An analysis by Zhang et al. [21] defined a between-treatment minimal clinically important difference for ADHD-RS-IV total scores as 6.6 points; this suggests that the ∼7 point reduction in ADHD-RS-IV total score on GXR compared with ATX demonstrated by the MAIC analyses is not only statistically significant but may also be clinically meaningful. The potential clinical impact

of these statistically significant differences should be explored further. Also, some researchers have attempted to correlate ADHD-RS-IV scores with scores on the Clinical Global Impressions-Improvement (CGI-I) and CGI-S scales, which correspond to global judgments regarding disease severity made by clinicians in practice [22]. In an evaluation of ADHD clinical trials of the effect of lisdexamfetamine dimesylate (Vyvanse®; Shire LLC, Wayne, PA, USA), Goodman et al. [22] found that differences in ADHD-RS-IV total score of 8–10 points were approximately correlated to a 1-level difference in CGI-S severity [e.g., from 4 (moderately ill) to 3 (mildly ill)]. However, the generalizability of those results to the current analysis of GXR and ATX is not known.

Our analysis has several limitations. As shown in Fig. 1, 92 trials (82 ATX and ten GXR) were assessed for eligibility, but the majority had to be excluded for a variety of reasons. For example, one trial [23] contained a study population that met the criteria for inclusion into our study; however, the paper did not report mean baseline ADHD-RS-IV total scores (required for MAIC) and therefore could not be used. Had more clinical trials been available for inclusion, different results might have been obtained. As is the case in any comparison of nonrandomized treatment groups, including MAIC, cross-trial differences in unobserved characteristics could have biased the comparison of outcomes. Only a direct treatment comparison through a

a Total scores

b Hyperactivity/impulsivity subscale scores

c Inattention subscale scores
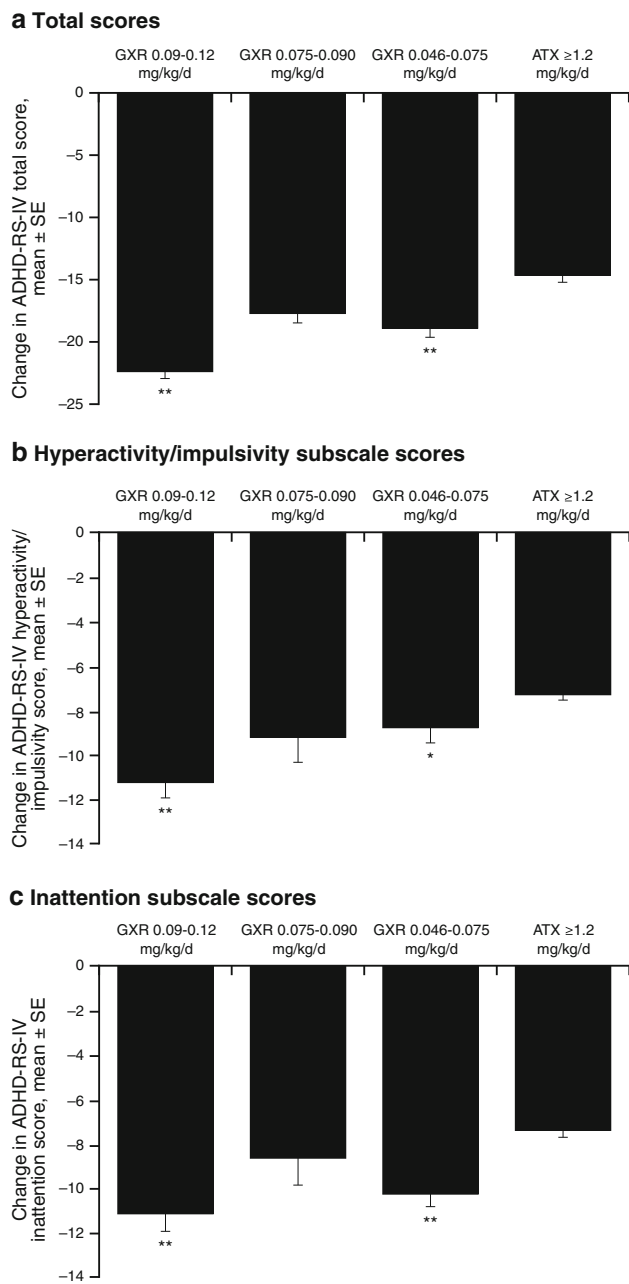
Fig. 4 Matching-adjusted indirect comparison of change from baseline at final on-treatment assessment in ADHD-RS-IV scores: six-trial sensitivity analysis. *ADHD* attention-deficit/hyperactivity disorder, *ADHD-RS-IV* ADHD Rating Scale IV, *ATX* atomoxetine, *GXR* guanfacine extended release, *SE* standard error. $*p < 0.05$ compared with ATX. $**p < 0.01$ compared with ATX

randomized trial could potentially avoid bias due to unobserved characteristics through to randomization and blinding. However, the fact that the average placebo effect was also matched reduces this concern. Another potential limitation is that the MAIC, like any adjustment procedure, reduces the effective sample size compared with the unmatched trial populations. This is an unavoidable

consequence of reducing variation between the different treatment groups. In addition, while the MAIC utilized IPD, the comparison between GXR and ATX was made at the population level. While our results show that GXR is more efficacious than ATX on average in the studied populations, particular individuals or subgroups may be more or less responsive to, or intolerant of, either medication. Individual responses indeed vary, and some individual patients may have better outcomes with ATX. Further patient-centered research would be necessary to better define patient characteristics associated with medication efficacy, including who may be a responder to GXR but not ATX and vice versa. As patient-centered medicine with a more personalized approach evolves, the MAIC method will be able to be potentially applied in subgroups to further help address such issues of patient heterogeneity. Also, an analysis of safety and tolerability with GXR versus ATX, while an important clinical consideration for treatment selection, was outside the scope of this CER. Additionally, the current study focuses on the short-term efficacy comparison between GXR and ATX. As ADHD is a chronic condition that may require long-term treatment, future research is needed to assess the long-term impact of these treatments. Lastly, our analysis compared GXR and ATX; CER studies using MAIC methodology between clonidine extended-release and either GXR or ATX have not yet been conducted to our knowledge.

