Schizophrenia Bulletin vol. 33 no. 1 pp. 183–191, 2007
doi:10.1093/schbul/sbl025
Advance Access publication on August 11, 2006

Is the Superior Efficacy of New Generation Antipsychotics an Artifact of LOCF?

Stefan Leucht[1,2], Rolf R. Engel[3], Josef Bäuml[2], and John M. Davis[4]

[2]Klinik für Psychiatrie und Psychotherapie der TU-München, Klinikum rechts der Isar, Ismaningerstr. 22, 81675 München, Germany; [3]Psychiatrische Klinik der Ludwig-Maximilians-Universität München; [4]Psychiatric Institute, Department of Psychiatry, University of Illinois at Chicago and Maryland Psychiatric Research Center, University of Maryland School of Medicine

It has been argued that the efficacy superiority found in meta-analyses for some of the atypical antipsychotics is an artifact of higher dropout rates due to side effects in the haloperidol group combined with last-observation-carried-forward (LOCF) analyses. We therefore reanalyzed a number of pivotal studies comparing new generation antipsychotics (NGAs) and conventional antipsychotics (CAs). A total of 5 studies ($n = 1271$) comparing amisulpride and 3 studies ($n = 2454$) comparing olanzapine with CAs were reanalyzed using original patient data. We applied 4 different models: LOCF, completer analysis, LOCF but excluding dropouts due to adverse events, and LOCF but excluding all dropouts with the exception of dropouts related to efficacy. Effect sizes expressed as standardized mean differences between NGAs and CAs based on the 4 different analysis models were compared. The overall results were not different irrespective of the model used. Single studies, however, showed higher effect sizes when LOCF instead of other models was used. Overall, it does not seem that higher dropout rates due to side effects in the haloperidol groups together with LOCF analyses consistently biased the results in favor of amisulpride and olanzapine. Because the results of the single studies, however, showed that this may occasionally be the case, future studies should look at the data from different angles applying sensitivity analyses, and they may use alternative statistics such as mixed models, which need to be developed further. Ultimately, strategies to reduce dropout rates are needed.

Key words: schizophrenia/olanzapine/amisulpride/bias/atypical antipsychotics

Introduction

There is currently a debate about the new generation antipsychotics (NGAs), so-called atypical antipsychotic drugs. While it is quite clear that the NGAs induce fewer extrapyramidal side effects (EPS) than high potency conventional antipsychotics (CAs) such as haloperidol,[1,2] meta-analyses have also shown that some of the new antipsychotics—amisulpride, clozapine, olanzapine, and risperidone—may be more efficacious than CAs.[2,3] Many have recently argued that this efficacy superiority of the NGA may just have been an artifact of the use of last-observation-carried-forward (LOCF) analyses, and this point has been developed most elegantly by Rosenheck.[4] The argument made concerns the following: when a patient terminates a study prematurely, in LOCF, his last observation is used (carried forward) as his endpoint evaluation. Given that the high potency CAs such as haloperidol induce more EPS than NGAs, it is assumed that more patients in the CA group leave the studies prematurely than in the NGA group. Thus, NGAs have more time to act on symptoms if the data are analyzed on an LOCF basis.[4]

We, therefore, reanalyzed original patient data from pivotal studies comparing amisulpride and olanzapine with CAs to assess whether the results differed if the calculations were based on LOCF or on 3 other models taking the dropout problem into account.

Methods

The Database

We reanalyzed original patient data of 5 published randomized controlled trials (RCT) that compared amisulpride with haloperidol/flupenthixol[5–9] and 3 studies comparing olanzapine with haloperidol[10–12] in a post hoc analysis. Important characteristics of the studies included are presented in table 1.

The 5 amisulpride studies represent the manufacturer's complete data set of trials comparing amisulpride with CAs with the exception of one trial[13] which did not use either the Brief Psychiatric Rating Scale (BPRS[14]) or the Positive and Negative Syndrome Scale (PANSS[15]) and could therefore not be included. A number of further old and small amisulpride vs CAs comparisons have been published, but the necessary original patient data are no

[1]To whom correspondence should be addressed; tel: +49-89-4140-4249, fax: +49-89-4140-4888, e-mail: stefan.leucht@lrz.tum.de.

