Second-generation antipsychotics for schizophrenia: can we resolve the conflict?

S. Leucht*, W. Kissling and J. M. Davis

1 Department of Psychiatry and Psychotherapy, Technische Universität München, Germany
2 Department of Psychiatry, University of Illinois at Chicago, USA

The initial enthusiasm about the second-generation or atypical antipsychotic drugs soon changed into criticism and debate, culminating in the controversial CATIE (Clinical Antipsychotic Trials of Intervention Effectiveness), CUtLASS (Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study) and EUFEST (European First-Episode Schizophrenia Trial) effectiveness trials. This review summarizes the results of three recent meta-analyses that compared second-generation antipsychotics (SGAs) with placebo, with conventional antipsychotics, and with SGAs head-to-head. We compare the meta-analyses with previous reviews and put them in the perspective of CATIE, CUtLASS and EUFEST. The data show that the SGAs are not a homogeneous group and that this confusing classification should be abandoned. We find that, overall, the data are consistent but experts interpret the same results differently. The debate seems to be driven more by values than by data; some place an emphasis on cost, others focus on extrapyramidal side-effects (EPS), weight gain, or efficacy. In our opinion, the SGAs are not the breakthrough that industry would like to maintain. They have different properties, so a clinician may individualize a treatment plan to a given patient’s problems, a decision that should be shared with the patient. However, these drugs are important contributions to treatment, and most psychiatrists, let alone patients, would probably not want to do without them.

Introduction

In 1988 the proof that clozapine was more efficacious than first-generation antipsychotic drugs (FGAs) for the treatment of schizophrenia led to enormous enthusiasm about the development of other clozapine-like atypical antipsychotics (Kane et al. 1988), each new drug being vigorously marketed as an atypical or second-generation antipsychotic (SGA). Since the introduction of risperidone (Mattes, 1997), there has been a heated debate among clinicians and researchers on which drug is best, culminating in the, again vigorously debated, CATIE (Clinical Antipsychotic Trials of Intervention Effectiveness), CUtLASS (Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study) and EUFEST (European First-Episode Schizophrenia Trial) effectiveness studies (Lieberman et al. 2005; Jones et al. 2006; Lewis et al. 2006; Kahn et al. 2008). There is indeed extreme controversy, with some experts lauding the new drugs and others saying that, overall, there is no difference. In this context, we have recently published updated meta-analyses on the following topics: SGAs versus placebo, SGAs versus typical antipsychotics, and head-to-head comparisons of SGAs (Leucht et al. 2008a, 2009a, b), where we focused on adherence to the evidence without opinion. Here we explore reasons for the controversy and integrate the results. Our aim is to compare our numerical results with those of other meta-analyses (Geddes et al. 2000; Adams et al. 2008), to put them in perspective with the CATIE, CUtLASS and EUFEST studies, and to resolve or at least understand the controversy.

Method

We present a narrative summary of our previous meta-analyses comparing SGAs with placebo (a meta-analysis based on 38 studies with 7323 participants; Leucht et al. 2008a), FGAs (a meta-analysis based on 215 studies, 150 of those double-blind with 21 533 participants; Leucht et al. 2009a, b), and SGAs head-to-head (a meta-analysis of 78 studies with 13 558 participants; Leucht et al. 2009b).

Results

Tables 1–3 summarize the main findings of our reviews.
SGAs versus placebo (Table 1)

Efficacy

All SGAs were more efficacious than placebo based on various outcomes. Nevertheless, two points were striking: first, the effect sizes for overall efficacy were medium sized. Hedges’ g for mean Positive and Negative Syndrome Scale (PANSS)/Brief Psychiatric Rating Scale (BPRS) reduction ranged from 0.41 (aripiprazole) to 0.59 (risperidone; a single old clozapine study yielded an effect size of 1.64), and was 0.51 across drugs (Cohen’s classification of effect sizes is 0.20 for small, 0.50 for medium and 0.80 for large). The number needed to treat (NNT) ranged from 4 to 7, and was 6 overall. It was surprising that haloperidol was efficacious for both negative symptoms and depression.

Side-effects

Overall, there were a few occasions where SGAs were more sedating than placebo, but there were no differences in drop-outs due to adverse events (except sertindole using risk difference). Most SGA ratings of extrapyramidal side-effects (EPS) were similar to placebo, but amisulpride was only studied in low doses (maximum 300 mg/day). Risperidone had a non-significant trend (p = 0.07) to produce more EPS, and 32% of risperidone patients received antiparkinsonian medication (compared to 26% of placebo patients), although clearly less so than haloperidol (48% compared to 20% in its placebo groups; Leucht et al. 2008a). A meta-analysis on SGAs for bipolar mania also suggested that some SGAs (aripiprazole, risperidone and ziprasidone; p = 0.06) do induce some EPS (see fig. 3 in Scherk et al. 2007), perhaps because manic patients are more sensitive towards EPS. A more plausible explanation is that manic patients are less likely to be on antipsychotic medication at the beginning of the trial than patients with schizophrenia, which can lead to carry-over effects. Similarly, in an aripiprazole study on adolescents with schizophrenia, EPS clearly occurred (see Table 3 in Findling et al. 2008).

