Quetiapine immediate release versus placebo for schizophrenia: Systematic review, meta-analysis and reappraisal.

Quetiapine immediate release versus placebo for schizophrenia: Systematic review, meta-analysis and reappraisal.

Paul Hutton[a]
Peter J. Taylor[b]
Lee Mulligan[c]
Sarah Tully[d]
Joanna Moncrieff[e]

[a]Department of Clinical Psychology
School of Health in Social Science
University of Edinburgh
Doorway 6, Medical Building,
Teviot Place
Edinburgh, EH8 9AG

[b]Department of Clinical Psychology
Institute of Psychology
Health & Society
University of Liverpool
Whelan Building
Brownlow Hill,
Liverpool, L69 3GB

[c]Department of Clinical Psychology
School of Psychological Sciences
University of Manchester
Manchester, M13 9PL.

[d]Psychosis Research Unit
Greater Manchester West Mental Health Foundation NHS Trust
Bury New Road
Prestwich,
Manchester, M25 3BL.

[e]Mental Health Sciences Unit
School of Life and Medical Sciences
University College London
London, WC1E 6BT.

Corresponding author: Dr Paul Hutton, Department of Clinical Psychology, School of Health in Social Science, University of Edinburgh, Doorway 6, Medical Building, Teviot Place, Edinburgh, EH8 9AG. Email: paul.hutton.cf@ed.ac.uk. Fax: +44 (0)131 651 3971. Tel: +44 (0)131 651 3972.

ABSTRACT

Background: Immediate-release (IR) quetiapine has been used to treat schizophrenia since 1997, despite all the principal placebo-controlled trials suffering from severe rates of missing outcome data (>50%), making their results difficult to interpret. New studies with relatively lower rates of drop-out have since been published.

Aims: Our objectives were; (1) to assess the efficacy and adverse effects of quetiapine IR for schizophrenia, with consideration of outcome quality and clinical meaningfulness of results; (2) to examine the potential impact of missing data on the main efficacy findings.

Method: We conducted a systematic review and meta-analysis of randomised controlled trials comparing quetiapine IR to placebo (or subtherapeutic dose in relapse prevention trials) for the acute and longer term treatment of schizophrenia. Primary outcomes were change in PANSS-rated overall symptoms and response rates (≥50% improvement in PANSS scores). We examined whether trials with severe drop-out (≥50%) had attenuated effect sizes and assessed the impact of making different assumptions about the outcome of those who left early.

Results: Our search identified 15 relevant trials, 11 of which assessed short-term efficacy (N=2259; 2-12 weeks). We obtained data for 2 unpublished trials, which together provided the first 12-week data for this drug, and the first data on self-reported quality of life. Previous research suggests patients require a mean reduction in PANSS total scores of at least 11 points to feel minimally better, whereas raters require at least 15 points. We found quetiapine IR has a Weighted Mean Difference (WMD) of 6.5 points (k=11; 95% Confidence Interval [CI] -8.9, -4; Standardised Mean Difference [SMD] -0.33, 95% CI -0.46, -0.21). The effect was not robust to changing assumptions about the outcome of the large number of people who left these trials early. Longer duration trials reported larger mean differences favouring quetiapine IR. Approximately 21 people needed to take quetiapine IR for one to experience at least a 50% improvement in PANSS scores (95% CI 13, 63). Response rates were smaller in more recent trials. No difference between quetiapine IR and placebo was observed on participant-reported quality of life (k=2; SMD 0.11, 95% CI -0.15, 0.36), according to previously unpublished data. Long-term quality of life data from two RCTs remains unpublished. Quetiapine IR caused sedation (NNH 9, 95% CI 7, 13) and clinically significant increases in weight (WMD 1.8kg, 95% CI 1.1, 2.4; SMD 0.64, 95% CI 0.43, 0.85; NNH 13, 95% CI 9, 23), but no extra-pyramidal effects were observed.

Conclusions: The original immediate release version of quetiapine has a small effect on overall psychotic symptoms over 2-12 weeks. Although larger benefits were observed in longer trials, the overall differences did not reach criteria for clinically significant change and were not robust to changing assumptions about the outcome of participants who left early. The probability of experiencing much improvement in symptoms was considerably smaller than the probability of clinically significant weight-gain or marked sedation.

Funding: There was no financial funding for this work.

Key words: Quetiapine, placebo, antipsychotics, meta-analysis, schizophrenia, psychosis.

Review Registration Prospero International Prospective Register of Systematic Reviews. Protocol No: CRD4201100165. Available from:
http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42011001651

Introduction

Quetiapine is a widely prescribed antipsychotic.(1) First manufactured by Astrazeneca, it was initially introduced for the treatment of schizophrenia and non-affective psychosis in the late 1990s, but is now licensed in several countries for the treatment of bipolar disorder and other conditions. The original immediate-release (IR) version was the third most frequently prescribed antipsychotic in the UK from 2004 to 2007 and, by 2008, had been taken by over 25 million people world-wide.(2) The patent for quetiapine IR expired in March 2012, and the generic version is now comparable in cost to haloperidol, leading to considerable cost-savings. Although older reviews of comparative and placebo-controlled trials have concluded it is an effective treatment for schizophrenia and non-affective psychosis, (3, 4) these were based on a limited number of trials which suffered from severe attrition. The Cochrane review, for example, noted that most of the original placebo-controlled studies were severely compromised by missing outcome data, and that their results were therefore *"impossible to interpret with confidence*".(3) Three of the 4 included studies were missing over 50% of their 6-week outcome data, and the remaining study had only 12 participants. In each case the outcomes of those leaving early were estimated by the method of carrying their last available observation forward (LOCF), an imputation strategy now regarded as unreliable.(5)

In their 2009 review, Leucht and colleagues found that, overall, second-generation antipsychotics had a moderate effect on symptoms (Hedges's g = 0.51).(4) However they suggested that the high drop-out in these studies may have attenuated the drug-placebo difference. Indeed, most placebo-controlled trials have more than 25% drop-out, and a significant number have over 50%.(6) Such high rates of missing data cannot be safely ignored. A recent survey by the Cochrane Schizophrenia Group found consultant psychiatrists, service users, carers and Cochrane researchers in agreement that trials with over 25% missing data lack credibility, (7) and there is largely a consensus that no statistical approach can produce reliable results when assumptions about the outcomes of participants carry more weight than actual observations (7, 8). Understanding the impact of missing data is particularly challenging if it is missing for non-random reasons which are related to outcome, as may be the case for antipsychotic trials. (9) The development of a sustained release version of quetiapine (quetiapine XR) has led to new randomised controlled trials (RCTs) comparing the IR version to placebo, some of which had relatively low rates of missing data. Due to the uncertainty introduced by high attrition in the older IR studies, we set out to perform a new systematic review and meta-analysis.

We had two main objectives. Our first was to provide a comprehensive assessment of the efficacy and adverse effects of quetiapine IR for schizophrenia when compared to placebo, with consideration of both outcome quality and the clinical meaningfulness of the results, as informed by recent advances in our understanding of what constitutes a Minimum Clinically Important Difference (MCID) in PANSS total scores.(10-12). Our second objective was to examine the potential impact of missing data on the primary outcomes. More specifically, we examined whether trials with high rates of missing data had smaller effect sizes (4) and we used a recently published approach to examine the impact on our efficacy estimates of changing assumptions about the likely outcomes of the high numbers of people who leave these trials early.(13)

Method

Search

Our search strategy and protocol detailing our inclusion and exclusion criteria are provided in a supplementary file. Two researchers independently searched publication databases, clinical trial registries and previous reviews for RCTs where participants with a schizophrenia diagnosis or early psychosis were randomly allocated to receive double-blind treatment with either placebo or quetiapine IR. No pre-specified limits were placed on study duration.

Data extraction and outcomes

Two reviewers independently extracted data from each study, using data extraction forms. We attempted to trace missing summary data by contacting first authors or the study sponsor. Our primary outcomes were the average reduction in total scores on the Positive and Negative Syndrome Scale (PANSS) at study endpoint or, if not available, Brief Psychiatric Rating Scale (BPRS) and the numbers of people achieving an important clinical response. We defined the latter as ≥50% reduction in PANSS / BPRS scores.(14) When these were not reported or provided, we imputed them from means and SDs using the validated method of Furukawa and colleagues (15) (see 'changes from protocol' section in the supplementary file).

Our secondary efficacy outcomes included relapse, positive symptoms, negative symptoms, depression, quality of life and needing additional antipsychotic medication or sedatives. We also examined the numbers leaving

early for any reason, need for hospital care and functioning. For adverse effects, we looked at use of antiparkinson medication, extrapyramidal side-effects, drop-out due to adverse events, sedation, total number of drug-attributable adverse events, insomnia, weight-gain and weight-loss.

We used a strict intention-to-treat (ITT) analysis for dichotomous outcomes, using the total numbers randomised to each group as the denominator in each case. Where possible, we assumed those leaving early or otherwise unaccounted for had an unchanged outcome from randomisation, but carried out sensitivity analyses to test this. Data incorporating last-observation carried forward (LOCF) assumptions were used only when there was no alternative.