Table 1. Characteristics of the Included Studies

Study	Antipsychotic Drugs Used (mg)	*n*	Weeks	Selected Patient Characteristics	Mean BPRS at Baseline
Amisulpride studies					
Möller et al[7]	AMI 800 HAL 20	95 96	6	Inpatients with paranoid, disorganized, or undifferentiated schizophrenia, *DSM-III-R*, BPRS psychotic subscore[a] ≥ 12, and at least 2 BPRS-psychosis items ≥ 4	61
Puech et al[8]	AMI 100, 400, 800, 1200 HAL 16	61, 64, 65, 65 64	4	Inpatients with acute exacerbations of paranoid, disorganized, or undifferentiated schizophrenia, *DSM-III-R*, BPRS psychotic subscore ≥ 12, and at least 2 BPRS-psychosis items ≥ 4	61
Colonna et al[5]	AMI 200–800 HAL 5–20	369 118	51	Inpatients or outpatients with acute exacerbations of paranoid, disorganized, or undifferentiated schizophrenia, *DSM-III-R*, and at least 2 BPRS-psychosis items ≥ 4	56
Carrière et al[6]	AMI 400–1200 HAL 10–30	97 105	17	Inpatients with paranoid schizophrenia or schizophreniform disorder, *DSM-IV*	65
Wetzel et al[9]	AMI 1000/600 FLU 25/15	70 62	6	Acutely admitted inpatients with paranoid or undifferentiated schizophrenia, *DSM-III-R*, BPRS total score ≥ 36 but no predominant negative symptoms defined as SANS composite score > 55	53
Olanzapine studies					
Beasley et al[10]	OLA (7.5 ± 2.5, 10 ± 2.5, 15 ± 2.5) HAL 15 ± 2.5 PBO	(65, 64, 69) 69 68	6	Inpatients with acute exacerbations of schizophrenia, *DSM-III-R*, BPRS total score ≥ 42	60
Tollefson et al[12]	OLA 5–20 HAL 5–20	1336 660	6	In- and outpatients with schizophrenia, schizophreniform disorder, or schizoaffective disorder, *DSM-III-R*, BPRS total score ≥ 36	52
Beasley et al[11]	OLA (1, 7.5 ± 2.5, 10 ± 2.5, 15 ± 2.5) HAL 15 ± 5	(88, 87, 86, 89) 81	6	Inpatients with acute exacerbations of schizophrenia, *DSM-III-R*, BPRS total score ≥ 42, CGI-S ≥ 4	59

Note: *n*, number of patients; AMI, amisulpride; CGI-S, Clinical Global Impressions – Severity Scale; *DSM-III-R(-IV), Diagnostic and Statistical Manual of Mental Disorders*, 3rd edition, revised (4th edition); FLU, flupenthixol; HAL, haloperidol; OLA, olanzapine; PBO, placebo; SANS, Scale for the Assessment of Negative Symptoms.
[a]Sum of conceptual disorganization, suspiciousness, hallucinatory behavior, and unusual thought content.

longer available because amisulpride has changed its owner several times.[16–21]

The 3 olanzapine studies are the largest trials comparing olanzapine with haloperidol and are the ones that were used for registering olanzapine at the Food and Drug Administration (FDA). All studies were randomized and all but one[5] were double blind. All trials examined patients with schizophrenia, schizoaffective disorder, or schizophreniform disorder according to *Diagnostic and Statistical Manual for Mental Disorders, Third Edition, Revised* (*DSM-III-R*) or (*DSM-IV*),[22,23] and with one exception,[6] all required various minimum scores as an inclusion criterion to assure that the patients had positive symptoms. One potentially ineffective 100 mg/d amisulpride dose group ($n = 61$) from the study by Puech and colleagues[8] and the potentially ineffective 1 mg and 5 ± 2.5 mg olanzapine dose groups were excluded a priori. Otherwise, if studies used several effective dose groups, these were pooled and considered as one group in the analysis.[8,10,11] The mean BPRS total score at baseline of all included patients in the amisulpride studies was 59.2 ± 12.5, and in the olanzapine studies it was 53.0 ± 11.2. The total number of patients in the amisulpride data set was 1271 (812 men, 459 women; 825 received amisulpride, 445 received haloperidol or flupenthixol; mean age 35.47 ± 10.83 years, weight

70.13 ± 14.39 kg, height 171 ± 9 cm). In all, 2454 patients were included in the olanzapine studies (1628 men, 826 women; 1645 received olanzapine, 809 received haloperidol; mean age 38.16 ± 11.23 years, weight 76.86 ± 17.10 kg, height 172 ± 10 cm).

Statistical Analysis

We used BPRS rather than the PANSS for our analyses because not all studies used the PANSS. When necessary, the BPRS items were extracted from the PANSS. All analyses were based on the difference between amisulpride/olanzapine and CAs concerning the change of the BPRS total score from baseline. In total, 4 different models were compared:

Model 1. The results were calculated based on an LOCF approach including all patients who had been randomized. In contrast to recent trials that often included only those patients who had at least one postbaseline rating, we included all randomized patients. This approach, a strict once randomized-analyzed model that is also applied in reviews of the Cochrane Schizophrenia Group,[24] should bias the findings even more in favor of the new antipsychotics because severe EPS such as acute dystonias induced by haloperidol frequently occur in the first days of treatment. Including only those patients with at least one postbaseline assessment would mean to exclude early dropouts due to haloperidol-induced EPS. In the amisulpride data set only 23 (1.8%) and in the olanzapine only 63 (2.6%) patients dropped out without a postbaseline rating, however, so that it is very unlikely that the "at least one postbaseline assessment approach" would have yielded different results. LOCF shows the efficacy of a drug on the whole-study population for the entire time when patients are on it. It is, however, a composite measure reflecting also factors other than efficacy. Assume that there are differences in dropouts due to side effects (or any other reason) between the drugs compared. Then the time the more side effect–associated drug had to act on symptoms will be shorter in the LOCF analysis. A major problem of LOCF is that it assumes that there would not have been any change after dropout if the assessments had been conducted (see later in "Discussion").