SGAs versus FGAs (Table 2)

Efficacy

Four (amisulpride, clozapine, olanzapine, risperidone) out of nine SGAs were more efficacious than FGAs with small to medium effect sizes in terms of the primary outcome overall symptoms (ranging from an effect size of 0.13 for risperidone to 0.52 for clozapine, NNT 6–15). Of note, the four SGAs that were more efficacious overall were also more efficacious for the specific positive symptoms and negative symptoms whereas the others were not (quetiapine was less effective for positive symptoms). We therefore concluded that negative symptoms cannot be a core component of ‘atypicality’. The relatively few data available for depression were slightly different; aripiprazole and quetiapine were more efficacious whereas risperidone was not. Few data on quality of life and relapse were available (Leucht et al. 2009a).

Side-effects

All SGAs induced fewer EPS than haloperidol, most even when the latter was used in low doses below 7.5 mg/day. In the relatively few studies that compared SGAs with low-potency antipsychotics, the difference in EPS was less clear. When compared to haloperidol, clozapine, olanzapine, sertindole and zotepine induced the most weight gain whereas quetiapine and risperidone caused intermediate weight gain and amisulpride stimulated yet a smaller amount of weight gain. However, aripiprazole and ziprasidone induced no significant weight gain. Importantly, the low-potency FGAs also induce weight gain. The SGAs also differed in their sedative properties; some were more sedating than haloperidol but some were less sedating than low-potency FGAs.

SGAs versus SGAs head-to-head (Table 3)

Head-to-head comparisons of SGAs are needed to compare the various SGAs because indirect comparisons relative to a common standard can be influenced by many confounders. Overall, a similar efficacy pattern was observed as we found in the meta-analysis with FGA as comparator. Olanzapine was more efficacious than aripiprazole, quetiapine, risperidone and ziprasidone; risperidone was more efficacious than quetiapine and ziprasidone. Amisulpride was not statistically different from olanzapine or risperidone, but was more efficacious than ziprasidone in drop-out due to inefficacy. Thus, the only notable difference was clozapine, which proved to be superior only to zotepine, and to risperidone in drop-out due to inefficacy. Possibly too low a dose of clozapine may explain this surprising finding (see discussion; Leucht et al. 2009b). (Our results on comparative side-effects of the various SGAs have not yet been published.)

Other reviews comparing SGAs with FGAs

The overall efficacy results of the reviews by Geddes et al. (2000), Davis et al. (2003) and the Cochrane Schizophrenia Group (Adams et al. 2008) are presented in Fig. 1. Overall, the effect sizes are comparable to those of our updated reviews.
Table 1. Second-generation antipsychotic drugs (SGAs) and haloperidol versus placebo (Leucht et al. 2008a)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Amisulpride</th>
<th>Aripiprazole</th>
<th>Clozapine</th>
