Quality of Reporting of Randomized Controlled Trials of Pharmacologic Treatment of Bipolar Disorders: A Systematic Review

Daniel Strech, MD, PhD; Bettina Soltmann, PhD; Beate Weikert, MD; Michael Bauer, MD, PhD; and Andrea Pfennig, MD, MSc

ABSTRACT

Objective: This study aimed to assess (1) the quality of reporting of randomized controlled trials of pharmacologic treatment of bipolar disorder, (2) the potential improvement in quality of reporting over time, and (3) differences in quality of reporting between journals that endorse or do not endorse the Uniform Requirements for Manuscripts Submitted to Biomedical Journals developed by the International Committee of Medical Journal Editors.

Data Sources: A systematic literature search was done to identify all randomized controlled trials published between 2000 and 2008 relevant to the pharmacologic treatment of bipolar disorder. The search strategy of the published National Institute for Health and Clinical Excellence guideline for management of bipolar disorders was used and adapted. All included and excluded clinical trials mentioned in the guideline and published from 2000 onward were reviewed for eligibility. For an update search from July 2004 through December 2008, an adapted search strategy was used in MEDLINE, EMBASE, PsycINFO, CINAHL, Ovid, and Cochrane Central Register of Controlled Trials. Titles and abstracts were scanned for relevance, and full texts were ordered in case of uncertainty to maximize sensitivity. Reference lists of retrieved systematic reviews were checked.

Study Selection: All full texts were checked for eligibility. Only relevant randomized controlled trials published between 2000 and 2008 were included. Abstracts, randomized controlled trials published before 2000, nonrandomized clinical studies, pooled analyses, editorials, reviews, case reports, observational studies, and unpublished reports were excluded.

Data Extraction: A checklist based on the Consolidated Standards of Reporting Trials (CONSORT) statement was used to assess quality of reporting of all included studies.

Results: A total of 105 randomized controlled trials were included in the analysis. Of the 72 applicable checklist items, 42% were generally reported adequately and 25% inadequately. Reporting was especially poor for randomization procedures, with, for example, 16% of studies defining generation of random allocation sequence and 15% defining method of allocation concealment. Inadequate randomization increases the potential for bias to influence the final results. Authors of clinical guidelines or health technology assessments are forced to exclude or downgrade trials with inadequate reporting on randomization. Also, information with essential clinical relevance was generally reported inadequately, such as the effect size (in 18% of studies) and the number needed to treat (in 8% of studies). Both effect measures are more important for clinicians than individual point estimates that have been reported adequately. No consistent trend could be shown for improvement in quality of reporting over time or for reporting of essential methodological items differently in journals that endorse the Uniform Requirements for Manuscripts (URM). The reporting of information on clinical relevance and generalizability of results, however, showed a consistent trend toward better reporting in journals endorsing the URM, with significant differences for the reporting of secondary outcomes (100% vs 89.9%; $P = .03$) and adverse events (93.2% vs 73.8%; $P = .011$) and interpretation of results with regard to totality of data (30.2% vs 11.5%; $P = .029$).

Conclusions: Our findings suggest that, while some trial-related information is well reported, a good part of the reporting quality of randomized controlled trials in bipolar disorder falls well below the required and also practically feasible level for many aspects essential for adequate interpretation of methodological quality and clinical relevance. Authors should be further encouraged to follow the CONSORT criteria.

J Clin Psychiatry 2011;72(9):1214–1221

Submitted: *April 7, 2010; accepted June 18, 2010.*
Online ahead of print: *January 25, 2011 (doi:10.4088/JCP.10r06166yel).*
Corresponding author: *Daniel Strech, MD, PhD, Hannover Medical School, CELLS–Centre for Ethics and Law in the Life Sciences, Institute for History, Ethics and Philosophy of Medicine, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany (strech.daniel@mh-hannover.de).*