Model 2. Completers (CO) were analyzed, which means only those patients who were still in the study at the last planned evaluation. Thus, these were the patients who in both groups received the full length of the planned treatment. Such a CO analysis can tell what the maximum effect is that a clinician could expect from a medication in people who are willing to continue taking it. This can to a certain extent rule out the problem that LOCF is not necessarily conservative when atypical and typical antipsychotics are compared. Because the tolerability of the former is better, more people on typical antipsychotics may dropout, so that atypical antipsychotics have more time to act on the symptoms. In CO analyses, both groups received treatment for the same amount of time.

Model 3. This was an LOCF analysis excluding all those patients who dropped out of the studies due to adverse events. This model directly addresses the criticism that the higher rate of adverse events under treatment with CAs may bias the results when LOCF is applied.[4] Excluding all dropouts due to adverse events eliminates this factor. Thus, this model analyses the efficacy of 2 drugs for all patients who tolerate them.

Model 4. This was an LOCF approach excluding all dropouts with the exception of those due to inefficacy of treatment or due to remission. This model is very similar to the CO analysis, but it keeps the efficacy-related dropouts in the analysis. A pure CO analysis may be overly positive toward the efficacy of the less efficacious compound because patients who dropped out due to inefficacy are excluded from such a CO model. Therefore, this model was applied in addition. It analyzes the efficacy of the antipsychotics for those patients who completed the study or discontinued it due to an efficacy-related reason.

Effect sizes as statistical measures of the magnitude of the difference between amisulpride/olanzapine and CAs were used to compare the results based on the 4 different models. Effect sizes were expressed as standardized mean differences (SMDs). Various formulas for the calculation of SMDs are available. In the primary analysis, we used Hedges g and its standard error SE(g) according to the formulas:

$$g = \sqrt{\mathrm{F}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\left(1 - \frac{3}{4df_\mathrm{e} - 1}\right)$$

and

$$\mathrm{SE}(g) = \sqrt{\frac{(n_1 + n_2)(\mathrm{F} + 2df_\mathrm{e})}{2n_1 n_2 df_\mathrm{e}}},$$

where n_1 and n_2 are the number of patients in the amisulpride/olanzapine and CAs groups, respectively, F is the F value of the treatment contrast between amisulpride/olanzapine and CAs, and df_e the number of degrees of freedom of its error term.[25(p22–23),65] Both were taken from an analysis of covariance using treatment as a factor and baseline BPRS as covariate. In the pooled database, "study" was used as a further covariate. The significance of the individual effect sizes was calculated as z = effect size/SE(effect size). P values below .05 (2-tailed) were considered to show statistical significance. All analyses

Table 2. Amisulpride vs Conventional Antipsychotics—Results Based on 4 Different Test Conditions

	g	CI-low	CI-up	*P*	F	n_1	n_2
Model 1: LOCF							
Pooled results	0.27	0.15	0.39	0.00	21.15	825	445
Puech et al[8]	0.22	−0.06	0.51	0.12	2.41	194	64
Möller et al[7]	0.23	−0.05	0.52	0.11	2.66	95	96
Colonna et al[5]	0.32	0.11	0.53	0.00	9.24	368	118
Carrière et al[6]	0.35	0.07	0.63	0.01	6.24	97	105
Wetzel et al[9]	0.29	−0.05	0.64	0.09	2.88	71	62
Model 2: Completers only							
Pooled results	0.20	0.05	0.34	0.01	6.86	560	261
Puech et al[8]	0.07	−0.27	0.41	0.68	0.17	150	43
Möller et al[7]	0.26	−0.09	0.60	0.15	2.16	71	59
Colonna et al[5]	0.35	0.07	0.63	0.02	5.95	215	62
Carrière et al[6]	0.28	−0.07	0.62	0.12	2.53	72	59
Wetzel et al[9]	−0.03	−0.45	0.38	0.87	0.03	52	38
Model 3: LOCF after exclusion of dropouts due to adverse events							
Pooled results	0.23	0.11	0.35	0.00	13.57	775	380
Puech et al[8]	0.10	−0.21	0.40	0.53	0.40	186	54
Möller et al[7]	0.20	−0.10	0.49	0.19	1.77	92	86
Colonna et al[5]	0.37	0.15	0.58	0.00	11.63	441	106
Carrière et al[6]	0.26	−0.04	0.55	0.09	2.88	93	83
Wetzel et al[9]	0.13	−0.24	0.49	0.49	0.48	66	51
Model 4: LOCF after exclusion of all dropouts except inefficacy or recovery							
Pooled results	0.25	0.11	0.39	0.00	13.09	624	307
Puech et al[8]	−0.02	−0.34	0.30	0.90	0.02	170	47
Möller et al[7]	0.20	−0.12	0.52	0.22	1.52	81	68
Colonna et al[5]	0.49	0.23	0.75	0.00	13.92	237	77
Carrière et al[6]	0.31	−0.02	0.63	0.06	3.52	78	69
Wetzel et al[9]	0.19	−0.19	0.58	0.32	0.99	58	46

Note: *g*, the effect size Hedges *g*; F, F value; $df = n_1 + n_2 - 3$; n_1, number of patients in the amisulpride group; n_2, number of patients in the conventional antipsychotic group; CI-low, lower limit of 95% confidence interval; CI-up, upper limit of 95% confidence interval; *P*, *P* value; LOCF, last-observation-carried-forward.

were made with SPSS for Windows version 12.0 and Microsoft Excel 2000.