<th>Haloperidol</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
<th>Risperidone</th>
<th>Sertindole</th>
<th>Ziprasidone</th>
<th>Zotepine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall symptoms</td>
<td>N = 1, n = 241</td>
<td>N = 7, n = 1556</td>
<td>N = 1, n = 22</td>
<td>N = 11, n = 1540</td>
<td>N = 6, n = 992</td>
<td>N = 5, n = 735</td>
<td>N = 7, n = 977</td>
<td>N = 3, n = 629</td>
<td>N = 4, n = 584</td>
<td>N = 3, n = 304</td>
</tr>
<tr>
<td>N</td>
<td>211</td>
<td>1550</td>
<td>21</td>
<td>1439</td>
<td>991</td>
<td>734</td>
<td>977</td>
<td>629</td>
<td>584</td>
<td>304</td>
</tr>
<tr>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressive symptoms</td>
<td>N = 2, n = 261</td>
<td>N = 6, n = 1489</td>
<td>N.A.</td>
<td>N = 1, n = 22</td>
<td>N = 2, n = 299</td>
<td>N = 3, n = 479</td>
<td>N = 2, n = 333</td>
<td>N = 2, n = 459</td>
<td>N = 3, n = 404</td>
<td>N = 1, n = 79</td>
</tr>
<tr>
<td>N</td>
<td>255</td>
<td>1470</td>
<td></td>
<td>22</td>
<td>298</td>
<td>479</td>
<td>333</td>
<td>459</td>
<td>404</td>
<td>79</td>
</tr>
<tr>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of antiparkinson medication</td>
<td>N = 3, n = 487</td>
<td>N = 6, n = 1310</td>
<td>N.A.</td>
<td>N = 1, n = 1608</td>
<td>N = 3, n = 481</td>
<td>N = 3, n = 509</td>
<td>N = 4, n = 323</td>
<td>N = 3, n = 661</td>
<td>N = 4, n = 598</td>
<td>N = 2, n = 227</td>
</tr>
<tr>
<td>RR</td>
<td>0.87</td>
<td>1.07</td>
<td></td>
<td>RR 2.34</td>
<td>RR 1.23</td>
<td>RR 0.79</td>
<td>RR 1.24</td>
<td>RR 0.79</td>
<td>RR 1.33</td>
<td>RR 1.49</td>
</tr>
<tr>
<td>(0.24–3.20)</td>
<td>(0.81–1.41)</td>
<td></td>
<td></td>
<td>(1.90–2.80)</td>
<td>(0.52–2.93)</td>
<td>(0.46–1.33)</td>
<td>(0.51–1.23)</td>
<td>(0.70–2.51)</td>
<td>(0.60–3.72)</td>
<td></td>
</tr>
<tr>
<td>Sedation</td>
<td>N = 4, n = 1107</td>
<td>N.A.</td>
<td>N = 4, n = 670</td>
<td>N = 3, n = 408</td>
<td>N = 3, n = 670</td>
<td>N = 4, n = 408</td>
<td>N = 2, n = 291</td>
<td>N = 2, n = 312</td>
<td>N = 3, n = 312</td>
<td>N = 2, n = 312</td>
</tr>
<tr>
<td>RR</td>
<td>1.87</td>
<td>1.07</td>
<td></td>
<td>RR 2.28</td>
<td>RR 1.93</td>
<td>RR 2.02</td>
<td>RR 2.13</td>
<td>RR 2.08</td>
<td>RR 4.60</td>
<td></td>
</tr>
<tr>
<td>(0.82–2.34)</td>
<td>(0.82–3.44)</td>
<td></td>
<td></td>
<td>(1.11–4.67)</td>
<td>(0.76–4.90)</td>
<td>(0.73–4.20)</td>
<td>(0.73–4.20)</td>
<td>(0.62–6.95)</td>
<td>(1.21–5.04)</td>
<td></td>
</tr>
<tr>
<td>Drop-out due to adverse events</td>
<td>N = 4, n = 514</td>
<td>N = 7, n = 1615</td>
<td>N = 1, n = 24</td>
<td>N = 10, n = 1476</td>
<td>N = 5, n = 914</td>
<td>N = 4, n = 521</td>
<td>N = 4, n = 599</td>
<td>N = 3, n = 661</td>
<td>N = 4, n = 598</td>
<td>N = 3, n = 312</td>
</tr>
<tr>
<td>RR</td>
<td>0.59</td>
<td>0.73</td>
<td></td>
<td>RR 1.31</td>
<td>RR 0.83</td>
<td>RR 1.30</td>
<td>RR 0.86</td>
<td>RR 1.49</td>
<td>RR 1.88</td>
<td></td>
</tr>
<tr>
<td>(0.20–1.78)</td>
<td>(0.89–2.04)</td>
<td></td>
<td></td>
<td>(0.80–2.12)</td>
<td>(0.80–2.12)</td>
<td>(0.80–2.12)</td>
<td>(0.80–2.12)</td>
<td>(0.80–2.12)</td>
<td>(0.80–2.12)</td>
<td></td>
</tr>
</tbody>
</table>

