Quetiapine immediate release v. placebo for schizophrenia: systematic review, meta-analysis and reappraisal

Paul Hutton, Peter J. Taylor, Lee Mulligan, Sarah Tully and Joanna Moncrieff

**Background**
Immediate-release (IR) quetiapine has been used to treat schizophrenia since 1997, although all the principal placebo-controlled trials have >50% missing outcome data. New studies with relatively lower rates of participant withdrawal have since been published.

**Aims**
To assess the efficacy and adverse effects of quetiapine IR for schizophrenia, with consideration of outcome quality and clinical meaningfulness of results, and 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 and placebo (or subtherapeutic dose in relapse prevention trials) for the treatment of schizophrenia (PROSPERO registration CRD4201100165). Primary outcomes were change in overall symptoms and response rates. We also examined whether high rates of participant withdrawal (≥50%) attenuated effect sizes, and assessed the impact of making different assumptions about these people’s outcomes.

**Results**
We identified 15 relevant trials (including 2 unpublished), providing the first 12-week data for this drug and the first data on self-reported quality of life. We found quetiapine IR to have a weighted mean difference (WMD) of 6.5 points (95% CI –8.9 to –4) on Positive and Negative Syndrome Scale (PANSS) total scores, which corresponds to a standardised mean difference (SMD) of –0.33 (95% CI –0.46 to –0.21). Longer trials reported larger mean differences favouring quetiapine IR, but the overall estimate was smaller if more conservative assumptions about the outcomes of people who left the trial early were made. Approximately 21 people needed to take quetiapine IR for 1 person to experience at least a 50% improvement in PANSS score. No difference in quality of life was observed (two RCTs), although small to moderate improvements in social functioning were found (three RCTs). Quetiapine IR caused sedation and increased rates of clinically significant weight gain, but no extrapyramidal effects were observed.

**Conclusions**
Quetiapine IR has a small beneficial effect on overall psychotic symptoms over 2–12 weeks, but also leads to weight gain and sedation.

**Declaration of interest**
None.

**Copyright and usage**
© The Royal College of Psychiatrists 2015.
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, and we used a recently published approach to examine the impact on our efficacy estimates of changing assumptions about the likely outcomes of the large numbers of people who leave these trials early.

### Method

Our search strategy and protocol detailing our inclusion and exclusion criteria are provided in an online data supplement. Two researchers independently searched publication databases, clinical trial registries and previous reviews for randomised controlled trials in which participants with a diagnosis of schizophrenia 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 at study end-point in total score on the PANSS or, if that were not available, on the Brief Psychiatric Rating Scale (BPRS), and the numbers of people achieving an important clinical response. We defined the latter as a greater than 50% reduction in PANSS or BPRS score. When these were not reported or provided, we imputed them from means and standard deviations using the validated method of Furukawa et al. (See 'Changes from protocol' section in the online supplement.) Our secondary efficacy outcomes included relapse, positive symptoms, negative symptoms, depression, quality of life and need for additional antipsychotic medication or sedatives. We also examined the numbers of participants leaving the study early for any reason, need for hospital care and functioning. For adverse effects, we looked at use of anti-Parkinsonian medication, extrapyramidal side-effects, withdrawal 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 participants leaving the study early 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 (MRRM) imputation method, followed by LOCF or observed case data if not available. Missing standard deviations were, where possible, calculated from t-test values, P-values, standard errors or confidence intervals. If no variance parameters were reported for a particular study, we imputed standard deviations using the medians of the other studies. Similarly to previous studies, we planned to use data from study arms where participants received an optimal drug dose of more than 250 mg. However, we carried out a sensitivity analysis excluding doses of more than 400 mg, as per the recent International Consensus on Antipsychotic Dosing and recent Leucht group analysis.