Results

Amisulpride

The results of the amisulpride studies are summarized in tables 2 and 3. In the pooled database, no dramatic differences between the 4 test conditions were found. Although the lowest effect size was found in the CO analysis (0.20), the range of the mean effect sizes was small (0.20–0.27), 95% confidence intervals overlapped broadly, and all 4 conditions yielded highly statistically significant superiorities of amisulpride ($P = 0.01$ or lower in all cases). The argument made in the literature that dropouts due to adverse events biased the results in favor of amisulpride was rejected because the exclusion of the dropouts due to adverse events in model 3 led to an only marginally lower effect size compared with the general LOCF model (0.23 vs 0.27). The lack of pronounced differences between models may be in part explained by a relatively small difference in terms of dropout rates between amisulpride and comparators of only 9.7%, although with the exception of uncooperativeness and recovery of the patients, the dropout rates for all the specific reasons were lower in the amisulpride group (see table 3).

When the single studies were analyzed separately, a rather heterogeneous picture was found, however. For example, Wetzel et al[9] yielded the same directions of effect in the 4 models as in the pooled analysis. Again, the CO analysis (model 2) showed the lowest effect size finding even a minimal trend in favor of the CA (effect size: −0.03). In Puech et al,[8] however, the lowest effect size was found in model 4 (LOCF but excluding all dropouts with the exception of inefficacy), possibly because in

Table 3. Dropout Rates in the Amisulpride Studies

	Amisulpride		Haloperidol[a]		All	
	n	%	*n*	%	*N*	%
Pooled studies						
Completed	590	66.6	253	56.9	843	63.3
Lost to follow-up	18	2.0	10	2.2	29	2.2
Inefficacy	83	9.4	52	11.7	135	10.1
Adverse event	50	5.6	65	14.6	115	8.6
Uncooperative	112	12.6	48	10.8	160	12.0
Recovery	4	0.5	2	0.4	6	0.5
Other	29	3.3	15	3.4	44	3.3
Total	886	100.0	445	100.0	1332	100.0
Wetzel et al[9]						
Completed	51	71.8	37	59.7	88	66.2
Lost to follow-up	0	0.0	0	0.0	0	0.0
Inefficacy	5	7.0	8	12.9	13	9.8
Adverse event	5	7.0	11	17.7	16	12.0
Uncooperative	6	8.5	4	6.5	10	7.5
Recovery	2	2.8	1	1.6	3	2.3
Other	2	2.8	1	1.6	3	2.3
Total	71	100.0	62	100.0	133	100.0
Puech et al[8]						
Completed	149	76.8	43	67.2	192	74.4
Lost to follow-up	3	1.5	0	0.0	3	1.2
Inefficacy	20	10.3	4	6.3	24	9.3
Adverse event	8	4.1	10	15.6	18	7.0
Uncooperative	8	4.1	7	10.9	15	5.8
Recovery	1	0.5	0	0.0	1	0.4
Other	5	2.6	0	0.0	5	1.9
Total	194	100.0	64	100.0	258	100.0
Möller et al[7]						
Completed	70	73.7	57	59.4	127	66.5
Lost to follow-up	1	1.1	2	2.1	3	1.6
Inefficacy	11	11.6	11	11.5	22	11.5
Adverse event	3	3.2	10	10.4	13	6.8
Uncooperative	8	8.4	11	11.5	19	9.9
Recovery	0	0.0	0	0.0	0	0.0
Other	2	2.1	5	5.2	7	3.7
Total	95	100.0	96	100.0	191	100.0
Colonna et al[5]						
Completed	203	55.2	57	48.3	260	53.4
Lost to follow-up	10	2.7	5	4.2	16	3.3
Inefficacy	33	9.0	20	16.9	53	10.9
Adverse event	30	8.2	12	10.2	42	8.6
Uncooperative	77	20.9	17	14.4	94	19.3
Recovery	1	0.3	0	0.0	1	0.2
Other	14	3.8	7	5.9	21	4.3
Total	368	100.0	118	100.0	487	100.0
Carrière et al[6]						
Completed	72	74.2	59	56.2	131	64.9
Lost to follow-up	2	2.1	3	2.9	5	2.5
Inefficacy	6	6.2	9	8.6	15	7.4
Adverse event	4	4.1	22	21.0	26	12.9
Uncooperative	8	8.2	9	8.6	17	8.4
Recovery	0	0.0	1	1.0	1	0.5
Other	5	5.2	2	1.9	7	3.5
Total	97	100.0	105	100.0	202	100.0

Note: *n*, number of patients.
[a]In Wetzel et al,[9] the comparator was flupenthixol.

contrast to the other studies, more patients in the amisulpride group than in the haloperidol group dropped out due to inefficacy. In Möller et al,[7] the highest effect size was found in the CO analysis, although again the results of the 4 models did not differ greatly. In the Colonna et al,[5] the conventional LOCF model yielded the lowest effect size so that again the efficacy superiority of amisulpride cannot be simply explained by LOCF. Carrière et al[6] was the study with the highest difference in terms of dropouts due to adverse events between amisulpride (4.1%) and haloperidol (21%). This may in part explain why here the effect size in model 3 excluding dropouts due to adverse events was indeed the lowest one.