N. Number of studies; n, number of participants; SMD, standardized mean difference (effect size) calculated as Hedges’ g; RR, relative risk; EPS, extrapyramidal side-effects measured by use of antiparkinson medication at least once; N.A., not available.

Numbers in parentheses are 95% confidence intervals. Negative SMDs reflect a superiority of SGAs.

* p < 0.05, ** p < 0.01.
Table 2. Second-generation antipsychotic drugs (SGAs) versus first-generation antipsychotic drugs (FGAs) (Leucht et al. 2009a)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Amisulpride</th>
<th>Aripiprazole</th>
<th>Clozapine</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
<th>Risperidone</th>
<th>Sertindole</th>
<th>Ziprasidone</th>
<th>Zotepine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall symptoms</td>
<td>N = 13, n = 1017</td>
<td>N = 5, n = 2049</td>
<td>N = 23, n = 1997</td>
<td>N = 28, n = 4966</td>
<td>N = 11, n = 2412</td>
<td>N = 34, n = 4173</td>
<td>N = 4, n = 1344</td>
<td>N = 5, n = 980</td>
<td>N = 15, n = 1125</td>
</tr>
<tr>
<td>Positive symptoms</td>
<td>N = 4, n = 703</td>
<td>N = 4, n = 1983</td>
<td>N = 10, n = 1080</td>
<td>N = 24, n = 4189</td>
<td>N = 9, n = 1742</td>
<td>N = 28, n = 3286</td>
<td>N = 3, n = 1145</td>
<td>N = 4, n = 728</td>
<td>N = 2, n = 192</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>N = 10, n = 929</td>
<td>N = 5, n = 2049</td>
<td>N = 17, n = 1603</td>
<td>N = 24, n = 4187</td>
<td>N = 10, n = 1926</td>
<td>N = 30, n = 3455</td>
<td>N = 4, n = 1198</td>
<td>N = 3, n = 691</td>
<td>N = 5, n = 450</td>
</tr>
<tr>
<td>Depression</td>
<td>N = 9, n = 900</td>
<td>N = 1, n = 1278</td>
<td>N = 6, n = 462</td>
<td>N = 12, n = 2893</td>
<td>N = 4, n = 442</td>
<td>N = 11, n = 1611</td>
<td>N = 2, n = 574</td>
<td>N = 3, n = 691</td>
<td>N = 2, n = 134</td>
</tr>
<tr>
<td>Quality of Life</td>
<td>N = 1, n = 194</td>
<td>N = 1, n = 206</td>
<td>N = 1, n = 311</td>
<td>N = 5, n = 1450</td>
<td>N = 2, n = 166</td>
<td>N = 4, n = 330</td>
<td>N = 1, n = 105</td>
<td>N = 1, n = 72</td>
<td>N = 1, n = 122</td>
</tr>
<tr>
<td>EPS*</td>
<td>N = 11, n = 985</td>
<td>N = 5, n = 2094</td>
<td>N = 6, n = 322</td>
<td>N = 14, n = 3758</td>
<td>N = 10, n = 2031</td>
<td>N = 28, n = 3430</td>
<td>N = 4, n = 1472</td>
<td>N = 4, n = 807</td>
<td>N = 9, n = 720</td>
</tr>
<tr>
<td>RR 0.61</td>
<td>RR 0.49</td>
<td>RR 0.16</td>
<td>RR 0.38</td>
<td>RR 0.56</td>
<td>RR 0.62</td>
<td>RR 0.36</td>
<td>RR 0.65</td>
<td>RR 0.76</td>
<td></td>
</tr>
<tr>
<td>(0.48–0.77)**</td>
<td>(0.36–0.66)*</td>
<td>(0.05–0.49)**</td>
<td>(0.40–0.79)**</td>
<td>(0.55–0.71)**</td>
<td>(0.29–0.54)**</td>
<td>(0.39–1.07)*</td>
<td>(0.54–1.07)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sedation*</td>
<td>N = 5, n = 622</td>
<td>N = 3, n = 1902</td>
<td>N = 15, n = 1583</td>
<td>N = 8, n = 2879</td>
<td>N = 9, n = 1834</td>
<td>N = 20, n = 2533</td>
<td>N = 3, n = 1127</td>
<td>N = 2, n = 607</td>
<td>N = 7, n = 554</td>
</tr>
<tr>
<td>RR 0.86</td>
<td>RR 0.63</td>
<td>RR 1.31</td>
<td>RR 0.93</td>
<td>RR 0.93</td>
<td>RR 0.92</td>
<td>RR 0.77</td>
<td>RR 0.99</td>
<td>RR 1.39</td>
<td></td>
</tr>
<tr>
<td>(0.41–1.79)</td>
<td>(0.44–0.89)**</td>
<td>(1.10–1.57)**</td>
<td>(0.49–1.76)</td>
<td>(0.77–1.09)</td>
<td>(0.44–1.34)</td>
<td>(0.42–2.32)</td>
<td>(1.01–1.93)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight gain (kg)*</td>
<td>0.28</td>
<td>0.19</td>
<td>2.08</td>
<td>3.46</td>
<td>1.26</td>
<td>1.72</td>
<td>2.38</td>
<td>-0.49</td>
<td></td>
</tr>
<tr>
<td>(−0.94 to 1.50)</td>
<td>(−0.06 to 0.97)</td>
<td>(0.36–3.81)</td>
<td></td>
<td>(2.46–4.46)*</td>
<td>(0.65–1.87)**</td>
<td>(1.08–2.36)**</td>
<td>(−0.19 to 4.95)</td>
<td>(−1.63 to 0.64)</td>
<td>(1.58–3.24)**</td>
</tr>
</tbody>
</table>

N, Number of studies; n, number of participants; SMD, standardized mean difference (effect size) calculated as Hedges’ g; RR, relative risk; EPS, extrapyramidal side-effects measured by use of antiparkinson medication at least once.

Numbers in parentheses are 95% confidence intervals.

Negative SMDs reflect a superiority of SGAs.

*aSGAs compared to any FGA (haloperidol or low-potency), the text sometimes mentions the results compared to haloperidol or low-potency FGAs.

*p < 0.05, **p < 0.001.
Table 3. Second-generation antipsychotics (SGAs) head-to-head, primary outcome PANSS/BPRS total score (modified from Leucht et al. 2009b)