We also wished to use a strict ITT analysis for continuous data, but expected to be limited to summary data derived from smaller samples excluding early drop-outs or those without at least one post-baseline assessment. For all outcomes, we intended to use summary data based on the mixed-model repeated measures (MMRM) imputation method, followed by LOCF or observed case (OC) data if not available. Missing standard deviations (SDs) were, where possible, calculated from t-values, p-values, standard errors or confidence intervals.(16) If no variance parameters were reported for a particular study, we imputed SDs using the median SD of the other studies. Based on previous studies, (4, 17) we planned to use data from study arms where participants received an optimal dose of >250mg. However we carried out a sensitivity analysis excluding doses of <400mg, as per the recent International Consensus on Antipsychotic Dosing (18) and recent Leucht group analysis.(19)

Meta-analytic calculations
For continuous data, we calculated the Hedges's g standardised mean difference (SMD) using Comprehensive Meta-Analysis Version 2. For the primary analysis of 2-12 week study endpoint data, we converted BPRS means and SDs to PANSS means and SDs using recently published conversion charts (PANSS total scores = ~1.538 x BPRS total scores) (20) thus allowing us to also present the unstandardized weighted mean difference (WMD) in PANSS total scores for all the studies combined. When a trial had two or more relevant arms, we combined the data following procedures in the Cochrane Handbook.(16)

For binary data, we calculated the relative risk (RR) of the unfavourable outcome, together with 95% confidence intervals, as well as the absolute risk difference (RD) and numbers needed to treat (NNT) or harm (NNH). If a trial had eligible binary data from two or more active treatment arms, we combined these into one. We used a random-effects analysis for all outcomes. For the primary outcomes we also performed a sensitivity analysis using fixed-effects, but not if heterogeneity was moderate or more, defined as an I^2 statistic of ≥40%.(16)

Assessing impact of missing data
We tested the hypothesis that trials with severe rates of missing data (≥50% at endpoint) had smaller drug-placebo differences on our primary outcomes than trials with less severe rates (<50%). The 50% cut-off was chosen because it marks the point at which estimated data carries more weight than actual observations, and because NICE, the Cochrane Schizophrenia Group and others often exclude trials with this degree of missing data from their reviews. (6, 21, 22) We also wished to compare studies with <25% and ≥25% attrition at endpoint,(7) but we were unable to do so because no 6-12 week studies had <25% attrition.

When observed case data were available, we were also able to examine the impact of missing data on the primary outcome by imputing values for those who left the trial early, using new guidelines provided by Ebrahim and colleagues.(13) Their method involves testing whether the overall treatment effect is robust under four increasingly more conservative strategies – two of which we applied here. Strategy 1 is non-extreme and involves replacing missing data in both arms of each trial with the observed case mean of the control arm. Strategy 2 is more conservative yet plausible, and uses the highest observed control arm mean to replace missing control arm data, and the lowest observed intervention arm mean to replace missing intervention arm data. For both approaches, we imputed the missing data treatment and placebo SDs with the median SD of the control arms of all the included trials, as recommended.(13)

Analysis of clinical significance
The Minimum Clinically Important Difference (MCID) has been defined by Jaeschke and colleagues as: *"the smallest difference in a score in the domain of interest which patients [or providers] perceive as beneficial and which would mandate in the absence of troublesome side-effects and excessive cost, a meaningful change in the patient's management"*.(11, 23) An analysis of data from 14 antipsychotic trials (10) (N=5970) found a rater-determined MCID on the PANSS of roughly 15 points, a criterion which has since been replicated by separate analyses of two large non-industry effectiveness trials (N=1650).(11, 12) Data from a large naturalistic study (N=398) suggested a lower criterion of 10 points, (24) which is similar to the patient-rated MCID of 11 points

derived from the CATIE trial.(12) We tested the validity of these definitions by comparing them to the median of mean changes the included trials were designed to detect. We assumed that trial sponsors had provided enough resources to detect with adequate power what they regarded, a priori, as the smallest difference between the groups that is important to detect.(25)

Risk of bias and study quality
Two raters independently assessed both study-level risk of bias with the Cochrane Collaboration Risk of Bias tool, (16) and outcome quality using the GRADE approach.(26) Further details on methods and ratings are provided in a supplementary file. We tested for publication bias using funnel plots for the PANSS/BPRS total score effect sizes (Hedges's g) of all studies. Ratings of bias and quality were used to inform interpretation of reliability and magnitude of effects.

Pre-registration of review protocol and subsequent changes
The review protocol was registered in advance with PROSPERO (International Prospective Register of Systematic Reviews). Subsequent changes, in addition to those outlined above, are detailed in the supplementary file. We abandoned the use of the response-rate hierarchy used by Leucht et al (4), given their recently expressed concerns that response rate estimates are particularly vulnerable to selective reporting bias (19), and used only the top of the hierarchy instead (50% or more reduction in PANSS / BPRS scores). This criteria is now recommended for use in studies of acutely ill, non-refractory patients (27), and we used the Furukawa et al method to impute this data when not reported or provided (15). This method has been recently validated using individual patient-level data from 16 antipsychotic trials (28), and has the additional advantage of allowing for the use of adjusted PANSS and BPRS total scores when calculating percentage change, thus avoiding underestimation of response (27). Additional changes included using meta-regression to assess the association between (i) study duration (measured in weeks) and (ii) year of publication on total symptoms and clinically significant improvement. These were conducted in STATA version 9 using the Metareg command and Knapp-Hartung variance estimator (29).

Results

Search
The process of selecting studies is detailed in Figure 1. We identified 15 relevant trials, 11 of which assessed short-term efficacy (N=2259). Lundbeck provided us with summary reports for two unpublished 12-week placebo-controlled studies, (30, 31) both of which were terminated early due to inefficacy of the investigational drug (bifeprunox; quetiapine IR was an active comparator in these trials). Astrazeneca, the makers of quetiapine, provided us with a considerable amount of additional unpublished data in relation to many of their trials. They decided not to provide us with the report for one unpublished long-term trial comparing therapeutic and sub-therapeutic doses of quetiapine (32), arguing the lack of a placebo control meant it did not meet our original inclusion criteria. However we managed to acquire an extract detailing the main results, and other published summaries allowed us to partially assess risk of bias. We therefore included data from a total of 15 studies. An overview of included studies in provided in Table 1, and excluded studies are listed in the supplementary file, together with a table of trial characteristics and baseline demographics.

Risk of bias and GRADE
Table 1 also provides the main risk of bias ratings and the right-hand columns of Tables 2 and 3 provide the outcome quality ratings for the main primary and safety outcomes. Ratings for secondary outcomes and additional safety outcomes are provided in the supplementary file. Our rationale for the ratings is also provided in the supplementary file, alongside ratings produced by other research groups (where available). In our judgement, the main problem with these trials is a somewhat high risk of selective reporting bias in relation to secondary outcomes and adverse effects, coupled with a very high risk of attrition bias for most outcomes. We also judge unblinding due to sedative effects to be likely (33), and the double-blind design might not protect against the risk of researchers adopting a high threshold for recording effects (eg., adverse effects) where the desired outcome is 'no difference' (34). There is also evidence from documents released through legal proceedings in the US that Astrazeneca have historically not published all active-comparator quetiapine trials, or have not reported all outcomes. (35)

Validation of Minimal Clinically Important Difference criterion
Researchers and trial sponsors designed their trials to detect with adequate power a mean change in PANSS total scores, or equivalent, of approximately 12 points (range 9 to 15.5), which corresponds to an SMD of 0.55, and is similar in magnitude to the empirically derived estimate of Minimal Clinically Important Difference of 11-15 points. (10-12).

Outcomes

Primary efficacy outcomes (Table 2 and Figures 2-3)
Moderate to high quality evidence suggested quetiapine IR was statistically superior to placebo from 2 to 12 weeks in terms of reducing overall symptoms, but the effect was small (WMD -6.5 points, 95% CI -8.89, -4.00; SMD -0.33, 95% CI -0.46, -0.21) and the 95% confidence intervals excluded the MCID of 11-15 points. Low to moderate quality evidence suggested the NNT for much improvement was 21 (95% CI 13, 63).

Sensitivity analyses and meta-regression
These analyses identified a significant effect of study duration (weeks) on the effect size for total PANSS, B = -0.04, (95% CI: -0.08, -0.01), p = .02, with a more treatment-favourable outcome associated with longer duration. Treatment duration did not significantly moderate the effect for treatment response, B < -0.01, RR = 1.00 (95% CI 0.98, 1.01), p = .70. Excluding the two 2-week studies (36, 37) was associated with a marginal increase in average PANSS change (WMD -7.7 points, 95% CI -10.0, -5.3; SMD -0.38, 95% CI -0.50, -0.26) and response rates (NNT 19, 95% CI 11, 59). Year of publication did not significantly predict outcomes for total PANSS, B = 0.01 (95% CI -0.01, 0.04), p = .28, but there was a small association for treatment response, B = .01, RR = 1.01 (95% CI 1.00, 1.02), p = .01, with a less treatment favourable outcome associated with a more recent publication date. Meta-regression bubble-plots are provided in the supplementary file.