### Meta-analytic calculations

For continuous data we calculated the Hedges’ g standardised mean difference (SMD) using Comprehensive Meta-Analysis version 2 for Windows 7. For the primary analysis of the 2–12 week study end-point data we converted BPRS scores (mean and s.d.) to PANSS scores using recently published conversion charts (PANSS total score = 1.538 × BPRS total score),20 thus allowing us to present also the unstandardised 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.21 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 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 I² statistic of ≥40%.22

### Impact of missing data

We tested the hypothesis that trials with severe rates of missing data (≥50% at end-point) 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 carry more weight than actual observations, and because the National Institute for Health and Care Excellence (NICE), the Cochrane Schizophrenia Group and others often exclude trials with this degree of missing data from their reviews.23 We also wished to compare studies with <25% and ≥25% attrition at end-point, but were unable to do so because no study of 6–12 weeks’ duration had less than 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 et al.24 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 standard deviations with the medians of the control arms of all the included trials, as recommended.

### Analysis of clinical significance

The minimum clinically important difference (MCID) has been defined by Jaeschke et al 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’.25,26 An analysis of data from 14 antipsychotic trials (n = 5970) found a rater-determined MCID on the PANSS of roughly 15 points, thus allowing us to present also the unstandardised 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. 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 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 I² statistic of ≥40%.22
We tested the validity of these definitions by comparing them with the median of mean changes that 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 was important to detect.15

**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 Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.26 Further details on method and ratings are provided in the online supplement. We tested for publication bias using funnel plots for the PANSS/BPRS total score effect sizes (Hedges’ g) of all studies. Ratings of bias and quality were used to inform interpretation of reliability and magnitude of effects.

**Registration of review protocol and subsequent changes**

The review protocol was registered in advance with the International Prospective Register of Systematic Reviews (PROSPERO), protocol CRD4201100165. Subsequent changes, in addition to those outlined above, are detailed in the online supplement. 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 score). This criterion is now recommended for use in studies of acutely ill patients with non-refractory illness,27 and we used the method of Furukawa et al 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 study duration (measured in weeks) and 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**

The process of selecting studies is detailed in Fig. 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 owing to the 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 subtherapeutic doses of quetiapine,32 arguing that 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 partly to assess risk of bias. We therefore included data from a total of 15 studies. An overview of included studies is provided in Table 1, and excluded studies are listed in the online supplement, together with a table of trial characteristics and baseline demographic data.

**Risk of bias and quality ratings**

Table 1 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 online supplement. Our rationale for the ratings is also provided online, 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 (e.g. adverse effects) where the desired outcome is ‘no difference’.34 There is also evidence from documents released through legal proceedings in the USA that AstraZeneca have historically not published all...
active-comparator quetiapine trials, or have not reported all outcomes.35

Validation of MCID criterion

Researchers and trial sponsors designed their trials to detect with adequate power a mean change in PANSS total score or equivalent of approximately 12 points (range 9.0–15.5), which corresponds to an SMD of 0.55, and is similar in magnitude to the empirically derived estimate of MCID of 11–15 points.10–12

Primary efficacy outcomes

Moderate- to high-quality evidence suggested that quetiapine IR was statistically superior to placebo from 2 weeks to 12 weeks in terms of reducing overall symptoms, but the effect was small (WMD = 7.7 points, 95% CI −9.9 to −2.3; SMD = −0.19, 95% CI −0.43 to −0.12) and an NNT of 13 (95% CI 7 to 250). The three 6-week studies with less than 50% attrition had a mean advantage of 6.1 points (95% CI −9.9 to −2.3; SMD = −0.27, 95% CI −0.43 to −0.12) and a non-significant NNT of 35 (95% CI 12 to 42H).10–12

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 −8.0 to −2.9; SMD = −0.29, 95% CI −0.44 to −0.15) and an NNT of 42 (19, 250H; H = harm), whereas the four 6-week studies with 50% or more attrition had a mean PANSS advantage of 9.2 points (95% CI −8.9 to −3.6; SMD = −0.32, 95% CI −0.45 to −0.19; NNT = 22, 95% CI 13 to 83).40