5 Conclusions

Using the MAIC method, adjustments for observable cross-trial differences at baseline were made using IPD. Compared with patients administered ATX at the target dose of 1.2 mg/kg/day, patients administered GXR at the target dose of 0.09–0.12 mg/kg/day demonstrated statistically significant and clinically meaningful mean reductions from baseline in ADHD-RS-IV total and subscale scores [21]. Results were consistent across several sensitivity analyses, demonstrating the results' generalizability to a wider more heterogeneous trial population.

References

1. American Psychiatric Association. Attention-deficit and disruptive behavior disorders. Diagnostic and statistical manual of mental disorders, 4th ed, text revision (DSM-IV-TR). Washington, DC: American Psychiatric Association, 2000, p. 85–93.
2. Wolraich M, Brown L, Brown RT, et al. Implementing the key action statements: an algorithm and explanation for process of care for the evaluation, diagnosis, treatment, and monitoring of ADHD in children and adolescents. Pediatrics. 2011;128(5, supplemental information):SI1–21.
3. Wilens TE, Spencer TJ. The stimulants revisited. Child Adolesc Psychiatr Clin N Am. 2009;9(3):573–603 (viii).
4. STRATTERA® (atomoxetine hydrochloride) capsules for oral use [package insert]. Indianapolis: Eli Lilly and Company; 2011.
5. INTUNIV® (guanfacine) extended-release tablets [package insert]. Wayne: Shire Pharmaceuticals LLC. 2013.
6. Congressional Budget Office. Research on the comparative effectiveness of medical treatments: issues and options for an expanded federal role. 2007; 1–35.
7. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1. Value Health. 2011;14(4):417–28.
8. Ali R, Hanger M, Carino T. Comparative effectiveness research in the United States: a catalyst for innovation. Am Health Drug Benefits. 2011;4(2):68–72.
9. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. Value Health. 2012;15(6):940–7.
10. Piantadosi S, Byar DP, Green SB. The ecological fallacy. Am J Epidemiol. 1988;127(5):893–904.
11. Signorovitch JE, Wu EQ, Yu AP, et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. Pharmacoeconomics. 2010;28(10):935–45.
12. Biederman J, Melmed RD, Patel A, et al. A randomized, double-blind, placebo-controlled study of guanfacine extended release in children and adolescents with attention-deficit/hyperactivity disorder. Pediatrics. 2008;121(1):e73–84.
13. Sallee FR, McGough J, Wigal T, Donahue J, Lyne A, Biederman J. Guanfacine extended release in children and adolescents with attention-deficit/hyperactivity disorder: a placebo-controlled trial. J Am Acad Child Adolesc Psychiatry. 2009;48(2):155–65.
14. Kelsey DK, Sumner CR, Casat CD, et al. Once-daily atomoxetine treatment for children with attention-deficit/hyperactivity disorder, including an assessment of evening and morning behavior: a double-blind, placebo-controlled trial. Pediatrics. 2004;114(1):e1–8.
15. Michelson D, Faries D, Wernicke J, et al. Atomoxetine in the treatment of children and adolescents with attention-deficit/hyperactivity disorder: a randomized, placebo-controlled, dose-response study. Pediatrics. 2001;108(5):E83.
16. Michelson D, Allen AJ, Busner J, et al. Once-daily atomoxetine treatment for children and adolescents with attention deficit hyperactivity disorder: a randomized, placebo-controlled study. Am J Psychiatry. 2002;159(11):1896–901.
17. Spencer T, Heiligenstein JH, Biederman J, et al. Results from 2 proof-of-concept, placebo-controlled studies of atomoxetine in children with attention-deficit/hyperactivity disorder. J Clin Psychiatry. 2002;63(12):1140–7.
18. Efron B, Tibshirani RJ. An introduction to the bootstrap. Boca Raton: CRC Press; 1993.
19. Report to the President and the Congress on Comparative Effectiveness Research. http://www.hhs.gov/recovery/programs/cer/execsummary.html. Accessed 7 August 2012.
20. Clancy C, Collins FS. Patient-Centered Outcomes Research Institute: the intersection of science and health care. Sci Transl Med. 2010;2(37):37cm18.
21. Zhang S, Faries DE, Vowles M, Michelson D. ADHD Rating Scale IV: psychometric properties from a multinational study as a clinician-administered instrument. Int J Methods Psychiatr Res. 2005;14(4):186–201.
22. Goodman D, Faraone SV, Adler LA, Dirks B, Hamdani M, Weisler R. Interpreting ADHD rating scale scores: linking ADHD rating scale scores and CGI levels in two randomized controlled trials of lisdexamfetamine dimesylate in ADHD. Prim Psychiatry. 2010;17(3):44–52.
23. Block SL, Kelsey D, Coury D, et al. Once-daily atomoxetine for treating pediatric attention-deficit/hyperactivity disorder: comparison of morning and evening dosing. Clin Pediatr (Phila). 2009;48(7):723–33.