Removing data from arms employing <400mg dose reduced the contribution made by two trials (38, 39) but this had little effect on the overall estimate of average change (WMD -6.3, 95% CI -8.7, -3.8; SMD -0.32, 95% CI -0.44, -0.20) or response rates (NNT 20, 95% CI 13, 53). Excluding the adolescent study (40) also had little effect on estimates (WMD -6.3, 95% CI -8.9, -3.6; SMD -0.32, 95% CI -0.45, -0.19; NNT 22, 95% CI 13, 83).

The impact of missing data
Overall, the seven 2-12 week trials with less than 50% attrition had a mean PANSS advantage of -5.4 points (95% CI -8, -2.9; SMD -0.29, 95% CI -0.44, -0.15) and an NNT of 42 (19, 250H), whereas the four 6-week studies with 50% or more attrition had a mean PANSS advantage of 9.2 points (95% CI -15, -3.4; SMD -0.39, 95% CI -0.62, -0.17) and NNT of 13 (95% CI 7, 250). The three 6-week studies with less than 50% attrition had a mean advantage of 6.1 points (95% CI -9.9, -2.3; SMD -0.27, 95% CI -0.43, -0.12) and a nonsignificant NNT of 35 (95% CI 12, 42H).

Strategy 1 of the Ebrahim approach (13) involved testing whether the overall results would be different if we assumed that participants who dropped out early from both groups had the same degree of change as participants in the control group who stayed until the end. To illustrate this, consider the Small et al trial (41). Here the mean change for the 49 quetiapine and 39 placebo group completers was, after conversion of BPRS to PANSS scores, -23.3 points (SD 17.7) and -14.9 points (SD 17.7) respectively – a between-group difference of around 8.4 points. Carrying forward the last available scores of the 104 non-completers in this trial reduced the quetiapine estimate to -13.5 points (SD 24.5; N=94) and the placebo estimate to -1.5 points (SD 24.0; N=94), and increased the between-group difference to around 12 points. These are the figures we used in the main analysis. Introducing the Strategy 1 assumption that non-completers had a similar outcome to the placebo group completers (-14.9 points) reduced the overall estimate for the quetiapine group to -19.1 points (SD 18.2), and reduced the advantage over placebo to 4.2 points. We carried out the same procedure for the other 5 trials for which we had completer data and where no usable MMRM estimates were provided (38, 39, 42-44), and entered the revised estimates into the overall meta-analysis. Table 2 shows the overall advantage for quetiapine over placebo fell to 4.3 points (95% CI -6.5, -2; SMD -0.23; 95% CI -0.35, -0.11).

Strategy 2 of the Ebrahim approach involved testing whether the overall results were robust to assuming that (a) quetiapine non-completers had the smallest treatment response observed and (b) placebo non-completers had the largest placebo response observed. In Small et al 1997, this involved assuming the 47 quetiapine non-completers had the same degree of response as quetiapine completers in the Lindenmayer 2008 study (-17.4 points) (39), and that the 57 placebo non-completers had the same degree of response as placebo completers in the Kahn 2007 study (-23.1 points) (43). The revised quetiapine and placebo estimates were -20.4 (SD 18.2) and -19.8 points (SD 18.3) respectively, leading to a between-group difference of 0.6 points. As shown in Table 2, applying Strategy 2 to the 6 trials where we had completer data reduced the overall advantage for quetiapine to 2.7 points (95% CI -5.5, 0.2; SMD -0.15, 95% CI -0.30, 0.01). Revised forest-plots for Strategies 1 and 2 are provided in the online supplementary file.

Publication Bias

We detected some asymmetry in the funnel-plot of clinically significant change, but not in relation to mean change in overall symptoms or most other outcomes. Funnel-plots for the primary outcomes are provided in the supplementary file.

Secondary efficacy outcomes (Supplementary file, sections H & K)

Relapse, exacerbation and need for hospital care
There was evidence from one study suggesting quetiapine IR was effective for prevention of symptom exacerbation in people with early psychosis who have responded to quetiapine (45) but an unpublished study suggested there was no effect of therapeutic dose (300-600mg) over subtherapeutic dose (75mg) in relapse prevention in chronic schizophrenia.(32) The combined estimate was therefore heterogeneous (I^2=87%) and not significant (NNT 5; 95% CI 2, 13H). Quetiapine IR was associated with a marginally reduced need for hospital care after 2-6 weeks (3 RCTs) (36, 37, 44) (NNT 19, 95% CI 10, 143). One trial suggested quetiapine IR had a small effect over 52 weeks in relation to reducing rehospitalisation due to relapse (45) (NNT 11, 95% CI 6, 143) but the results were not robust to changing assumptions about the outcome of those leaving early. Overall the relapse and rehospitalisation data were very low to low in quality.

Other outcomes
There was a small effect on positive symptoms (SMD -0.32, 95% CI -0.44, -0.20; moderate quality evidence) and a marginal to small effect on negative symptoms (SMD -0.21, 95% CI -0.32, -0.10; moderate quality evidence) over 2-12 weeks, and a marginal effect on depression over 2-6 weeks (SMD -0.13, 95% CI -0.23, -0.02; low quality evidence). Forest-plots are provided in the supplementary file. We did not investigate whether these estimates were robust to missing data, but the sensitivity analyses for the primary outcome of total symptoms suggests this is unlikely. Those taking quetiapine IR had a marginally reduced need for additional sedative medication after 2-6 weeks (NNT 34; 95% CI 13, 53H; 6 RCTs), but we judged the evidence as low quality because of selective reporting and missing data, and no reduced need for antipsychotic medication was observed in the two 6-week RCTs where additional medication was not restricted (NNT 24, 95% CI 7, 19H; moderate quality evidence) (36, 37).

Pooled self-report endpoint data from the two 12-week trials (30, 31) did not indicate any benefit of quetiapine IR on quality of life, as measured by the Schizophrenia Quality of Life (S-QoL) scale (SMD 0.11, 95% CI -0.15, 0.36), but we judged the evidence to be very low in quality due to early termination of the trials, missing data and possible selective reporting from the other trials. No significant effects were observed on any of the subscales, including psychological well-being (SMD -0.02 95% CI -0.28, 0.24) or family relationships (SMD 0.01, 95% CI -0.25, 0.28). Since only observed case S-QoL data were reported, we imputed missing data using Strategy 1 from Ebrahim and colleagues (13). This reduced the overall effect from 0.11 to 0.06 (95% CI -0.14, 0.27). Long-term quality of life data from two RCTs (32, 45) remains unpublished.

An analysis of data from 3 RCTs (1 adolescent; 2 adult) covering a 6-12 week period (30, 31, 40) found quetiapine IR had a small-moderate benefit on functioning as assessed by a combination of Children's Global Assessment Scale (CGAS) data and Personal and Social Performance (PSP) data (SMD 0.39, 95% CI 0.18, 0.60). GAF data was also reported, but unlike the PSP this assesses symptom severity as well as functioning. After imputing missing PSP data using Strategy 1, the effect size was small (SMD 0.28, 95% CI 0.09, 0.46). One study found no benefit of 12 months of quetiapine IR maintenance treatment over placebo in relation to employment status.(45) Overall the functioning and employment data were very low in quality due to selective reporting, early termination of studies, imprecision and missing data. High quality evidence from 11 trials suggested quetiapine IR had a marginal effect on rates of early discontinuation over a 2-6 week period (NNT 21, 95 CI 10, 333H).

Safety outcomes (Table 3, Figures 4-6)
There was low quality evidence that quetiapine IR was associated with a small to moderate increased risk of non-serious adverse effects over the short-term (NNH 11, 95% CI 8, 22). There was no evidence of extrapyramidal side-effects, and no evidence of an increased risk of serious adverse events. Moderate quality evidence suggested no need for additional antiparkinson medication in quetiapine-treated participants over the short-term, but longer-term data was not reported.

Data from 12 trials suggested quetiapine IR had a moderate to large effect on weight-gain over 2-12 weeks (SMD 0.64, 95% CI 0.43, 0.85). Participants gained an extra 1.75kg (95% CI 1.10kg, 2.40kg) on average (~4lbs), but we rated the evidence as very low quality because of non-reporting of variance parameters in 7 out of 12 studies, high rates of drop-out, and high heterogeneity. Moderate quality evidence suggested around 12%

of quetiapine-treated participants experienced a clinically significant increase in weight over 2-12 weeks, compared to 4% of those taking placebo (NNH 13, 95% CI 9, 23) and 35% reported sedation or somnolence as an adverse effect, compared to 6% of those taking placebo (NNH 9, 95% CI 7, 13). Details on additional safety outcomes and forest-plots are reported in sections I & K in the supplementary file.