Strategy 1 of the Ebrahim approach involved testing whether the overall results would be different if we assumed that participants who withdrew early from both groups had the same degree of change as participants in the control group who stayed until the end.13 To illustrate this, consider the study by Small et al.41 Here the mean change for the 49 quetiapine and 39 placebo group participants who completed the trial was, after conversion of BPRS to PANSS scores, −23.3 points (s.d. = 17.7) and −14.9 points (s.d. = 17.7) respectively – a between-group difference of around 8.4 points. Carrying forward the last available scores of the 104 people who did not complete this trial reduced the quetiapine estimate to −13.5 points (s.d. = 24.5; n = 94) and the placebo estimate to −1.5 points (s.d. = 24.0; n = 94), and increased the
date. Meta-regression bubble plots are provided in the online supplement.

Removing data from arms employing doses smaller than 400 mg reduced the contribution made by two trials36,39 but this had little effect on the overall estimate of average change (WMD = 6.3, 95% CI −8.7 to −3.8; SMD = −0.32, 95% CI −0.44 to −0.20) or response rates (NNT = 20, 95% CI 13 to 53). Excluding the study with an adolescent sample (Findling et al) also had little effect on estimates (WMD = −6.3, 95% CI −8.9 to −3.6; SMD = −0.32, 95% CI −0.45 to −0.19; NNT = 22, 95% CI 13 to 83).40

Table 1 included studies and Cochrane risk of bias ratings

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>n</th>
<th>Publication</th>
<th>Clinical report</th>
<th>Random sequence generation</th>
<th>Allocation concealment</th>
<th>Performance bias (masking of participants and personnel)</th>
<th>Detection bias (masking of assessments)</th>
<th>Incomplete outcome data (attrition bias)</th>
<th>Selective reporting (reporting bias)</th>
<th>Other bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arvanitis &amp; Miller</td>
<td>1997</td>
<td>Yes</td>
<td>Yes38</td>
<td>Yes3</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>Unclear</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Small et al</td>
<td>1997</td>
<td>Yes31</td>
<td>Yes39</td>
<td>Yes9</td>
<td>Low</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Borison et al</td>
<td>1996</td>
<td>Yes3</td>
<td>Yes3</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Kahn et al</td>
<td>2007</td>
<td>Yes3</td>
<td>Yes3</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low</td>
<td>Unclear</td>
<td>Low</td>
<td>Unclear</td>
<td></td>
</tr>
<tr>
<td>Canuso et al</td>
<td>2009</td>
<td>Yes3</td>
<td>Yes9</td>
<td>Low</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low</td>
<td>Unclear</td>
<td>Low</td>
<td>Unclear</td>
<td></td>
</tr>
<tr>
<td>Potkin et al</td>
<td>2006</td>
<td>Yes3</td>
<td>No</td>
<td>Low</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>Low</td>
<td>Unclear</td>
<td></td>
</tr>
<tr>
<td>Chen et al</td>
<td>2010</td>
<td>Yes3</td>
<td>No</td>
<td>Low</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Lindenmayer et al</td>
<td>2008</td>
<td>Yes3</td>
<td>Yes3</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Cutler et al</td>
<td>2010</td>
<td>Yes3</td>
<td>Yes3</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Hough et al</td>
<td>2011</td>
<td>Yes3</td>
<td>Yes</td>
<td>Low</td>
<td>Unclear</td>
<td>High</td>
<td>Low</td>
<td>Unclear</td>
<td>High</td>
<td>Unclear</td>
<td></td>
</tr>
<tr>
<td>Chapel et al</td>
<td>2009</td>
<td>Yes3</td>
<td>No</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>Unclear</td>
<td>High</td>
<td>Uncler</td>
<td></td>
</tr>
<tr>
<td>Fingding et al</td>
<td>2012</td>
<td>Yes3</td>
<td>Yes9</td>
<td>Low</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low</td>
<td>Unclear</td>
<td>High</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Study 11915A</td>
<td>2009</td>
<td>No</td>
<td>Yes3</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Study 11916A</td>
<td>2009</td>
<td>No</td>
<td>Yes3</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Arvanitis &amp; Scott (Study 15)</td>
<td>1995</td>
<td>No</td>
<td>Partially32</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
</tbody>
</table>

a. Not including financial conflict of interest of sponsor or researcher.