Olanzapine

The results of the olanzapine studies are summarized in tables 4 and 5. Pooling all studies, the difference in terms of global dropout rates between olanzapine and haloperidol was somewhat more pronounced (17%) than in the amisulpride database (9.3%). Fewer patients treated with olanzapine than with haloperidol dropped out in all major categories. However, pooling the studies, the results based on the 4 models did not differ to any important extent with mean effect sizes between 0.23 and 0.28 and all *P* values below 0.001. Although the lowest effect size was indeed found in model 3 excluding dropouts due to adverse events, it is noteworthy that the highest difference between olanzapine and haloperidol was found in the CO model 2 rather than in the LOCF model. Therefore, in contrast to the assumption that the inclusion of dropouts in LOCF affected the results in favor of olanzapine, it seems the longer olanzapine was given, the higher its efficacy superiority became.

Looking at the single studies, again a heterogeneous pattern was found. While in Tollefson et al[12] the effect sizes of all 4 models were virtually identical, in Beasley et al,[11] the highest effect size was found in model 4 excluding dropouts due to inefficacy and the lowest one in model 3 excluding dropouts due to adverse events. In Beasley et al,[10] the highest effect size was found in the COs and the lowest one in model 4 excluding dropouts due to inefficacy.

Discussion

Our analyses show that higher rates of dropouts due to adverse events in the haloperidol groups and using LOCF alone do not explain the statistically significantly higher efficacy of amisulpride and olanzapine compared with CAs found in meta-analyses.[2,3,26] The pooled effect sizes found in all models analyzed were similar and showed statistically significant superiorities of the 2 NGAs even when dropouts due to adverse events were removed from the analyses.

Table 4. Olanzapine vs Haloperidol—Results Based on 4 Different Test Conditions

	g	CI-low	CI-up	*P*	F	n_1	n_2
Model 1: LOCF							
Pooled results	0.25	0.16	0.33	0.00	33.03	1639	805
Beasley et al[11]	0.16	−0.10	0.43	0.23	1.46	174	80
Beasley et al[10]	0.01	−0.29	0.30	0.95	0.00	130	67
Tollefson et al[12]	0.29	0.20	0.39	0.00	38.15	1335	658
Model 2: Completers only							
Pooled results	0.28	0.17	0.40	0.00	22.77	1068	387
Beasley et al[11]	0.19	−0.16	0.55	0.29	1.16	107	43
Beasley et al[10]	0.36	−0.08	0.80	0.11	2.62	62	30
Tollefson et al[12]	0.29	0.16	0.42	0.00	19.28	899	314
Model 3: LOCF after exclusion of dropouts due to adverse events							
Pooled results	0.23	0.14	0.32	0.00	26.82	1560	739
Beasley et al[11]	0.06	−0.23	0.34	0.69	0.16	160	68
Beasley et al[10]	−0.05	−0.36	0.26	0.75	0.10	125	61
Tollefson et al[12]	0.30	0.20	0.39	0.00	36.04	1275	610
Model 4: LOCF after exclusion of all dropouts except inefficacy or recovery							
Pooled results	0.25	0.16	0.35	0.00	27.56	1399	624
Beasley et al[11]	0.23	−0.08	0.54	0.14	2.20	134	58
Beasley et al[10]	−0.08	−0.42	0.27	0.67	0.18	100	47
Tollefson et al[12]	0.30	0.19	0.40	0.00	32.05	1165	519

Note: *g*, the effect size Hedges *g*; F, F value; $df = n_1 + n_2 - 3$; n_1, number of patients in the olanzapine group; n_2, number of patients in the haloperidol group; CI-low, lower limit of 95% confidence interval; CI-up, upper limit of 95% confidence interval; *P*, *P* value; LOCF, last-observation-carried-forward.

The argument that LOCF biases study results in favor of NGAs assumes that adverse events are the main reason why patients terminate trials prematurely. In the olanzapine studies, however, most patients dropped out due to lack of efficacy because it has also been reported for a larger set of studies.[27] In the amisulpride studies, the most common dropout reason was patients' uncooperativeness, although the rates of inefficacy of treatment and adverse events were in the same range. Obviously, the effect of dropouts due to adverse events in an LOCF analysis depends on the proportion of patients terminating a study for this reason and whether the proportions differed between NGA and CA groups. For example, in the pooled amisulpride studies, 2.6 times more patients treated with CAs dropped out due to adverse events (5.6% vs 14.6%). If these numbers and their difference had been higher, they should have had a stronger impact on the results.

A number of methodological limitations of our analysis must be discussed. How representative are our results? For amisulpride, we analyzed all randomized comparisons with CAs available in the manufacturers' database. Only some older studies for which original patient data are no longer available due to changes of ownership of amisulpride could not be included. The latter studies were small and therefore did not have a strong impact on the results of meta-analyses.[2,3] The olanzapine studies are not complete, but the 3 studies included are the largest comparisons of olanzapine with haloperidol and they were used for registration at the FDA. We want to emphasize that we did not specifically select favorable studies. Tollefson et al[12] was especially important for our purposes. It has recently been argued by Rosenheck[4] that in contrast to their own trial,[28] olanzapine was more efficacious than haloperidol in Tollefson et al[12] because in the latter, prophylactic antiparkinson medication was not used. Therefore, Rosenheck[4] speculated that many patients in the haloperidol group dropped out due to adverse events, and LOCF biased the results. However, the results in Tollefson et al[12] were virtually identical irrespective of the model we used. Only 4.5% (olanzapine) vs 7.3% (haloperidol) patients dropped out to adverse events so that this factor did not explain olanzapine's efficacy superiority. Another reason explaining the different results of Rosenheck et al[28] and Tollefson et al[12] may be that the population in the latter study was more chronic, which finds its expression in the minimal reduction of symptoms (about 10 PANSS points) during the 1-year course of the study.