Discussion

Main findings
Using published and unpublished data, the average change in PANSS total scores attributable to quetiapine IR over a 2-12 week period was small. Although the 95% confidence intervals excluded the minimum clinically important difference of 11-15 points, it should be noted that few treatments for psychosis reach this threshold (19). Furthermore, as study duration increased, so did the effect size. Marginal advantages were observed at 2 weeks, whereas moderate effects that approached the threshold for change of at least minimal clinical importance were observed in two as yet unpublished 12-week trials. On the other hand, the overall effect size was smaller if we assumed the substantial number of participants who left early would have had the same degree of change as placebo-treated participants who stayed. There was no evidence to suggest that drug-attributable benefits have been underestimated because of severe rates of drop-out in the older studies. Approximately 21 people needed to take quetiapine IR for one to experience much improvement, defined in accordance with recent recommendations as ≥50% reduction in PANSS total scores (27).

Null to small effects were observed for depression and negative symptoms, respectively. Although moderate effects on positive symptoms were observed in the two unpublished 12-week trials, the pooled effect over 2-12 weeks was small. The two 12-week trials also reported small to moderate effects on functioning but found no difference between quetiapine IR and placebo on participant-reported quality of life. Quetiapine IR caused weight-gain and sedation, but did not lead to extrapyramidal side-effects. Although there was no evidence of increased serious adverse effects, the evidence was very low quality due to imprecision and incomplete reporting.

We found some evidence that estimates of clinically significant response derived from more recent trials were smaller in magnitude than older trials, which is consistent with results from other antipsychotic meta-analyses (4, 46), although no relationship between publication year and total symptoms was observed. It remains unclear whether reduced antipsychotic response in recent, large multisite trials with multiple treatment arms reflects a change in the characteristics of participants taking part, improvements in study quality or reporting, increased variability due to an increasing number of sites, (47) or better masking of treatment allocation due to reduced expectation of receiving placebo in such trials. (48)

Older reviews of quetiapine IR have reported effect sizes of around 0.4 when compared to placebo (3, 4, 19), and NNTs of around 10 to 11. (3, 4) However most of the trials available at the time had severe rates of drop-out and examined quetiapine IR as the target drug for regulatory approval, rather than as a control for other preparations. Previous reviews have not been able to account for selective reporting in relation to response rates, examine the impact of changing assumptions about missing data on estimates, or consider the clinical significance of the quetiapine IR-attributable change. A more recent review pooled 4-12 week outcome data for quetiapine IR with outcome data for the more recent 'extended release' version of quetiapine (quetiapine XR) and reported an overall moderate effect size of 0.44 (19). Since XR was judged by its manufacturer to be sufficiently novel to warrant a separate patent application and significantly greater cost, our a priori view was that pooling the XR and IR data would give an inaccurate appraisal of both formulations. In relation to duration, we planned to include 2-week quetiapine IR data in our review because this was the approach favoured by preceding reviews which were available at the time of protocol writing (4). We adhered to this decision because several prescribing guidelines recommend a minimum 2-week trial and evidence on prescribing practices suggests psychiatrists normally wait only 3 to 3.5 weeks before switching to another antipsychotic because of non-response. (49) Nonetheless, it is important to consider that overall efficacy was positively associated with trial duration in our review, and may have been larger still had we included quetiapine XR data. Our data might help explain why a recent meta-analysis found that quetiapine IR was significantly less effective at reducing positive symptoms than first generation antipsychotics (9 RCTs).(17) In the unpublished Study 15,(32) therapeutic dose quetiapine IR was significantly less effective than haloperidol in preventing relapse.

Missing data was high in the included trials. In order to reduce it, trial researchers should continue to assess participants who stop treatment early, as this will inform realistic estimates of likely outcome had they stayed, both in relation to efficacy and adverse effects. Since many early studies of other second-generation antipsychotics also suffered from severe attrition, (6) the robustness of their effects to changing assumptions

about missing data may also need to be examined. Although meta-regression has been used to examine the relationship between drop-out and effect size, (19) such analyses are inevitably limited by the fact that very few trials have low rates of missing data.(6) Application of the Ebrahim approach (13) to this data would help prescribers and patients appreciate the extent of uncertainty in estimates of antipsychotic benefits and costs.

In addition to attrition bias, the proper assessment of both drug and non-drug treatments for psychosis continues to be limited by incomplete and selective reporting of outcomes, low external validity and non-publication of negative trials (8). Indeed, a recent meta-analysis found the median effect of currently available antipsychotics over placebo fell from moderate to small after adjusting for the tendency for small studies to report larger effects (19), and selective reporting bias is a particular concern when, as documented by Spielmans and Parry (35) and others, it biases our understanding of the severity of adverse effects.

Limitations
We took advantage of several important developments in methodology which were published after we registered our protocol (13, 20, 28). Post hoc changes do raise a risk of bias, but we had to balance this against taking the opportunity to increase the quality, robustness and usefulness of our estimates, and we hope we have provided enough information for readers to judge the merit of these decisions.

Our claim that 11-15 points is required for minimal clinical improvement might be controversial, not least because few treatments achieve such change in psychosis (19). However we note the evidence supporting this minimum threshold is now quite consistent across different populations (10-12) and we demonstrated that quetiapine trial researchers only designed their trials to be able to detect, with adequate power, differences of approximately 12 points. Although it has been argued that small benefits might have value at a public health level, (4) there is clearly a need for further debate on this issue. As with non-inferiority and equivalence trials, (34) researchers planning superiority trials might consider stating explicitly, in advance, what they believe constitutes a minimal important difference on continuous outcomes. Although this can be inferred from power calculations, it needs to be stated explicitly.

We were unable to access the full clinical study reports (CSR) for each trial, which is problematic given a recent study found a much better quality of reporting in these documents, when compared to registry reports or peer-reviewed publications (50). Although we have acquired a large amount of previously unpublished data, access to the CSRs would have raised the quality of many of the outcomes, in particular the assessments of mean weight gain and response rates. Acquiring unpublished data was challenging, and we doubt we would have been successful had a public debate on publication bias in clinical trials not been taking place at the time. This is an unsatisfactory and unsustainable situation, and a change in the law is clearly required to ensure that *"all trials past and present [are] registered, and the full methods and the results reported"* (Alltrials campaign, 2012)."

Acknowledgements

We thank the following individuals and their organisations for providing us with unpublished data from various clinical trials: Dr Carla Canuso from Johnson & Johnson, Rakesh Kantaria, John Ramsey, Jasmine Lichfield and Craig Shering from Astrazeneca and Mads Kronborg and Dr Andrew Roberts from Lundbeck. We also thank Giovana Pezzi from Astrazeneca, Jan Yonge and Dr Chris Bushe from Eli Lilly, Sanofi (Medical Information), the UK Medicines and Healthcare Regulatory Authority, the Danish Medicines Agency, Dr Ben Goldacre, Edd Howard and Kerry Roberts.

Funding: There was no financial funding for this work.

Competing interests: PH has been a co-investigator on a National Institute of Health Research (NIHR) funded pilot trial of cognitive therapy (CT) for people with psychosis who refuse antipsychotics, and is a co-investigator on another NIHR-funded pilot trial of CT versus antipsychotics in early psychosis; JM is co-chairperson of the Critical Psychiatry Network. PJT, LM and ST declare they have no competing interests.