<table>
<thead>
<tr>
<th>Amisulpride</th>
<th>Aripiprazole</th>
<th>Clozapine</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
<th>Risperidone</th>
<th>Sertindole</th>
<th>Ziprasidone</th>
<th>Zotepine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aripiprazole</strong></td>
<td><strong>--</strong></td>
<td><strong>OLA †</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
</tr>
<tr>
<td><strong>Clozapine</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
</tr>
<tr>
<td><strong>Olanzapine</strong></td>
<td><strong>N=4, n=701</strong></td>
<td><strong>N=2, n=794</strong></td>
<td><strong>N=10, n=731</strong></td>
<td><strong>SMD −0.07</strong></td>
<td><strong>SMD −0.22</strong></td>
<td><strong>SMD −0.11</strong></td>
<td><strong>−0.28 to 0.14</strong></td>
<td><strong>−0.36 to 0.08)</strong></td>
</tr>
<tr>
<td><strong>Quetiapine</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>OLA †</strong></td>
<td><strong>N=5, n=299</strong></td>
<td><strong>N=10, n=1449</strong></td>
<td><strong>SMD 0.23</strong></td>
<td><strong>−0.30 to 0.15</strong></td>
<td><strong>(0.12–34)</strong>**</td>
</tr>
<tr>
<td><strong>Risperidone</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>OLA †</strong></td>
<td><strong>RIS †</strong></td>
<td><strong>N=16, n=2438</strong></td>
<td><strong>N=10, n=1978</strong></td>
<td><strong>RIS †</strong></td>
</tr>
<tr>
<td><strong>Sertindole</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>N=2, n=493</strong></td>
<td><strong>SMD 0.09</strong></td>
<td><strong>(−0.36 to 0.54)</strong></td>
</tr>
<tr>
<td><strong>Ziprasidone</strong></td>
<td><strong>--</strong></td>
<td><strong>--</strong></td>
<td><strong>OLA †</strong></td>
<td><strong>N=1, n=146</strong></td>
<td><strong>N=5, n=1542</strong></td>
<td><strong>N=2, n=710</strong></td>
<td><strong>N=3, n=1016</strong></td>
<td><strong>RIS †</strong></td>
</tr>
<tr>
<td><strong>Zotepine</strong></td>
<td><strong>N=1, n=122</strong></td>
<td><strong>N=1, n=146</strong></td>
<td><strong>N=5, n=1542</strong></td>
<td><strong>N=2, n=710</strong></td>
<td><strong>N=2, n=731</strong></td>
<td><strong>SMD 0.11</strong></td>
<td><strong>SMD −0.19</strong></td>
<td><strong>(−0.06 to 0.32)</strong></td>
</tr>
<tr>
<td><strong>SMD 0.07</strong></td>
<td><strong>SMD 0.07</strong></td>
<td><strong>SMD 0.13</strong></td>
<td><strong>SMD 0.11</strong></td>
<td><strong>SMD 0.08</strong></td>
<td><strong>SMD 0.00</strong></td>
<td><strong>SMD 0.11</strong></td>
<td><strong>SMD 0.21</strong></td>
<td><strong>(0.07–0.34)</strong></td>
</tr>
<tr>
<td><strong>−0.10 to 0.24</strong></td>
<td><strong>−0.28 to 0.14</strong></td>
<td><strong>−0.06 to 0.32</strong></td>
<td><strong>0.03–0.19)</strong></td>
<td><strong>−0.10 to 0.28)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PANSS, Positive and Negative Syndrome Scale; BPRS, Brief Psychiatric Rating Scale; N, number of studies; n, number of participants; SMD, standardized mean difference (effect size) calculated as Hedges’ g; CLO, clozapine; OLA, olanzapine; RIS, risperidone.

Numbers in parentheses are 95% confidence intervals.

Negative SMDs reflect a superiority of the SGA listed in the first column.

Blank fields reflect that no study is available.

† Statistically significantly superior, ↔ no significant difference between groups.

*p < 0.05, **p < 0.001.
Comparison with CATIE

Fig. 2 shows the primary efficacy outcomes reported in the primary publication of CATIE (Lieberman et al. 2005).

Discussion

Are the data consistent?

Fig. 1 shows that the results of our recent meta-analysis are consistent with prior meta-analyses and Cochrane Reviews (Geddes et al. 2000; Davis et al. 2003; Adams et al. 2008). All reviews found that amisulpride, clozapine, olanzapine and risperidone were significantly more efficacious than FGAs whereas for the other SGAs there were no significant differences. The data are consistent with CATIE-I (Lieberman et al. 2005) in that olanzapine was superior in several efficacy outcomes (drop-out due to inefficacy, time on effective treatment, see Fig. 2), and (open-label) clozapine was better than other SGAs in CATIE-II (McEvoy et al. 2006) and in CUTLASS (Lewis et al. 2006). Our meta-analyses found a somewhat better result for risperidone, possibly because in CATIE the
mean modal dose was 3.9 mg/day and only 40% received the 6 mg/day dose, meaning that as many patients as 30–40% received 3 mg or 1.5 mg/day, less efficacious doses according to randomized dose studies (Davis & Chen, 2004) (guidelines suggest 2–8 mg/day; Lehman et al. 2004). Clozapine did not turn out to be superior to other SGAs as had been expected, but doses were usually well below 400 mg/day (five were <210 mg/day) and thus clearly lower than in the pivotal studies showing superiority to FGAs (Kane et al. 1988; Rosenheck et al. 1997; 600 and 523 mg/day respectively) and lower than the optimum doses (Simpson et al. 1999; Davis & Chen, 2004). A sufficiently dosed double-blind clozapine versus other SGAs trial is still needed.

There is minimal controversy about side-effects and the meta-analytic results are fairly consistent with CATIE.

Are the data flawed?

Several possible methodological flaws have been put forward. Here we discuss industry bias, EPS artifact, statistical methods, blinding and general problems of current clinical trials.

Industry bias

In an analysis of 33 industry-sponsored head-to-head comparisons of SGAs, our blind ratings of abstracts found that 90% favored the sponsor’s drug, which provides an answer to our title ‘why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine’ (Heres et al. 2006). However, Davis et al. (2008) examined the efficacy effect sizes and found no difference between industry and non-industry sponsored trials, a finding consistent with our two later meta-analyses (SGA versus FGA and SGA versus SGA; Leucht et al. 2009a, b). We conclude that most of the ‘industry sponsorship’ effect (Heres et al. 2006) is due to the spin that the authors put on the data, which contributes to enormous confusion and creates misinformation. That being said, there are of course some studies with obviously flawed designs such as using too low clozapine doses or omitting a result. Failure to report a specific finding can blur interpretation; for example, positive symptoms have never been published in the single comparison of ziprasidone with amisulpride (Olie et al. 2006).