We hasten to emphasize that our analysis focused solely on the effects of dropouts in combination with LOCF. A number of other biases might have potentially

Table 5. Dropout Rates in the Olanzapine Studies

	Olanzapine		Haloperidol		Total	
	n	%	*n*	%	*N*	%
Pooled results						
Complete	1056	64.2	382	47.2	1438	58.6
Adverse event	79	4.8	66	8.2	145	5.9
Criteria not met/ compliance	60	3.6	33	4.1	93	3.8
Lack of efficacy	342	20.8	246	30.4	588	24.0
Lost to follow-up	22	1.3	18	2.2	40	1.6
Patient decision	77	4.7	62	7.7	139	5.7
Satisfactory response	5	0.3	0	0.0	5	0.2
Sponsor decision	4	0.3	2	0.2	6	0.2
Total	1645	100.0	809	100.0	2454	100.0
Beasley et al[11]						
Completed	108	61.7	43	53.1	151	59.0
Adverse event	14	8.0	12	14.8	26	10.2
Criteria not met/ compliance	8	4.6	2	2.5	10	3.9
Lack of efficacy	22	12.6	16	19.8	38	14.8
Lost to follow-up	3	1.7	2	2.5	5	2.0
Patient decision	15	8.6	6	7.4	21	8.2
Satisfactory response	5	2.9	0	0.0	5	2.0
Total	175	100.0	81	100.0	256	100.0
Beasley et al[10]						
Completed	60	45.1	30	43.5	90	44.6
Adverse event	5	3.8	6	8.7	11	5.4
Criteria not met/ compliance	8	6.0	2	2.9	10	5.0
Lack of efficacy	42	31.6	19	27.5	61	30.2
Lost to follow-up	4	3.0	5	7.2	9	4.5
Patient decision	14	10.5	7	10.1	21	10.4
Total	133	100.0	69	100.0	202	100.0
Tollefson et al[12]						
Completed	888	66.4	309	46.9	1197	60.0
Adverse event	60	4.5	48	7.3	108	5.4
Criteria not met/ compliance	44	3.3	29	4.4	73	3.7
Lack of efficacy	278	20.8	211	32.0	489	24.5
Lost to follow-up	15	1.1	11	1.7	26	1.3
Patient decision	48	3.6	49	7.4	97	4.9
Sponsor decision	4	0.3	2	0.3	6	0.3
Total	1337	100.0	659	100.0	1996	100.0

led to artificial efficacy superiorities of the new antipsychotics. For example, side effects could have played a role. Haloperidol induces EPS such as parkinsonism or akathisia that can mimic negative symptoms or agitation. In the absence of prophylactic antiparkinson medication, these side effects can have artificially inflated the BPRS ratings of the haloperidol-treated patients.[4] In contrast to Tollefson and colleagues,[12] Rosenheck et al[28] did use prophylactic antiparkinson medication. Another potential source of bias is that many patients in the studies had previously been partial nonresponders to haloperidol. If such patients are randomized to haloperidol, their chance to respond is lower than if they are randomized to a new compound with a different receptor-binding profile. Indeed, first-episode studies have usually shown small or no efficacy superiorities of NGAs,[29–31] but first-episode patients respond so well that ceiling effects are a problem. To assess these and other factors, further analyses are needed.

A further important limitation of our analysis is that the reasons for dropping out are not validated in such studies. In our experience, the description of the dropout reason is often straightforward. But there are also occasions when, eg, a patient experiences a side effect without complaining about it, simply saying he no longer wants to be part of the study, and this is classified as "consent withdrawn," or a patient decides to participate in a study because he hopes to receive treatment with a new antipsychotic. If he then realizes that he is receiving haloperidol because he experiences EPS, he may withdraw from the study and give some other reason. How often such problems occurred in our studies is unclear. This is not a problem for the newer mixed-effects model techniques because they ignore the reason for withdrawal anyway and simply include all cases as far as they go (see below). This issue has received little research attention. Perhaps a probing question could sharpen these data and increase the precision of measuring this effect.

We also do not conclude that LOCF should be the primary model used to analyze antipsychotic drug trials. Regulatory authorities such as the FDA preferred LOCF in the last decade, but they also asked for sensitivity analyses. In our individual studies, a variety of patterns were seen and occasionally CO analyses yielded the worst result. We have the impression that currently the choice of LOCF or CO analysis is often based on which model showed the best results for the manufacturer's drug.[32] There is also a trend to use more complex mixed-effects designs to analyze antipsychotic drug trials. Their general idea is to use information from the observed data to provide information on the missing data, but missing data are not explicitly imputed.[33] We did not apply such an approach because our question was not what the "best" way of analyzing such studies is but only whether higher dropout rates due to adverse events in the haloperidol group combined with the use of LOCF biased the findings in favor of atypical antipsychotics. Mixed-effects models would not have been able to answer this specific question because missing values are estimated in these models, irrespective of the reason for dropout. Thus, for the purpose of our specific question, our "pragmatic" method was more appropriate.