REFERENCES

1. NHS. Schizophrenia Data Focussed Commentary: NHS National Prescribing Centre; 2009. Available from: http://www.npci.org.uk/therapeutics/cns/schiz/resources/dfc_schizophrenia.pdf.
2. MHRA. UKPAR Seroquel XL 50mg, 200mg, 300mg & 400mg prolonged-release tablets: Medicines and Healthcare products Regulatory Authority; 2008. Available from: http://www.mhra.gov.uk/home/groups/l-unit1/documents/websiteresources/con030966.pdf.
3. Srisurapanont M, Maneeton B, Maneeton N. Quetiapine for schizophrenia. Cochrane Database Syst Rev. 2004(2):CD000967.
4. Leucht S, Arbter D, Engel RR, Kissling W, Davis JM. How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. Molecular Psychiatry. 2009 Apr;14(4):429-47.
5. Hamer RM, Simpson PM. Last observation carried forward versus mixed models in the analysis of psychiatric clinical trials. Am J Psychiatry. 2009 Jun;166(6):639-41.
6. Hutton P, Morrison AP, Yung AR, Taylor PJ, French P, Dunn G. Effects of drop-out on efficacy estimates in five Cochrane reviews of popular antipsychotics for schizophrenia. Acta Psychiatr Scand. 2012 Apr 6.
7. Xia J, Adams CE, Bhagat N, Bhagat V, Bhoopathi P, El-Sayeh H, et al. Losing participants before the end of the trial erodes credibility of findings. Psychiatric Bulletin. 2009;33:254-7.
8. Leucht S, Heres S, Hamann J, Kane JM. Methodological issues in current antipsychotic drug trials. Schizophrenia Bulletin. 2008 Mar;34(2):275-85.
9. Rabinowitz J, Davidov O. The association of dropout and outcome in trials of antipsychotic medication and its implications for dealing with missing data. Schizophr Bull. 2008 Mar;34(2):286-91.
10. Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR. Linking the PANSS, BPRS, and CGI: clinical implications. Neuropsychopharmacology. 2006 Oct;31(10):2318-25.
11. Thwin SS, Hermes E, Lew R, Barnett P, Liang M, Valley D, et al. Assessment of the minimum clinically important difference in quality of life in schizophrenia measured by the quality of well-being scale and disease-specific measures. Psychiatry Res. 2013 Mar 6.
12. Hermes ED, Sokoloff D, Stroup TS, Rosenheck RA. Minimum clinically important difference in the Positive and Negative Syndrome Scale with data from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE). J Clin Psychiatry. 2012 Apr;73(4):526-32.
13. Ebrahim S, Akl EA, Mustafa RA, Sun X, Walter SD, Heels-Ansdell D, et al. Addressing continuous data for participants excluded from trial analysis: a guide for systematic reviewers. J Clin Epidemiol. 2013 Sep;66(9):1014-21 e1.
14. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? Schizophr Res. 2005 Nov 15;79(2-3):231-8.
15. Furukawa TA, Cipriani A, Barbui C, Brambilla P, Watanabe N. Imputing response rates from means and standard deviations in meta-analyses. Int Clin Psychopharmacol. 2005 Jan;20(1):49-52.
16. Higgins JPT, Green S, Collaboration C. Cochrane handbook for systematic reviews of interventions. Chichester, England ; Hoboken, NJ: Wiley-Blackwell; 2011.
17. Leucht S, Corves C, Arbter D, Engel RR, Li C, Davis JM. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. Lancet. 2009 Jan 3;373(9657):31-41.
18. Gardner DM, Murphy AL, O'Donnell H, Centorrino F, Baldessarini RJ. International consensus study of antipsychotic dosing. Am J Psychiatry. 2010 Jun;167(6):686-93.
19. Leucht S, Cipriani A, Spineli L, Mavridis D, Orey D, Richter F, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet. 2013 Jun 26.
20. Leucht S, Rothe P, Davis JM, Engel RR. Equipercentile linking of the BPRS and the PANSS. Eur Neuropsychopharmacol. 2013 Feb 21.
21. NICE. Schizophrenia: core interventions in the treatment and management of schizophrenia in adults in primary care and secondary care (update): National Institute for Health and Clinical Excellence; 2009.
22. Bagnall AM, Jones L, Ginnelly L, Lewis R, Glanville J, Gilbody S, et al. A systematic review of atypical antipsychotic drugs in schizophrenia. Health Technol Assess. 2003;7(13):1-193.
23. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989 Dec;10(4):407-15.
24. Schennach-Wolff R, Obermeier M, Seemuller F, Jager M, Schmauss M, Laux G, et al. Does clinical judgment of baseline severity and changes in psychopathology depend on the patient population? Results of a CGI and PANSS linking analysis in a naturalistic study. J Clin Psychopharmacol. 2010 Dec;30(6):726-31.
25. Man-Son-Hing M, Laupacis A, O'Rourke K, Molnar FJ, Mahon J, Chan KB, et al. Determination of the clinical importance of study results. J Gen Intern Med. 2002 Jun;17(6):469-76.

26. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008 Apr 26;336(7650):924-6.
27. Leucht S, Davis JM, Engel RR, Kissling W, Kane JM. Definitions of response and remission in schizophrenia: recommendations for their use and their presentation. Acta Psychiatr Scand Suppl. 2009(438):7-14.
28. Samara MT, Spineli LM, Furukawa TA, Engel RR, Davis JM, Salanti G, et al. Imputation of response rates from means and standard deviations in schizophrenia. Schizophr Res. 2013 Nov 18.
29. Harbord RM, Higgins JPT. Meta-regression in STATA. The Stata Journal. 2008;8:493-519.
30. Lundbeck. Synopsis - Study 11915A: A one-year multi-national, multi-centre, randomised, double-blind, parallel-group, fixed-dose bifeprunox study combining a 12-week placebo-controlled, quetiapine-referenced phase with a 12-month quetiapine-controlled phase in patients with schizophrenia.2010.
31. Lundbeck. Synopsis - Study 11916A: A one-year multi-national, multi-centre, randomised, double-blind, parallel-group, fixed-dose bifeprunox study combining a 12-week placebo-controlled, quetiapine-referenced phase with a 12-month quetiapine-controlled phase in patients with schizophrenia.2010.
32. Arvanitis L, Scott M. A Multicenter, Double Blind, Randomized, Controlled, Multiple Fixed Dose and Dose Regimen Comparison of SEROQUELTM (ICI 204,636) and Haloperidol in the Prevention of Psychotic Relapse in Outpatients with Chronic or Subchronic Schizophrenia (5077IL/0015): AstraZeneca1996.
33. Perlis RH, Ostacher M, Fava M, Nierenberg AA, Sachs GS, Rosenbaum JF. Assuring that double-blind is blind. Am J Psychiatry. 2010 Mar;167(3):250-2.
34. Treadwell JR, Uhl S, Tipton K, Shamliyan T, Viswanathan M, Berkman ND, et al. Assessing equivalence and noninferiority. J Clin Epidemiol. 2012 Nov;65(11):1144-9.
35. Spielmans GI, Parry PI. From evidence-based medicine to marketing-based medicine: Evidence from internal industry documents. Journal of Bioethical Inquiry. 2010;7(1):13-29.
36. Potkin SG, Gharabawi GM, Greenspan AJ, Mahmoud R, Kosik-Gonzalez C, Rupnow MF, et al. A double-blind comparison of risperidone, quetiapine and placebo in patients with schizophrenia experiencing an acute exacerbation requiring hospitalization. Schizophr Res. 2006 Jul;85(1-3):254-65.
37. Canuso CM, Dirks B, Carothers J, Kosik-Gonzalez C, Bossie CA, Zhu Y, et al. Randomized, double-blind, placebo-controlled study of paliperidone extended-release and quetiapine in inpatients with recently exacerbated schizophrenia. Am J Psychiatry. 2009 Jun;166(6):691-701.
38. Arvanitis LA, Miller BG. Multiple fixed doses of "Seroquel" (quetiapine) in patients with acute exacerbation of schizophrenia: a comparison with haloperidol and placebo. The Seroquel Trial 13 Study Group. Biol Psychiatry. 1997 Aug 15;42(4):233-46.
39. Lindenmayer JP, Brown D, Liu S, Brecher M, Meulien D. The efficacy and tolerability of once-daily extended release quetiapine fumarate in hospitalized patients with acute schizophrenia: a 6-week randomized, double-blind, placebo-controlled study. Psychopharmacol Bull. 2008;41(3):11-35.
40. Findling RL, McKenna K, Earley WR, Stankowski J, Pathak S. Efficacy and safety of quetiapine in adolescents with schizophrenia investigated in a 6-week, double-blind, placebo-controlled trial. J Child Adolesc Psychopharmacol. 2012 Oct;22(5):327-42.
41. Small JG, Hirsch SR, Arvanitis LA, Miller BG, Link CG. Quetiapine in patients with schizophrenia. A high- and low-dose double-blind comparison with placebo. Seroquel Study Group. Arch Gen Psychiatry. 1997 Jun;54(6):549-57.
42. Borison RL, Arvanitis LA, Miller BG. ICI 204,636, an atypical antipsychotic: efficacy and safety in a multicenter, placebo-controlled trial in patients with schizophrenia. U.S. SEROQUEL Study Group. J Clin Psychopharmacol. 1996 Apr;16(2):158-69.
43. Kahn RS, Schulz SC, Palazov VD, Reyes EB, Brecher M, Svensson O, et al. Efficacy and tolerability of once-daily extended release quetiapine fumarate in acute schizophrenia: a randomized, double-blind, placebo-controlled study. J Clin Psychiatry. 2007 Jun;68(6):832-42.
44. Cutler AJ, Tran-Johnson T, Kalali A, Astrom M, Brecher M, Meulien D. A failed 6-week,randomized, double-blind, placebo-controlled study of once-daily extended release quetiapine fumarate in patients with acute schizophrenia: lessons learned. Psychopharmacol Bull. 2010;43(4):37-69.
45. Chen EY, Hui CL, Lam MM, Chiu CP, Law CW, Chung DW, et al. Maintenance treatment with quetiapine versus discontinuation after one year of treatment in patients with remitted first episode psychosis: randomised controlled trial. BMJ. 2010;341:c4024.
46. Agid O, Siu CO, Potkin SG, Kapur S, Watsky E, Vanderburg D, et al. Meta-Regression Analysis of Placebo Response in Antipsychotic Trials, 1970-2010. Am J Psychiatry. 2013 Jul 30.
47. Leucht S, Heres S, Davis JM. Increasing placebo response in antipsychotic drug trials: let's stop the vicious circle. Am J Psychiatry. 2013 Nov 1;170(11):1232-4.
48. Mallinckrodt CH, Zhang L, Prucka WR, Millen BA. Signal detection and placebo response in schizophrenia: parallels with depression. Psychopharmacol Bull. 2010;43(1):53-72.