High potency FGA comparator artifact in the evaluation of efficacy and side-effects

Could the EPS from high-dose haloperidol or other high-potency FGAs mimic (negative) symptoms and artificially inflate any SGA superiority? Geddes et al. (2000) observed the same effect sizes as we did but interpreted them differently, attributing the observed superiority to occur as an artifact of high haloperidol doses. Other larger meta-analyses could not replicate the dose effect with further studies (Davis et al. 2003; Leucht et al. 2009a; Table 1 in Davis et al. 2008). A Veterans Affairs (VA) collaborative study provided an intriguing investigation of the EPS artifact hypothesis by comparing prophylactic antiparkinsonian medication combined with haloperidol versus olanzapine and found little difference (Rosenheck et al. 2003). In our dataset there were 11 studies involving only three SGAs on the use of prophylactic anti-parkinsonian medication, and these failed to clearly demonstrate this effect for clozapine and olanzapine (Leucht et al. 2009a).

In addition, a dose–response analysis by Davis & Chen (2004) found that higher doses of haloperidol were not associated with less efficacy than lower doses. Marder et al. 1997; see Table 5) and Davis & Chen (2001, p. 769) examined the original patient data and found no significant correlation between observed EPS and efficacy. We hasten to add that EPS can, and probably does, influence ratings of negative symptoms (or vice versa), but firm proof is not available and such an effect would have to be large enough to fully explain the difference.

Nevertheless, haloperidol was used in 95 out of 150 studies; partial justification being that it was the standard antipsychotic at the time of the SGAs’ introduction. We find that high-potency FGAs clearly cause more EPS. Low-potency FGAs did not induce more EPS than some, but not all, SGAs, but they lead to weight gain and are sedating (Leucht et al. 2009a). It is a major limitation that only a few studies used mid-potency FGA comparators. We recommend that each new drug is compared with a low-potency, a mid-potency, and a high-potency FGA. However, in the absence of such data, ‘no evidence of effect does not mean evidence of no effect’ (Tarnow-Mordi & Healy, 1999).

Statistical methods

Many have suggested that the efficacy superiority of some SGAs may have been an artifact of the use of last observation carried forward (LOCF) analyses. The argument made concerns the following: when a patient terminates a study prematurely, in LOCF their last observation is used (‘carried forward’) as their end-point evaluation. If this happens more frequently with haloperidol, SGAs have more time to act on symptoms in an LOCF analysis. We had the opportunity to reanalyze original patient data from pivotal studies comparing amisulpride and olanzapine with
conventional antipsychotics, but did not find a clear LOCF bias (Leucht et al. 2007). Nevertheless, it would be of interest to do such reanalyses on head-to-head comparisons of SGAs. Participants on drugs with side-effects that occur early on, such as EPS (e.g. risperidone) or sedation (e.g. quetiapine), may drop out earlier than those on drugs with side-effects that appear later, such as weight gain (e.g. olanzapine).

Blinding

We found that open studies do have a bias in favor of the SGAs but our meta-analysis was based on just double-blind (Leucht et al. 2009a).

General methodological problems of current antipsychotic drug trials

The populations of patients who enter studies change. For example, risperidone and olanzapine studies usually occurred in subjects who had never received an SGA. Patients who entered later trials may have already been on one or more SGAs. This might make it more difficult to show a difference between SGAs introduced later (e.g. aripiprazole) and FGAs. Nevertheless, we found the same efficacy pattern in head-to-head comparisons of SGAs, where more participants would have already received risperidone/olanzapine.

There are other methodological problems with current antipsychotic drug trials, including: high drop-out rates, too short wash-out phases, chronicity of the participants, selected patient populations, and lack of standardized response criteria (for a review, see Leucht et al. 2008b). These limitations could either inflate or reduce differences, and we could find no evidence that the observed difference among some drugs was a methodological artifact.

Why is there a conflict?

If we accept that, overall, the data are fairly consistent and not entirely biased, we need to discuss the factors producing the controversy.

Promoting all new drugs as one group

The new drugs were all vigorously promoted as being better drugs, similar to clozapine but with pharmacological reasons why they should be better. However, many have never been shown to be more efficacious from the first registrational studies onwards. Such promotion keeps the drugs constantly in the physicians’ minds as a class that is essentially different from anything available before. Companies consistently emphasize the disadvantages of their competitors’ drugs (sometimes referred to as ‘counter-marketing’).

However, the most insidious effect of promotion is what is not said. All this promotion contributes to enormous confusion.

Cost

The SGAs cost US$7.5 billion in the USA in 2003, as much as the cost of all US psychiatrists (Rosenheck, 2005). Without the high cost of the new agents as a key driver, the debate might not even exist. To be cost-effective, higher acquisition costs must be counterbalanced by better efficacy for a long enough time, leading to fewer hospitalizations. As CATIE and CUI LASS did not find important efficacy differences, it was not surprising that their cost-effectiveness analyses found the cheaper drugs to be superior. To find cost-effectiveness, the patients need to be on the initial drug long enough so that fewer hospitalizations can produce the cost offset. Some may suggest that the money spent on the SGAs should rather be spent on psychosocial therapies. Unfortunately, however, the money might be allocated to other areas of medicine such as cardiology or cancer.