Although we go beyond the scope of our article here, we want to briefly summarize the current debate around newer statistical methods for randomized clinical trials. The general problem is how dropouts in RCTs need to be addressed. Theoretically, there are 3 different situations: (1) Data are missing completely at random (MCAR) if the missingness is related neither to observed

nor to unobserved outcomes. (2) Data are missing at random (MAR) if the missingness is explained by observed outcomes but not by unobserved outcomes. (3) Data are not missing at random if the missingness depends on the unobserved outcomes.[33] Leon et al[34] classified the multiple strategies for coping with dropouts into 3 groups—"complete case analysis," "imputation strategies," and "analyses of incomplete data" (mixed-effects models, pattern mixture models, and propensity adjustment). "COs-only" analyses assume MCAR, although this assumption is unlikely for RCTs in schizophrenia. Nevertheless, they can be useful sensitivity analyses as carried out in our analysis.[34] LOCF is a crude method of imputation which also assumes MCAR and which assumes that the subjects' responses would have been constant after dropping out. This is unlikely to be true. Indeed, LOCF has been shown sometimes to overestimate and sometimes to underestimate treatment effects.[33] For example, in a recent first-episode study in schizophrenia, LOCF did not find olanzapine to be more efficacious than haloperidol, while a mixed-effects model did; thus, LOCF was more conservative.[30] Much more sophisticated imputation strategies than LOCF exist, such as "multiple imputation." Among the analyses of incomplete data, mixed-effects models have recently been advocated for.[33] Here, the observed information is used to provide information about the missing data. In contrast to LOCF, mixed models assume only MAR; they have been shown to be more robust in the face of the biases of missing data and to control better for type I and type II errors, although in many occasions the results were the same.[33] A number of questions on mixed-effects models remain open. The phenomena we have been exploring, namely, the different rates of dropouts due to adverse events, may affect certain mixed models. We plan a parallel analysis using mixed models to examine this question. In contrast to LOCF or CO analyses, mixed models are hard to understand intuitively. There are a number of different approaches, but authors rarely explain among which models they chose, why they decided on a specific one, and to what extent the assumptions of the model were met. This leads to an uncertainty in the reader, and the concern about potential misuse (by picking the model that fits one's purposes best) has been expressed.[35] Guidelines describing the different models and the situations for which they are best suited along with standards for reporting are needed to enhance the transparency of these complex methods.[35] Again, for our specific question, they were not the most appropriate tool.

We conclude that higher dropout rates due to adverse events combined with using LOCF did not consistently bias the results of RCTs in favor of amisulpride and olanzapine. The results of single studies, however, varied. Future analyses should look at the data from different angles. A further elaboration of alternative statistical approaches, such as mixed-effects models, for dealing with dropouts is needed. Finally, strategies to reduce the dropout rates in such studies are urgently needed.

Acknowledgments

We wish to thank EliLilly and SanofiAventis for allowing us to use their original patient data. This study was supported by the American Psychiatric Association/ AstraZeneca Young Minds in Psychiatry Award 2004. The study received no other funding.

References

1. Leucht S, Pitschel-Walz G, Abraham D, Kissling W. Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta-analysis of randomized controlled trials. *Schizophr Res*. 1999;35:51–68.
2. Leucht S, Pitschel-Walz G, Engel R, Kissling W. Amisulpride—an unusual atypical antipsychotic. A meta-analysis of randomized controlled trials. *Am J Psychiatry*. 2002;159:180–190.
3. Davis JM, Chen N, Glick ID. A meta-analysis of the efficacy of second-generation antipsychotics. *Arch Gen Psychiatry*. 2003;60:553–564.
4. Rosenheck RA. Effectiveness versus efficacy of second-generation antipsychotics: haloperidol without anticholinergics as a comparator. *Psychiatr Serv*. 2005;56:85–92.
5. Colonna L, Saleem P, Dondey-Nouvel L, Rein W. Amisulpride study group. Long-term safety and efficacy of amisulpride in subchronic or chronic schizophrenia. *Int Clin Psychopharmacol*. 2000;15:13–22.
6. Carrière P, Bonhomme D, Lempérière T. Amisulpride has superior benefit: risk profile to haloperidol in schizophrenia: results of a multicentre, double-blind study (the Amisulpride study group). *Eur Psychiatry*. 2000;15:321–329.
7. Möller HJ, Boyer P, Fleurot O, Rein W. Improvement of acute exacerbations of schizophrenia with amisulpride: a comparison with haloperidol. *Psychopharmacology*. 1997;132:396–401.
8. Puech A, Fleurot O, Rein W. Amisulpride, an atypical antipsychotic, in the treatment of acute episodes of schizophrenia: a dose-ranging study vs haloperidol. *Acta Psychiatr Scand*. 1998;98:65–72.
9. Wetzel H, Grunder G, Hillert A, et al. Amisulpride versus flupentixol in schizophrenia with predominantly positive symptomatology—a double-blind controlled study comparing a selective D-2-like antagonist to a mixed D-1-/D-2-like antagonist. *Psychopharmacology*. 1998;137:223–232.
10. Beasley CM, Tollefson GD, Tran P, et al. Olanzapine versus haloperidol and placebo. Acute phase results of the american double-blind olanzapine trial. *Neuropsychopharmacology*. 1996;14:111–123.
11. Beasley CM, Hamilton SH, Crawford AM, et al. Olanzapine versus haloperidol: acute phase results of the international double-blind olanzapine trial. *Eur Neuropsychopharmacol*. 1997;7:125–137.
12. Tollefson GD, Beasley CM, Tran PV, et al. Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: results of an international collaborative trial. *Am J Psychiatry*. 1997;154:457–465.