Besides clinical experience and patient preferences, valid and applicable clinical evidence is needed to guide health care decision-making. Randomized controlled trials are widely accepted as the "gold standard" for accumulating strong evidence for any health care intervention.[1] The quality and the reporting of randomized controlled trials, however, can be suboptimal. Within the design of randomized controlled trials there is the inherent risk of bias skewing results at various stages and minimizing internal and external validity.[2] First, there is empirical evidence to suggest that lack of, or inadequate attention to, random allocation, allocation concealment, blinding, and intention-to-treat principle can lead to bias.[3,4] Second, setting, participants, demographic data, and comedication, for example, can limit the generalizability of the trial results.[5,6] There is also increasing evidence on the selective reporting in clinical trial findings, with some recent disreputable examples in pharmacologic treatment for depression and other psychiatric disorders.[7–10]

A lack of standardization in reporting trial methods and results can limit the interpretation of results for individual medical decision-making and the pooling of trial results for meta-analysis. Furthermore, authors of evidence-based clinical guidelines or health technology assessment are forced to exclude or downgrade trials with low quality of reporting.[11]

In an attempt to improve reporting of trials, the Consolidated Standards of Reporting Trials (CONSORT) statement was first published in 1996 and revised in 2001 and 2010 on the basis of the empirical evidence regarding trial bias (www.consort-statement.org).[12,13] The CONSORT statement has been adopted by the Uniform Requirements for Manuscripts Submitted to Biomedical Journals (URM) developed by the International Committee of Medical Journal Editors.[14] Journals adhering to the URM are encouraged to state in their instructions to authors that their requirements are in accordance with the URM and to cite this version.[14] Journals following the URM (and therefore following the CONSORT statement) are, for example, *Archives of General Psychiatry*, *The American Journal of Psychiatry*, *The Journal of Clinical Psychiatry*, *BMJ*, *JAMA*, *The New England Journal of Medicine*, and *Lancet*.

© Copyright 2011 Physicians Postgraduate Press, Inc.

During the last decade, many studies have assessed the quality of reporting of randomized trials within several clinical domains.[15,16] There are very limited data, however, examining the quality of reporting of clinical trials in the field of psychiatry. Thornley et al[17] assessed the quality of 2,000 controlled trials in schizophrenia published up to 1997 by using a summary measure (the Jadad scale[18]) based on each trial's description of randomization, blinding, and withdrawal from treatment. Cipriani et al[19] assessed the quality of 39 randomized controlled trials published up to 2004 that compared fluoxetine with any other antidepressant agent by using the Jadad scale[18] and a subset of the CONSORT statement (randomization, allocation concealment, and power calculation) as proxy measures of trial quality, according to Schulz and Grimes.[20] Besides the proxy measures of trial quality, both reviews provided no further information on quality of reporting. Recently, Han et al[21] compared psychiatric randomized controlled trials published up to 2007 in 7 leading medical journals to evaluate whether the CONSORT statement has improved the completeness of randomized controlled trial reporting. This recent review,[21] however, has several methodological shortcomings. First, this review analyzes only a subset of core clinical journals and therefore has limited generalizability. Second, and most important, it is unclear how the trial rating was conducted, for Han et al[21] used a dichotomous scale (described vs not described) to rate CONSORT items. However, a good part of the CONSORT items ask for several methodological aspects that cannot be answered by a dichotomous scale.[12] Item 10 (implementation), for example, addresses the description of (1) who generated the allocation sequence, (2) who enrolled participants, and (3) who assigned participants to their groups. To rate item 10 as "described" indicates that all independent aspects have been described adequately. The rather optimistic findings in Han et al[21] with respect to item 10 and others indicate that they did not apply this rather rigorous rule of rating. As a result, the findings of Han et al[21] differ importantly for several CONSORT items (in a rather optimistic direction) in comparison with studies on other clinical domains,[15] as well as with our findings.

In this study, the focus was on randomized controlled trials of pharmacologic treatment of bipolar disorders. Studying quality of reporting in this area is important for a number of reasons. First, bipolar disorder is a severe, often recurrent psychiatric disorder with the need for long-term treatment management. With a life-time prevalence between 1.0% and 6.5%[22–24] and an age at onset of around 17 years,[25] bipolar disorder is rather frequently encountered in clinical practice and results in enormous health and economic burden.[26] Clinicians and authors of clinical guidelines and health technology assessments should be aware of reporting quality in bipolar disorder trials so as to better interpret the findings and apply them to patient care.