How large are the efficacy differences?

Cohen classified effect sizes of 0.20 as small, 0.50 medium, and 0.80 large, but cautioned against overinterpretation. The effect size of clozapine versus FGAs was 0.52 (NNT = 6). The other significant effect sizes ranged between 0.11 and 0.31, with NNTs between 6 and 15. For perspective, the effect size of haloperidol versus placebo was 0.53 and the NNT equaled 9 (Leucht et al. 2008a). There are many differences between placebo-controlled and active comparator-controlled trials, such as higher drop-out rates in the former (Kemmler et al. 2005), that make it impossible to say that clozapine ‘doubles’ the efficacy compared with placebo (haloperidol versus placebo 0.53 + clozapine versus FGA 0.52). For another perspective, the NNT to reduce mortality by statins is 88, for aspirin it is 1000, and the effect size of antihypertensive drugs on blood pressure is 0.50. In our opinion, highly disturbed patients are rarely included in trials, and those who are allowed have undergone some partial stabilization. We speculate that this would decrease the effect sizes. Schizophrenia onset is in adolescence/early adulthood and afflicts patients for life, and even a small benefit for a long period of time could be important.

Is the debate more driven by values than by data?

Philosophers of science classify knowledge into three categories: true by definition (logic), empirically true (science), and other (values, ethics, religion). This
distinction is useful here because we feel that many statements on treatment of choice fall in the last category and are value driven; that is, which outcome is more important cannot be proven right or wrong by empirical evidence. For example, those interested in efficacy might value efficacy in particular. Those interested in, for example, weight gain or type 2 diabetes, might emphasize avoiding drugs with these side-effects; those concerned about cost may emphasize the importance of using low-cost drugs; and those who believe that we need better psychosocial treatments will want funds for psychosocial treatments, not expensive medication. The problem is that it is impossible to conclude which value is the most important, as they are qualitatively different. What an expert’s value is should not be confused with what the data are.

CATIE, CUtLASS and EUFEST stimulated the conflict

CATIE. We described earlier how many CATIE results are compatible with the meta-analyses. CATIE is frequently misrepresented as having different results and the assumed reason is ‘industry independent versus industry sponsored’, but we did not find a clear sponsor effect. The misperception was that all SGAs are more efficacious, but even the registrational studies failed to find quetiapine and ziprasidone to be more efficacious.

In CATIE, there were more discontinuations due to EPS and a greater use of antiparkinsonian agents in the perphenazine group. Nevertheless, the EPS differences were smaller than in our meta-analysis, which mainly used haloperidol, and in CATIE there were no differences in rated EPS. It could be that EPS, which are complex phenomena, are not adequately measured in large multi-site trials. For example, the measurement of EPS is altered by variable use of antiparkinsonian medication or different skills of raters. A considerable number of CATIE patients were on prophyllactic antiparkinsonian medication that was not withdrawn before the study. Perphenazine induces few dystonic reactions and was a wise choice, but we need more studies comparing it with the SGAs to define the exact EPS differences. We speculate that the EPS risk of perphenazine is close to that of risperidone (Hoyberg et al., 1993), but higher than that of quetiapine or olanzapine. Finally, there is no perfect study, and CATIE’s limitations have been discussed extensively elsewhere (e.g. Kasper & Winkler, 2006), reminding us of an old joke in the theatre community: How many actors does it take to change a light bulb? Twenty-six: one to change it and the other 25 to say they could have done it better. We highlight that the primary analysis based on survival analysis is valid, but the 74% overall discontinuation rate creates problems to continuous variables. With so many non-random drop-outs, the groups become progressively no longer randomized. Seventy-four per cent of the data must be guessed by statistical modeling. As there were differences between some drugs in drop-outs due to poor efficacy, analysis of observed data at time points late in the study will differ because of a disproportional retention confounding efficacy evaluation. To illustrate this, consider a cancer trial where the drug reduces death over placebo; at the end-point there will still be patients on placebo who are healthy, but the few healthy controls will be as healthy as the many drug-treated patients.

CUtLASS. CUtLASS found no difference between SGAs and FGAs apart from the higher SGA cost. This finding was not entirely unexpected because, in CUtLASS, clinicians could choose among SGAs and FGAs (the only depots available at that time were FGAs), making conclusions about individual drugs impossible, and we found that not all SGAs are more efficacious. In addition, 60% of the participants of the FGA group were started on sulpiride, a drug chemically similar to amisulpride and therefore probably an old ‘atypical’. It seems that you can do reasonably well when you choose carefully among old drugs. We would hasten to add that sulpiride was not well investigated, indeed even the optimum dose range is not known. Neither sulpiride nor perphenazine was the mainstay of treatment (and not even available in some countries). We fear that an interpretation of CUtLASS that all drugs are equal would make psychiatrists return to old bad habits, such as high-dose haloperidol.