b. AstraZeneca supplied extract.
c. Pfizer supplied extract.
### Table 2 Primary efficacy outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Time (weeks)</th>
<th>No. of included studies</th>
<th>Quetiapine events/n</th>
<th>Placebo events/n</th>
<th>Hedges' g (95% CI)</th>
<th>Difference Mean (95% CI)</th>
<th>Risk ratio (95% CI)</th>
<th>Absolute difference (95% CI)</th>
<th>NNTB/H (95% CI)</th>
<th>Heterogeneity</th>
<th>Quality (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall symptoms (mean change in PANSS total score) based on LOCF or MMRM</td>
<td>2–12</td>
<td>11</td>
<td>1346</td>
<td>912</td>
<td>−0.33 (−0.44, −0.21)</td>
<td>−6.44 (−8.89, −4.00)*</td>
<td>0.047 (0.001, 0.097)</td>
<td>21B (13B, 63B)*</td>
<td>0.005 (0.000, 0.010)*</td>
<td>100H (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
</tr>
<tr>
<td>Overall symptoms (mean change in PANSS total score) using strategy 1 imputations</td>
<td>2–12</td>
<td>11</td>
<td>1373</td>
<td>931</td>
<td>−0.23 (−0.35, −0.11)</td>
<td>−4.25 (−6.46, −2.04)*</td>
<td>0.115 (0.078, 0.151)</td>
<td>9H (7H, 13H)*</td>
<td>0.013 (0.012, 0.014)</td>
<td>100H (1008, 314)*</td>
<td>0.001 (0.000, 0.002)</td>
</tr>
<tr>
<td>Overall symptoms (mean change in PANSS total score) using strategy 2 imputations</td>
<td>2–12</td>
<td>11</td>
<td>1373</td>
<td>931</td>
<td>−0.15 (−0.30, 0.01)</td>
<td>−2.66 (−5.46, 0.15)</td>
<td>0.016 (0.015, 0.018)</td>
<td>250H (40B, 31H)</td>
<td>0.004 (0.003, 0.005)</td>
<td>9H (7H, 13H)*</td>
<td>0.001 (0.000, 0.002)</td>
</tr>
<tr>
<td>Significant improvement (≥50% reduction in PANSS/BPRS score) based on LOCF</td>
<td>2–12</td>
<td>11</td>
<td>1126/1375</td>
<td>816/933</td>
<td>0.95</td>
<td>−0.047</td>
<td>21B</td>
<td>0.023 (0.016, 0.016)</td>
<td>13B (63B)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>Low to moderate</td>
</tr>
</tbody>
</table>

BPRS, Brief Psychiatric Rating Scale; LOCF, last observation carried forward; MMRM, mixed-models repeated measures; NNTB/H, number needed to treat (benefit/harm); PANSS, Positive and Negative Syndrome Scale; RR, relative risk.

*P < 0.05.