49. Hamann J, Kissling W, Leucht S. How long do psychiatrists wait for response before they switch to another antipsychotic? Psychopharmacol Bull. 2007;40(3):149-54.
50. Wieseler B, Kerekes MF, Vervoelgyi V, McGauran N, Kaiser T. Impact of document type on reporting quality of clinical drug trials: a comparison of registry reports, clinical study reports, and journal publications. BMJ. 2012;344:d8141.
51. Hough DW, Natarajan J, Vandebosch A, Rossenu S, Kramer M, Eerdekens M. Evaluation of the effect of paliperidone extended release and quetiapine on corrected QT intervals: a randomized, double-blind, placebo-controlled study. Int Clin Psychopharmacol. 2011 Jan;26(1):25-34.
52. Chapel S, Hutmacher MM, Haig G, Bockbrader H, de Greef R, Preskorn SH, et al. Exposure-response analysis in patients with schizophrenia to assess the effect of asenapine on QTc prolongation. J Clin Pharmacol. 2009 Nov;49(11):1297-308.

Table 1 Included studies and Cochrane risk of bias ratings

	Year published / completed	Primary publication available?	Clinical study report synopsis or extract available?	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Performance bias (blinding of participants and personnel)	Detection bias (blinding of assessments)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias (not including financial conflict of interest of sponsor or researcher)
Arvanitis	1997	Yes (38)	Yes[a]	Unclear	Unclear	Unclear	Unclear	High	Unclear	High
Small	1997	Yes (41)	Yes[a]	Low	Unclear	Unclear	Unclear	High	Unclear	High
Borison	1996	Yes (42)	Yes[a]	Unclear	Unclear	Unclear	Unclear	High	High	High
Kahn	2007	Yes (43)	Yes	Unclear	Unclear	Unclear	Unclear	High	High	Unclear
Canuso	2009	Yes (37)	Yes[b]	Low	Low	Unclear	Unclear	Low	Unclear	Unclear
Potkin	2006	Yes (36)	No	Low	Low	Unclear	Unclear	Low	High	Unclear
Chen	2010	Yes (45)	No	Low	Low	Unclear	Unclear	High	High	High
Lindenmayer	2008	Yes (39)	Yes	Unclear	Unclear	Unclear	Unclear	High	High	Unclear
Cutler	2010	Yes (44)	Yes	Unclear	Unclear	Unclear	Unclear	High	High	Unclear
Hough	2011	Yes (51)	Yes	Low	Low	Unclear	High	Low	High	Unclear
Chapel	2009	Yes (52)	No	Unclear	Unclear	Unclear	High	Unclear	High	Unclear
Findling	2007	Yes (40)	Yes	Low	Low	Unclear	Unclear	High	Low	Low
Study 11915A	2009	No	Yes (30)	Unclear	Unclear	Unclear	Unclear	High	High	High
Study 11916A	2009	No	Yes (31)	Unclear	Unclear	Unclear	Unclear	High	High	High
Study 15	1995	No	Partially (32)	Unclear	Unclear	Low	Low	High	High	High

[a] Astrazeneca supplied extract; [b] Pfizer supplied extract; Note: see supplementary file for detailed ratings for each reviewer, and for ratings from published reviews (where available).

Table 2 Primary efficacy outcomes

Outcome	Time-point (weeks)	No of included studies	Que N events / N	Pla N events / N	Hedges's g (95% CI)	Mean difference (95% CI)	Risk ratio (95% CI)	Absolute difference (95% CI)	NNTB/H (95% CI)	Heterogeneity for g or RR: I^2; Chi^2 (p-value)	Quality (GRADE)
Overall symptoms (mean change in PANSS total) based on LOCF or MMRM	2-12	11	1346	912	-0.33 (-0.44, -0.21)	-6.44 (-8.89, -4.00)	-	-	-	47%; 18.9 (p=0.040)	Moderate-high
Overall symptoms (mean change in PANSS total) using Strategy 1 imputations.	2-12	11	1373	931	-0.23 (-0.35, -0.11)	-4.25 (-6.46, -2.04)	-	-	-	52%; 20.7 (p=0.023)	
Overall symptoms (mean change in PANSS total) using Strategy 2 imputations	2-12	11	1373	931	-0.15 (-0.30, 0.01)	-2.66 (-5.46, 0.15)	-	-	-	70%; 32.9 (p<0.001)	
Significant improvement (≥50% reduction in PANSS / BPRS) based on LOCF	2-12	11	1126 / 1375	816 / 933	-	-	0.95 (0.91, 0.98)	-0.047 (-0.016, -0.016)	21B (13B, 63B)	43%; 17.5 (p=0.070)	Low - moderate

Shaded rows indicates result statistically significant at p<0.05

Table 3 Main safety outcomes

Outcome (definition, imputation strategy)	Time-point (weeks)	No of included studies	Que N events / N	Pla N events / N	Hedges's g (95% CI)	Mean difference (95% CI)	Risk ratio (95% CI)	Absolute difference (95% CI)	NNTB/H (95% CI)	Heterogeneity for g or RR: I^2; Chi^2 (p-value)	Quality (GRADE)
Serious adverse events	2-12	8	50/851	45/658	-	-	0.94 (0.64, 1.39)	-0.001 (-0.023, 0.021)	1000B (44B, 48H)	0%; 3.3 (p=0.885)	Very low
Any adverse event	2-12	9	754/1112	438/756	-	-	1.14 (1.06, 1.22)	0.089 (0.045, 0.134)	11H (22H, 8H)	0%; 6.6 (p=0.583)	Low
Simpson-Angus Scale (worsening)	6	7	128/869	83/596	-	-	0.97 (0.73, 1.29)	0.007 (-0.033, 0.047)	143H (30B, 21H)	15%; 7 (p=0.317)	Low
Abnormal Involuntary Movement Scale (worsening)	6	4	88/534	56/265	-	-	0.694 (0.521, 0.924)	-0.047 (-0.130, 0.037)	21B (8B, 27H)	0%; 2.6 (0.460)	Low
Barnes Akathesia Scale (worsening)	2-6	7	67/973	49/643	-	-	0.866 (0.609, 1.234)	-0.005 (-0.030, 0.020)	200B (33B, 50H)	0%l 3.5 (p=0.745)	Low
Needing medication for extra-pyramidal side-effects	2-6	9	97/1071	65/698	-	-	0.838 (0.597, 1.176)	0.004 (-0.025, 0.032)	250H (40B, 31H)	12%; 9.1 (p=0.334)	Moderate
Mean weight change	2-12	12	1410	948	0.640 (0.428, 0.852)	1.753kg (1.104, 2.402)	-	-	-	83%; 65.4 (p<0.001)	Very low
Significant weight-gain (≥7% or recorded as adverse effect)	2-12	10	140/1220	32/863	-	-	2.988 (2.048, 4.362)	0.076 (0.044, 0.109)	13H (23H, 9H)	0%; 6.9 (p=0.648)	Moderate
Sedation or somnolence	2-12	12	247/1419	57/958	-	-	2.818 (1.963, 4.047)	0.115 (0.078, 0.151)	9H (7H, 13H)	31%; 16.1 (p=0.138)	Moderate

Outcome (definition, imputation strategy)	**Time-point (weeks)**	**No of included studies**	**Que N events / N**	**Pla N events / N**	**Hedges's g (95% CI)**	**Mean difference (95% CI)**	**Risk ratio (95% CI)**	**Absolute difference (95% CI)**	**NNTB/H (95% CI)**	**Heterogeneity for g or RR: I^2; Chi^2 (p-value)**	**Quality (GRADE)**
Leaving early due to adverse effects	2-12	11	97/1263	74/885	-		1.009 (0.753, 1.351)	0.010 (-0.010, 0.031)	100H (100B, 32H)	0%; 9.4 (p=0.495)	Moderate

Shaded rows indicates result statistically significant at $p<0.05$; *indicates at least moderate heterogeneity ($I^2 \geq 40\%$); bold text indicates primary outcome