The design used in many effectiveness studies (CUtLASS), many US VA collaborative studies and some CATIE analyses keeps patients in the drug group initially randomized to evaluate the patients after switching to other treatments. Although useful in certain situations, this design may blur drug efficacy and side-effects because many of the patients are on a different drug for part of the study. To illustrate this, consider an automobile race from Alaska to the tip of Argentina, where drivers were randomized to an very expensive BMW or the most inexpensive Ford (disclosure: BMW is a major funder of the Technical University of Munich). If the car broke down during the race, the driver could choose either a BMW or a Ford as a replacement car and could even keep the car at the end of the race. Many Fords quickly broke down and the drivers invariable replaced them with BMWs. The drivers assigned randomly to Fords did almost as well as the BMWs, possibly because most were driving BMWs for most of the race.
EUFESt. The EUFESt study found a similar rank order in that olanzapine and amisulpride turned out best in discontinuation due to inefficacy, but not identical because haloperidol did worse than ziprasidone. Side-effects were again even more consistent with previous evidence. Our finding that unblinded trials favored the SGAs may in part explain why those SGA equal to haloperidol in blinded studies were intermediate but statistically superior to haloperidol (Leucht et al. 2009a).

What are the implications of the data?

SGAs are not a homogeneous category and not a class

As SGAs differ in many properties, including efficacy, side-effects (even in the occurrence of EPS), cost (some are becoming generic), and pharmacology (amisulpride is not a serotonin blocker), they do not form a homogeneous class. Humans like to believe in a consistent body of evidence, avoiding cognitive dissonance. They focus on evidence consistent with their beliefs, and ignore or minimize evidence inconsistent with their beliefs. They therefore distort or simplify information to make it consistent with their dominant cognitions and beliefs. Therefore, many may have initially thought that clozapine was a more efficacious drug, they concluded that all so-called SGAs were more efficacious. As recent evidence has shown that some new drugs are not more efficacious, others may try to make this dissonance consistent by saying that all antipsychotics are the same. As clozapine seems to be effective for negative symptoms, they attribute this to all SGAs, even though most were not shown to be more effective for negative symptoms from the registrational studies onwards. Some confuse empirical data with their conclusions, making a psychoanalytical interpretation as to what the data really show. Most importantly, forcing improper generalizations that all SGAs are the same creates confusion and the classification might have to be abandoned (Leucht et al. 2009a).

Implications for choice of drug

Our meta-analyses do not suggest that haloperidol should be used first line, and so does the Cochrane review on haloperidol versus placebo (Joy et al. 2007). Even in very low doses (2–4 mg) it led to more EPS than several FGAs (Zimbroff et al. 1997; Schouler et al. 2005; Kahn et al. 2008). Low-potency FGAs have a lower EPS risk but they induce weight gain, sedation and other side-effects [hypotension, QT-prolongation, sudden death (thioridazine)]. Some mid-potency antipsychotics may avoid problems of both low- and high-potency antipsychotics, but these drugs have not been sufficiently studied (dose range, etc.). To avoid the side-effects of the more efficacious drugs clozapine, olanzapine (weight gain, etc.), amisulpride and risperidone (some EPS and more prolactin increase than haloperidol), physicians could start patients on aripiprazole or ziprasidone, which are just as efficacious as haloperidol, yet more tolerable. Quetiapine has very low EPS but intermediate weight gain. Nevertheless, it is useful to have many options available and we feel it is not justified to conclude that FGAs should never be used. We still use them in many circumstances, although usually second line.

There is substantial variability between individual patients in how they respond to antipsychotic drugs. Not all patients will develop weight gain on olanzapine or EPS on haloperidol. Therefore, the drug of choice must be tailored to the individual patient. Furthermore, we strongly recommend shared decision making with the patient (see review by Hamann et al. 2003). After all, it is the patients who take the medication.

Conclusion

Marketing by pharmaceutical companies has often promoted SGAs by smoke and mirrors. Many hopes in the SGAs, such as dramatically better efficacy, compliance, quality of life and no side-effects, have not been fulfilled. Overall, the data are consistent, but value judgments and spin have led to different interpretations. Nevertheless, we have no doubt that the introduction of these compounds has contributed to the treatment of schizophrenia and, as an article in 2003 in The New York Times said, ‘few psychiatrists – and perhaps even fewer patients – would want to lose any of the newer generation of antipsychotics now on the market’ (Goode, 2003).

Acknowledgements

Financial support was provided by a grant from the German Federal Ministry of Education and Research (no. FKZ: 01KG 0606, GZ: GFKG01100506) to S.L.; and a grant from the National Institute of Mental Health, Advanced Center for Intervention and Services Research Center (MH-68580), Grant No. 1 P01MH68580-01 CFDA #93.242, the Maryland Psychiatric Research Center to J.D.

Declaration of Interest

S.L. has received speaker and/or consultancy honoraria from SanofiAventis, BMS, EliLilly, Janssen/Johnson and Johnson, Lundbeck and Pfizer; and he has received funding for research projects from
EliLilly and SanofiAventis. W.K. has received speaker and/or advisory board/consultancy honoraria from Janssen, Sanofi-Aventis, Johnson and Johnson, Pfizer, Bayer, BMS, Astra Zeneca, Lundbeck, Novartis and EliLilly.

References


Davis JM, Chen N, Glick ID (2003). A meta-analysis of the efficacy of second-generation antipsychotics. *Archives of General Psychiatry* 60, 553–564.

Davis JM, Chen N, Glick ID (2008). Issues that may determine the outcome of antipsychotic trials: industry sponsorship and extrapyramidal side effect. *Neuropsychopharmacology* 33, 971–975.