### Table 3 Main safety outcomes

<table>
<thead>
<tr>
<th>Outcome (definition, imputation strategy)</th>
<th>Time (weeks)</th>
<th>No. of included studies</th>
<th>Quetiapine events/n</th>
<th>Placebo events/n</th>
<th>Hedges’ g (95% CI)</th>
<th>Difference Mean (95% CI)</th>
<th>Risk ratio (95% CI)</th>
<th>Absolute difference (95% CI)</th>
<th>NNTB/H (95% CI)</th>
<th>Heterogeneity</th>
<th>Quality (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious adverse event</td>
<td>2–12</td>
<td>8</td>
<td>50/851</td>
<td>45/658</td>
<td>0.94 (0.64, 1.39)</td>
<td>−0.001 (−0.023, 0.021)</td>
<td>11H</td>
<td>1000B (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>0.001 (0.000, 0.002)</td>
<td>Very low</td>
</tr>
<tr>
<td>Any adverse event</td>
<td>2–12</td>
<td>9</td>
<td>754/1112</td>
<td>438/756</td>
<td>1.14 (1.06, 1.22)*</td>
<td>0.089 (0.045, 0.134)*</td>
<td>11H</td>
<td>1000B (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>0.001 (0.000, 0.002)</td>
<td>Low</td>
</tr>
<tr>
<td>Simpson–Angus Scale (worsening)</td>
<td>6–7</td>
<td>7</td>
<td>128/869</td>
<td>83/596</td>
<td>0.97 (0.73, 1.29)</td>
<td>0.007 (−0.033, 0.047)</td>
<td>143H</td>
<td>1000B (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>0.001 (0.000, 0.002)</td>
<td>Low</td>
</tr>
<tr>
<td>Abnormal Involuntary Movement Scale (worsening)</td>
<td>2–6</td>
<td>7</td>
<td>67/973</td>
<td>49/643</td>
<td>0.866 (0.609, 1.234)</td>
<td>−0.005 (−0.030, 0.020)</td>
<td>200B</td>
<td>1000B (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>0.001 (0.000, 0.002)</td>
<td>Low</td>
</tr>
<tr>
<td>Barnes Akathisia Rating Scale (worsening)</td>
<td>2–6</td>
<td>7</td>
<td>97/1071</td>
<td>65/698</td>
<td>0.838 (0.597, 1.176)</td>
<td>0.004 (−0.025, 0.032)</td>
<td>250H</td>
<td>1000B (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>0.001 (0.000, 0.002)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Needing medication for extrapyramidal side-effects</td>
<td>2–6</td>
<td>9</td>
<td>97/1071</td>
<td>65/698</td>
<td>0.838 (0.597, 1.176)</td>
<td>0.004 (−0.025, 0.032)</td>
<td>250H</td>
<td>1000B (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>0.001 (0.000, 0.002)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Mean weight change, kg</td>
<td>2–12</td>
<td>12</td>
<td>1410</td>
<td>948</td>
<td>0.640 (0.428, 0.852)*</td>
<td>1.753 (1.014, 2.402)*</td>
<td>13H</td>
<td>1000B (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>0.001 (0.000, 0.002)</td>
<td>Very low</td>
</tr>
<tr>
<td>Significant weight-gain (≥7% or recorded as adverse effect)</td>
<td>2–12</td>
<td>10</td>
<td>140/1220</td>
<td>32/863</td>
<td>2.988 (2.048, 4.362)*</td>
<td>0.076 (0.044, 0.109)*</td>
<td>13H</td>
<td>1000B (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>0.001 (0.000, 0.002)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Sedation or somnolence</td>
<td>2–12</td>
<td>12</td>
<td>247/1419</td>
<td>57/958</td>
<td>2.818 (1.963, 4.047)*</td>
<td>0.115 (0.078, 0.151)*</td>
<td>9H (7H, 13H)*</td>
<td>1000B (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>0.001 (0.000, 0.002)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Leaving early owing to adverse effects</td>
<td>2–12</td>
<td>11</td>
<td>97/1263</td>
<td>74/885</td>
<td>1.009 (0.753, 1.351)</td>
<td>0.010 (−0.010, 0.031)</td>
<td>100H</td>
<td>1000B (448, 481)*</td>
<td>0.001 (0.000, 0.002)</td>
<td>0.001 (0.000, 0.002)</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

GRADE, Grading of Recommendations Assessment, Development and Evaluation; NNTB/H, number needed to treat (benefit/harm); RR, risk ratio.