13. Speller JC, Barnes TRE, Curson DA, Pantelis C, Alberts JL. One-year, low-dose neuroleptic study of in-patients with chronic schizophrenia characterised by persistent negative symptoms—amisulpride v. haloperidol. *Br J Psychiatry*. 1997;171:564–568.
14. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep*. 1962;10:790–812.
15. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13:261–275.
16. Klein HE, Dieterle D, Rüther E, Eben E, Nedopil N, Hippius H. A double-blind comparison of amisulpride vs haloperidol in acute schizophrenic patients. In: Pichot P, Berner P, Wolf R, Thau K, eds. *Psychiatry, the State of the Art.* Cambridge, Mass: Plenum Press; 1985:687–691.
17. Pichot P, Boyer P. Etude multicentrique controlée en double insu: amisulpride (Solian 200) versus halopéridol à forte dose dans les états psychotiques aigus. *Ann Psychiatr*. 1988;3(3):326–332.
18. Pichot P, Boyer P. Controlled double-blind multi-centre trial of low dose amisulpride versus fluphenazine in the treatment of the negative syndrome of chronic schizophrenia. *Amisulpride.* Paris, France: Expansion scientifique francaise; 1989:125–137.
19. Costa e Silva JA. Comparative double-blind study of amisulpride versus haloperidol in the treatment of acute psychotic states. *Amisulpride.* Paris, France: Expansion scientifique francaise; 1989:93–104.
20. Delcker A, Schoon ML, Oczkowski B, Gaertner HJ. Amisulpride versus haloperidol in treatment of schizophrenic patients—results of a double-blind study. *Pharmacopsychiatry*. 1990;23:125–130.
21. Ziegler B. Study of the efficacy of a substituted benzamide amisulpride, versus haloperidol, in productive schizophrenia. *Amisulpride.* Paris, France: Expansion scientifique francaise; 1989:73–81.
22. American Psychiatric Association. *Diagnostic and Statistical Manual for Mental Disorders*. third revision, revised (*DSM-III-R*). Washington, DC: American Psychiatric Association; 1987.
23. American Psychiatric Association. *Diagnostic and Statistical Manual for Mental Disorders,* 4th ed. Washington, DC: American Psychiatric Association; 1994.
24. Adams CE, Coutinho E, Duggan L, Leucht S, Srisurapanont M, Tharyan P. *Cochrane Schizophrenia Group. The Cochrane Library.* Chichester, UK: John Wiley & Sons Ltd; 2005.
25. Rosenthal R. *Meta-analytic Procedures for Social Research.* 2nd ed. New York, NY: Sage Publications; 1991.
26. Geddes J, Freemantle N, Harrison P, Bebbington P. Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. *BMJ*. 2000;321: 1371–1376.
27. Kinon BJ, Liu-Seifert H, Adams DH. Predominance of psychiatric-based reasons for antipsychotic treatment discontinuation. *Biol Psychiatry*. 2005;57:105S–106S.
28. Rosenheck R, Perlick D, Bingham S, et al. Effectiveness and cost of olanzapine and haloperidol in the treatment of schizophrenia—a randomized controlled trial. *JAMA*. 2003;290:2693–2702.
29. Lieberman JA, Phillips M, Gu H, et al. Atypical and conventional antipsychotic drugs in treatment-naive first-episode schizophrenia: a 52-week randomized trial of clozapine vs chlorpromazine. *Neuropsychopharmacology*. 2003;28: 995–1003.
30. Lieberman JA, Tollefson G, Tohen M, et al. Comparative efficacy and safety of atypical and conventional antipsychotic drugs in first-episode psychosis: a randomized, double-blind trial of olanzapine versus haloperidol. *Am J Psychiatry*. 2003;160:1396–1404.
31. Emsley RA. Risperidone Working Group. Risperidone in the treatment of first-episode psychotic patients: a double-blind multicenter study. *Schizophr Bull*. 1999;25:721–729.
32. Heres S, Davis JM, Maino K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine and quetiapine beats olanzapine again. An exploratory analysis of head-to-head comparisons of second generation antipsychotics. *Am J Psychiatry*. 2006;163: 180–194.
33. Mallinckrodt CH, Watkin JG, Molenberghs G, Carroll RJ. Choice of the primary analysis in longitudinal clinical trials. *Pharma Stat*. 2004;3(3):161–169.
34. Leon AC, Mallinckrodt CH, Chuang-Stein CC, Archibald DG, Archer GE, Chartier K. Attrition in randomized clinical trials: methodological issues in psychopharmacology. *Biol Psychiatry*. 2006;59(11):1001–5.
35. Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the *Archives of General Psychiatry*. *Arch Gen Psychiatry*. 2004;61:310–7.