Figure 1 PRISMA Diagram

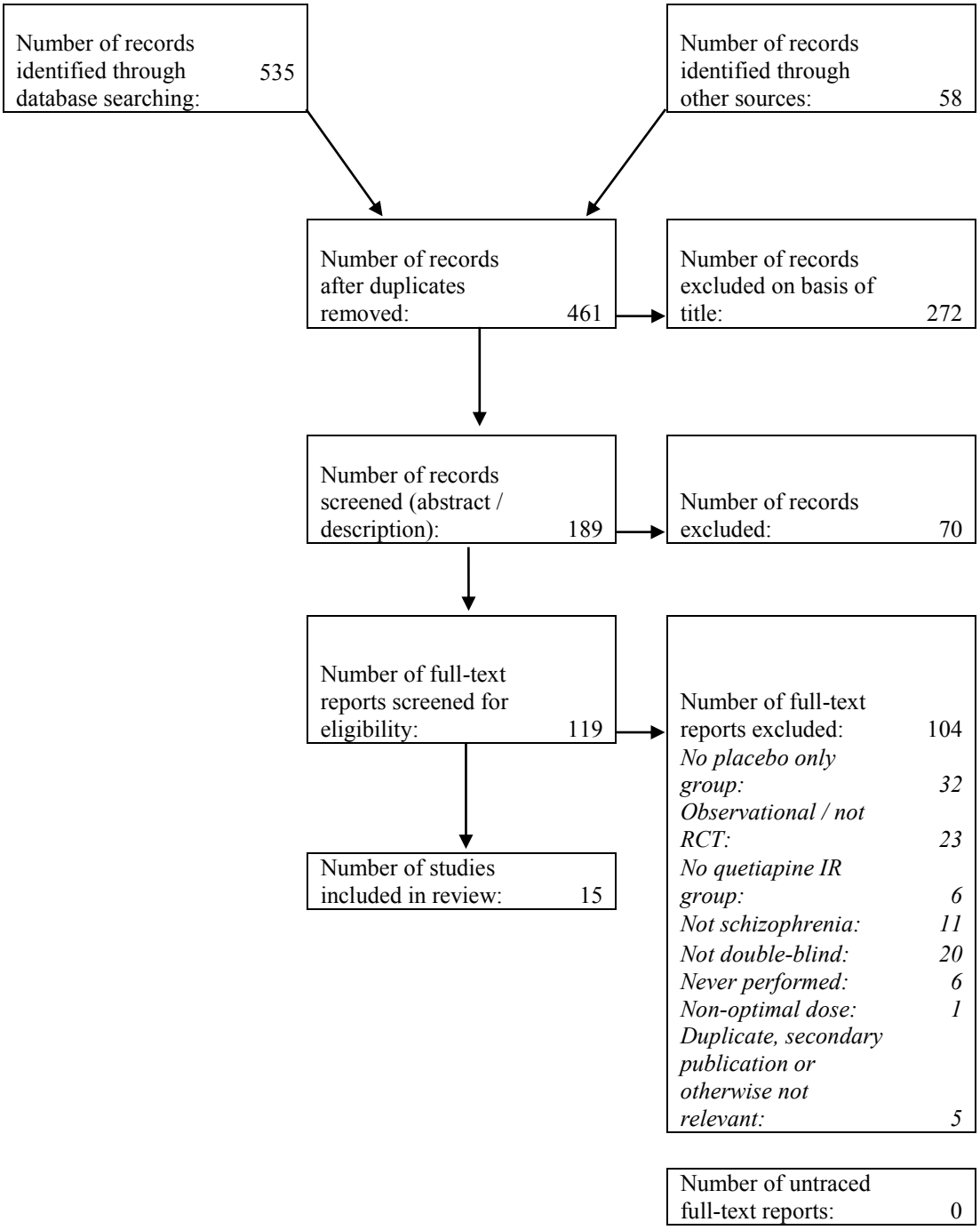


Figure 2 Mean change in PANSS total scores or equivalent, using mostly original LOCF estimates to impute missing data (*indicates severe attrition, ≥50%)

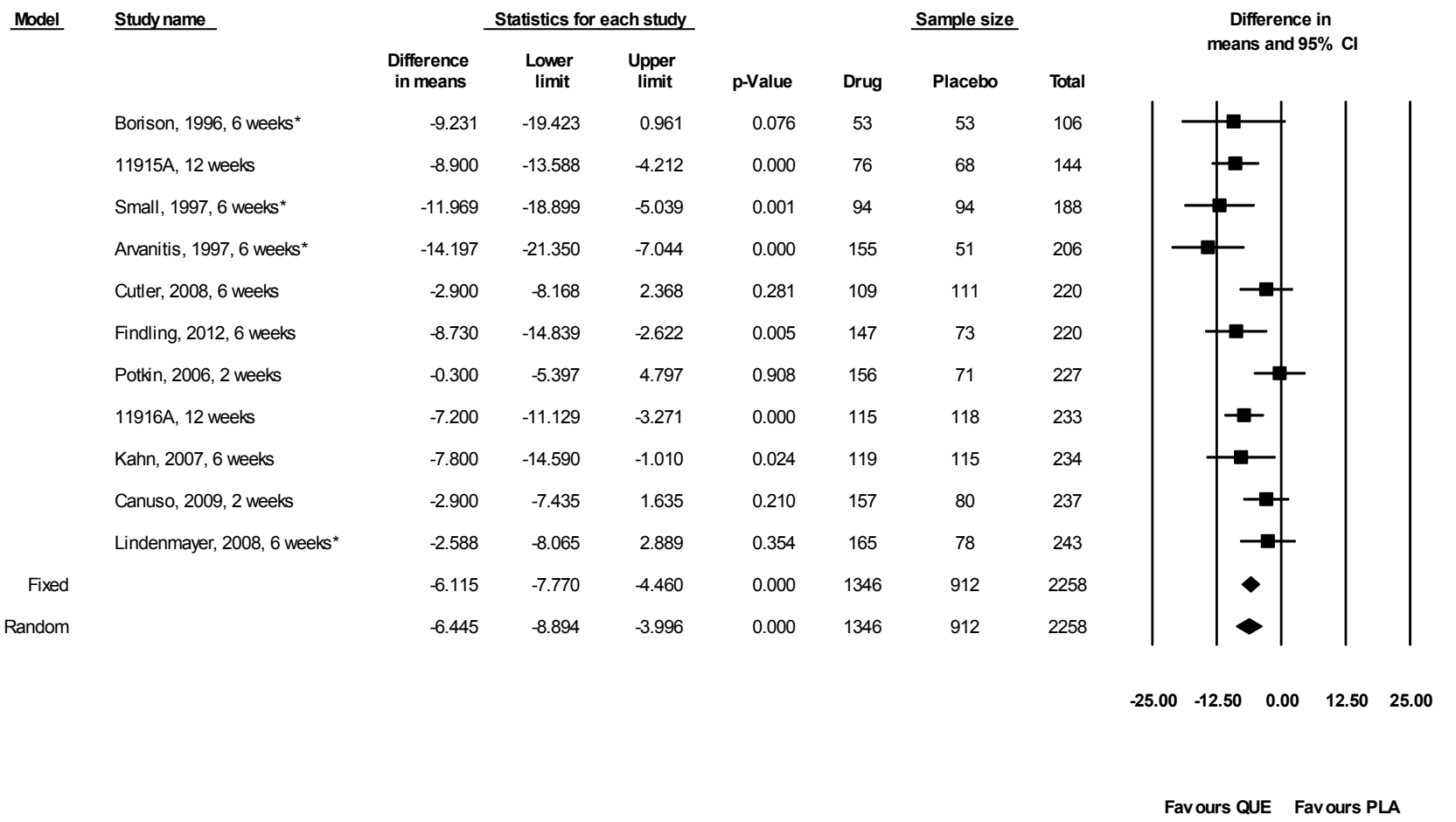


Figure 3. Relative risk of not achieving at least 50% reduction in PANSS total scores or equivalent, based on mostly LOCF data (F = estimated using Furukawa method; *indicates severe attrition ≥50%)

Model	Study name	Statistics for each study				Events			Sample size			Risk ratio and 95% CI
		Risk ratio	Lower limit	Upper limit	p-Value	Drug	Placebo	Total	Drug	Placebo	Total	
	Borison, 1996, 6 weeks, 50% PANSS - F*	0.912	0.771	1.080	0.286	43	48	91	54	55	109	
	11915A 12 weeks 50% PANSS - F	0.895	0.805	0.994	0.039	65	65	130	76	68	144	
	Small, 1997, 6 weeks, 50% PANSS - F*	0.827	0.689	0.992	0.040	62	75	137	96	96	192	
	Arvanitis, 1997, 6 weeks 60% BPRS*	0.877	0.814	0.945	0.001	135	50	185	157	51	208	
	Findling, 2012, 6 weeks 50% PANSS - F	0.869	0.716	1.054	0.155	91	52	143	147	73	220	
	Potkin, 2006, 2 weeks, 50% PANSS - F	0.986	0.845	1.151	0.859	118	56	174	156	73	229	
	Cutler 2010, 6 weeks 50% PANSS	1.000	0.959	1.042	0.992	113	114	227	116	117	233	
	11916A 12 weeks 50% PANSS - F	0.986	0.892	1.091	0.787	99	103	202	115	118	233	
	Canuso 2009 50% PANSS	0.976	0.860	1.108	0.705	128	66	194	159	80	239	
	Kahn 2007, 50% PANSS	0.932	0.841	1.032	0.177	102	105	207	123	118	241	
	Lindenmayer 2008, 6 weeks 50% PANSS*	0.989	0.947	1.033	0.633	170	82	252	176	84	260	
Fixed		0.965	0.943	0.988	0.004	1126	816	1942	1375	933	2308	
Random		0.948	0.913	0.984	0.005	1126	816	1942	1375	933	2308	