*P < 0.05.
between-group difference to around 12 points. These are the figures we used in the main analysis. Introducing the strategy 1 assumption that those who did not complete the trial had a similar outcome to those in the placebo group who did complete it (−14.9 points) reduced the overall estimate for the quetiapine group to −19.1 points (s.d. = 18.2) and reduced the advantage over placebo to 4.2 points. We repeated this procedure for the other five trials for which we had completer data and where no usable MMRM estimate was provided, and entered the revised estimates into the overall meta-analysis. Table 2 shows that the overall advantage for quetiapine over placebo fell to 4.3 points (95% CI: −6.5 to −2.0; SMD = −0.23, 95% CI: −0.35 to −0.11).
Strategy 2 of the Ebrahim approach involved testing whether the overall results were robust to assuming, first, that participants in the quetiapine non-completers group had the smallest treatment response observed, and second, that those in the placebo non-completers group had the largest placebo response observed. In the study by Small et al this involved assuming the 47 people in the quetiapine non-completers group had the same degree of response as the quetiapine completers group in the study by Lindenmayer et al (-17.4 points),32 and that the 57 people in the placebo non-completers group had the same degree of response as the placebo completers group in the 2007 study by Kahn et al (-23.1 points).43 The revised quetiapine and placebo estimates were -20.4 (s.d. = 18.2) and -19.8 points (s.d. = 18.3) respectively, leading to a between-group difference of 0.6 points. As shown in Table 2, applying strategy 2 to the six trials for which we had completer data reduced the overall advantage for quetiapine to 2.7 points (95% CI 5.5 to 0.2; SMD = -0.15, 95% CI -0.30 to 0.01). Revised forest plots for strategies 1 and 2 are provided in the online supplement.

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 online supplement.

Secondary efficacy outcomes

Full details concerning secondary efficacy outcomes are given in sections H and K of the online supplement.

Relapse, exacerbation and need for hospital care

Evidence from one study indicated that quetiapine IR was effective for prevention of symptom exacerbation in people with early psychosis who had responded to quetiapine,45 but an unpublished study suggested there was no effect of therapeutic dose (300–600 mg) over a subtherapeutic dose (75 mg) in relapse prevention in chronic schizophrenia.32 The combined estimate was therefore heterogeneous (I² = 87%) and not significant (NNT = 5, 95% CI 2 to 13H). Quetiapine IR was associated with a marginally reduced need for hospital care after 2–6 weeks in three RCTs (NNT = 19, 95% CI 10 to 143).36,37,44 One trial suggested quetiapine IR had a small effect over 52 weeks in relation to reducing readmission to hospital due to relapse (NNT = 11, 95% CI 6 to 143),45 but the results were not robust to changing assumptions about the outcome of those leaving early. Overall the relapse and readmission 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 to -0.20; moderate-quality evidence) and a marginal to small effect on negative symptoms (SMD = -0.21, 95% CI -0.32 to -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 to -0.02; low-quality evidence). Forest plots are provided in the online supplement. We did not investigate whether these estimates were robust to missing data, but the sensitivity analyses for the primary outcome of total symptoms suggest that 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 to 53H; six 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 to 19H; moderate-quality evidence).36,37

Pooled self-report end-point data from the two 12-week trials 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 to 0.36), but we judged the evidence to be very low in quality owing to early termination of the trials, missing data and possible selective reporting from the other trials. No significant effect was observed on any of the subscales, including psychological well-being (SMD = -0.02, 95% CI -0.28 to 0.24) or family relationships (SMD = 0.01, 95% CI -0.25 to 0.28). Since only observed case S-QoL data were reported, we imputed missing data using strategy 1 from Ebrahim et al.13 This reduced the overall effect from 0.11 to 0.06 (95% CI -0.14 to 0.27). Long-term quality of life data from two RCTs remain unpublished.32,45

An analysis of data from three RCTs (one studying adolescents, two with adult samples) covering a period of 6–12 weeks found quetiapine IR had a small to moderate benefit on functioning30,31,46 as assessed by a combination of Children’s Global Assessment Scale data and Personal and Social Performance (PSP) data (SMD = 0.39, 95% CI 0.18 to 0.60). Global Assessment of Functioning data were also reported, but unlike the PSP this measure 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 to 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 owing 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 period of 2–6 weeks (NNT = 21, 95% CI 10 to 333H).

Safety outcomes

Safety outcomes are detailed in Table 3 and Figs 4–6. There was low-quality evidence that quetiapine IR was associated with a small to moderately increased risk of non-serious adverse effects over the short term (NNH = 11, 95% CI 8 to 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 anti-Parkinsonian medication in quetiapine-treated participants in the short term, but longer-term data were 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 to 0.85). Participants gained an extra 1.75 kg (95% CI 1.10 to 2.40) on average, 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 withdrawal and high heterogeneity. Moderate-quality evidence suggested that around 12% of participants treated with quetiapine experienced a clinically significant increase in weight over 2–12 weeks, compared with 4% of those taking placebo (NNH = 13, 95% CI 9 to 23), and 35% reported sedation or somnolence as an adverse effect compared with 6% of those taking placebo (NNH = 9, 95% CI 7 to 13). Details on additional safety outcomes and forest plots are reported in sections I and K of the online supplement.

Discussion

Using published and unpublished data, we found the average change in PANSS total score attributable to quetiapine IR over 2–12 weeks to be 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. 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 that 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 had been underestimated because of the severe rates of withdrawal in the older studies. Approximately 21 people needed to take quetiapine IR for 1 person to experience much improvement, defined in accordance with recent recommendations as a 50% or greater reduction in PANSS total score.

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 owing to imprecision and incomplete reporting.

We found some evidence that estimates of clinically significant response derived from more recent trials were lower than in older trials, which is consistent with results from other antipsychotic meta-analyses, 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, or better masking of treatment allocation due to reduced expectation of receiving placebo in such trials.

Older reviews of quetiapine IR have reported effect sizes of around 0.4 when compared with placebo, and NNTs of around 10 or 11. However, most of the trials available at the time had high rates of participant withdrawal and examined quetiapine IR as a 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, or consider the clinical significance of the change attributable to quetiapine IR. 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. Since quetiapine 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 data for the two formulations would give an inaccurate appraisal of both. 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 that were available at the time of protocol writing. 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 3–3.5 weeks before switching to another antipsychotic because of non-response. Nonetheless, it is important to consider that overall efficacy was positively associated with trial duration in our review, and might have been larger still had we included quetiapine XR data. Our data may help explain why a recent meta-analysis found that quetiapine IR was significantly less effective at reducing positive symptoms than first-generation antipsychotics (nine RCTs). In an unpublished study therapeutic dose quetiapine IR was significantly less effective than haloperidol in preventing relapse.

### Missing data

Levels of missing data were high in the included trials. In order to reduce this, 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, 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 withdrawals and effect size, such analyses are inevitably limited by the fact that few trials have low rates of missing data. Application of the Ebrahim approach 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. 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, and selective reporting bias...
is a particular concern when, as documented by Spielmans & Parry and others,33 it biases our understanding of the severity of adverse effects.

**Study limitations**

We took advantage of several important developments in methodology which were published after we registered our protocol.13,20,28 Changes post hoc 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 a score of 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 designed their trials to be able to detect with adequate power only differences of approximately 12 points. Although it has been argued that small benefits might have value at a public health level,8 there is clearly a need for further debate on this issue. As with non-inferiority and equivalence trials,14 researchers planning superiority trials might consider stating in advance what they believe constitutes a minimum 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 for each trial, which is problematic given a recent study found a much better quality of reporting in these documents when compared with registry reports or peer-reviewed publications.50 Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents. Although this can be inferred from power calculations, it needs to be stated explicitly.

We were unable to access the full clinical study reports for each trial, which is problematic given a recent study found a much better quality of reporting in these documents when compared with registry reports or peer-reviewed publications.50 Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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 we doubt we would have been successful had a public debate on access to the reports would have raised the quality of many of the documents, when compared with registry reports or peer-reviewed publications. Although we have acquired a large amount of previously unpublished data, access to the reports 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.


Arvanitis L, Scott M. A multicenter, double blind, randomized, controlled, multiple fixed dose and dose regimen comparison of SEROQUELTM (ICI 264,636) and haloperidol in the prevention of psychotic relapse in outpatients with chronic or subchronic schizophrenia (5077IL/0015). AstraZeneca, 1996.