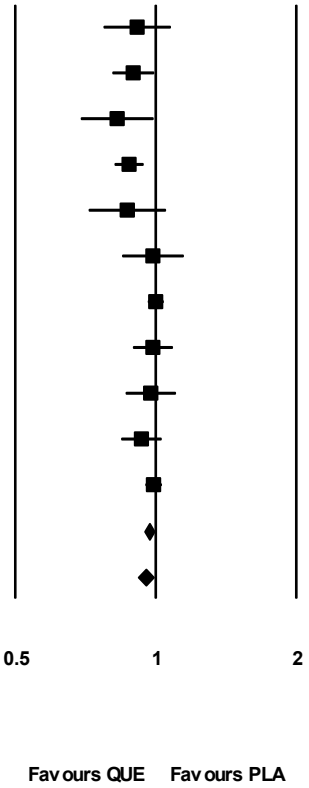
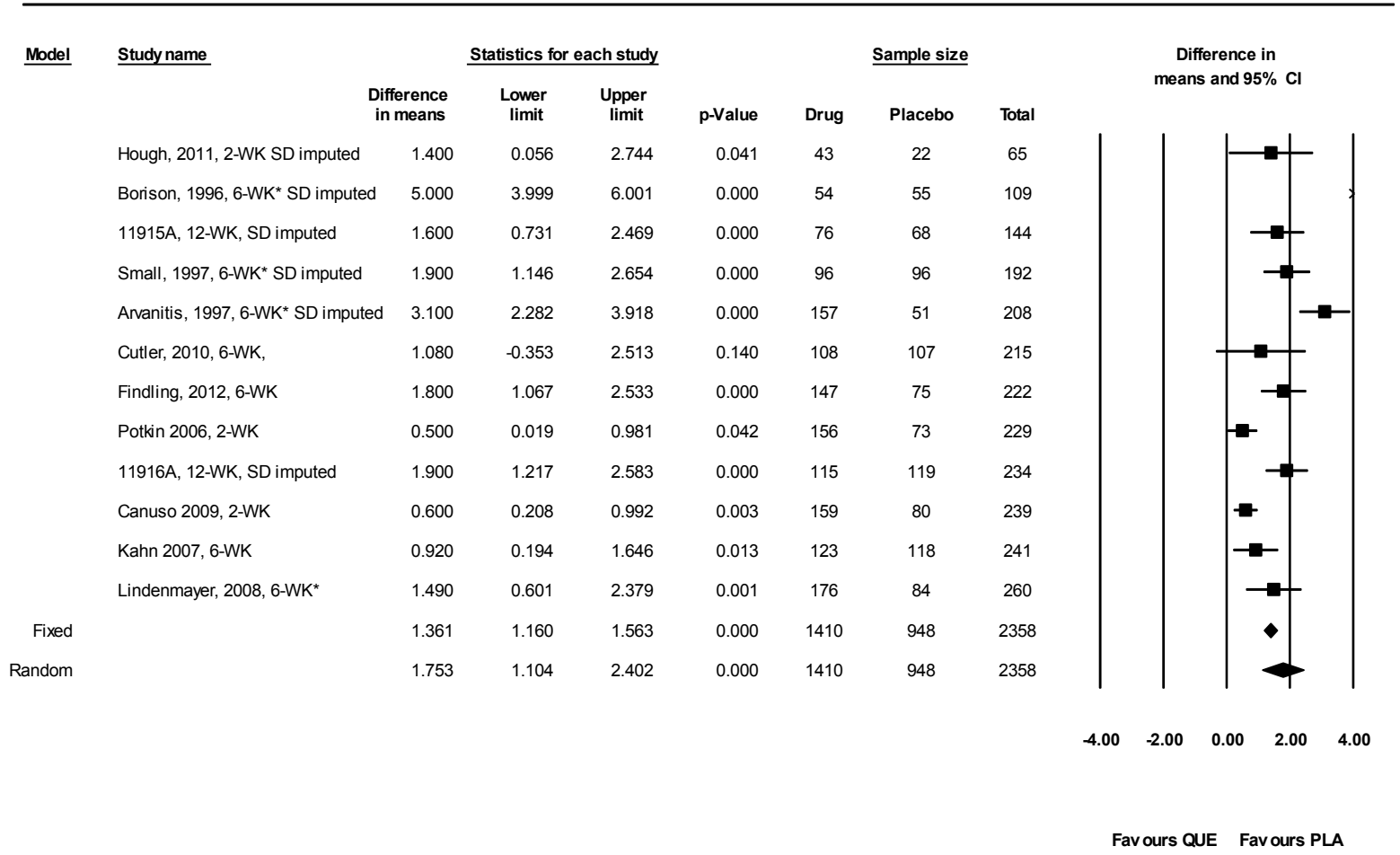


Figure 4. Mean weight-gain (kg; *indicates severe attrition, ≥50%)



Model	Study name	Difference in means	Lower limit	Upper limit	p-Value	Drug	Placebo	Total
	Hough, 2011, 2-WK SD imputed	1.400	0.056	2.744	0.041	43	22	65
	Borison, 1996, 6-WK* SD imputed	5.000	3.999	6.001	0.000	54	55	109
	11915A, 12-WK, SD imputed	1.600	0.731	2.469	0.000	76	68	144
	Small, 1997, 6-WK* SD imputed	1.900	1.146	2.654	0.000	96	96	192
	Arvanitis, 1997, 6-WK* SD imputed	3.100	2.282	3.918	0.000	157	51	208
	Cutler, 2010, 6-WK,	1.080	-0.353	2.513	0.140	108	107	215
	Findling, 2012, 6-WK	1.800	1.067	2.533	0.000	147	75	222
	Potkin 2006, 2-WK	0.500	0.019	0.981	0.042	156	73	229
	11916A, 12-WK, SD imputed	1.900	1.217	2.583	0.000	115	119	234
	Canuso 2009, 2-WK	0.600	0.208	0.992	0.003	159	80	239
	Kahn 2007, 6-WK	0.920	0.194	1.646	0.013	123	118	241
	Lindenmayer, 2008, 6-WK*	1.490	0.601	2.379	0.001	176	84	260
Fixed		1.361	1.160	1.563	0.000	1410	948	2358
Random		1.753	1.104	2.402	0.000	1410	948	2358

Figure 5. Relative risk of clinically significant weight-gain (normally ≥7% increase) (*indicates severe attrition, ≥50%)

Model	Study name	Statistics for each study				Events			Sample size		
		Risk ratio	Lower limit	Upper limit	p-Value	Drug	Placebo	Total	Drug	Placebo	Total
	Borison, 1996, 6-WK*	7.130	1.701	29.889	0.007	14	2	16	54	55	109
	11915A, 12-WK	11.649	0.668	203.016	0.092	6	0	6	76	68	144
	Small, 1997, 6-WK*	4.800	1.911	12.057	0.001	24	5	29	96	96	192
	Arvanitis, 1997, 6-WK*	2.166	0.671	6.989	0.196	20	3	23	157	51	208
	Findling, 2012, 6-WK	3.912	1.213	12.610	0.022	23	3	26	147	75	222
	Cutler, 2010, 6-WK	2.185	0.860	5.553	0.100	13	6	19	116	117	233
	11916A, 12-WK	8.207	1.043	64.589	0.046	8	1	9	116	119	235
	Canuso 2009, 2-WK	2.516	0.299	21.173	0.396	5	1	6	159	80	239
	Kahn 2007, 6-WK	1.812	0.841	3.904	0.129	17	9	26	123	118	241
	Lindenmayer, 2008, 6-WK*	2.386	0.535	10.650	0.254	10	2	12	176	84	260
Fixed		2.988	2.048	4.362	0.000	140	32	172	1220	863	2083
Random		2.988	2.048	4.362	0.000	140	32	172	1220	863	2083

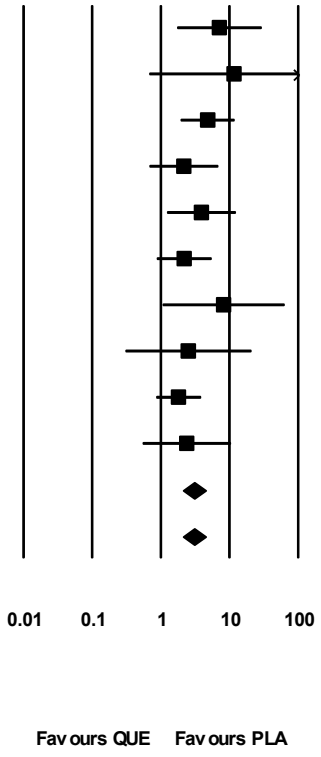


Figure 6. Relative risk of somnolence or sedation (*indicates severe attrition, ≥50%)