The objective of this study was to comprehensively assess the quality of reporting of randomized controlled trials of pharmacologic treatment of bipolar disorder by using a checklist based on explanations of and elaborations on the CONSORT 2001 statement.[2] The publication periods of January 2000 through June 2004 and July 2004 through December 2008 were compared to study potential improvement in quality of reporting. Trials published in journals that endorse the URM were compared to those published in journals that do not endorse the URM to study potential differences in quality of reporting.

METHOD

Literature Search

This study was conducted as an independent project within the development of the German evidence-based and consensus-based S3 Guideline for Diagnosis and Therapy of Bipolar Disorders.[27] For the independent project, a systematic literature search was used to identify all randomized controlled trials published between 2000 and 2008 relevant to the pharmacologic treatment of bipolar disorders. For this purpose, the search strategy of the published National Institute for Health and Clinical Excellence (NICE) guideline for management of bipolar disorders[28] was used and adapted. All included and excluded clinical trials mentioned in the NICE guideline and published from 2000 on were reviewed for eligibility. For the update search from January 2005 through December 2008, the adapted search strategy (eAppendix 1) was used in MEDLINE, EMBASE, PsycINFO, CINAHL, Ovid, and Cochrane Central Register of Controlled Trials. Titles and abstracts were scanned for relevance. Full texts were ordered in case of uncertainty to maximize sensitivity. Reference lists of retrieved systematic reviews were checked. All full texts were checked for eligibility. Only relevant randomized controlled trials published between 2000 and 2008 were included. Abstracts, randomized controlled trials published before 2000, nonrandomized clinical studies, pooled analyses, editorials, reviews, case reports, observational studies, and unpublished reports were excluded.

Inclusion Criteria

Participants in the studies had to be adults aged 18 years and above with a diagnosis of bipolar disorder. Studies including patients with other diagnoses were included only when less than 10% of the population had no other diagnoses than unipolar depression, schizoaffective disorder, or cyclothymia, or when baseline data for relevant clinical variables and results were presented for the bipolar subgroup separately.

Checklist

While the CONSORT statement alone is not applicable as a checklist to rate the reporting of trials, we developed checklist items by referring to the explanations for and elaborations on each CONSORT item given in the extended CONSORT 2001 statement and at the CONSORT Web site (www.consort-statement.org; retrieved on January 3, 2009). All checklist items are presented in Table 1. Before assessing the individual publications, we defined items of essential methodological and clinical relevance that should be adequately reported. Essential methodological information

Table 1. Checklist Items Based on the CONSORT Statement and Their Reporting in Randomized Controlled Trials on Bipolar Disorder

CONSORT Checklist	Item Number and Description	Bipolar Disorder Trials Reporting, %	
		Yes (if applicable)	Not Applicable
1. Title/abstract	1a Identifies study as a randomized trial	95	0
	1b Abstract has a structured format	79	0
2. Introduction	2a Reports the evidence of the benefits and harms of any active intervention included in a trial	55	0
	2b Includes references to previous trials or to a systematic review of previous trials, or a note of the absence of such trials	100	0
3. Participants	3a Gives inclusion criteria	99	0
	3b Gives exclusion criteria	90	0
	3c Eligibility criteria: defines method of recruitment (eg, by referral or self-selection, advertisements in public or in other clinics)	21	0
	3d Defines setting and location in which the study was carried out (country, city, and immediate environment such as inpatient or outpatient unit)	57	0
4. Interventions	4a Defines each intervention thoroughly, including control interventions	98	0
	4b Describes the "usual care" given to both the control group and the intervention group (open/closed ward, other psychotherapy or social therapies)	8	0
	4c Gives timing of interventions	90	0
5. Objectives	5a States prospectively defined clinical objectives or hypothesis	97	0
6. Outcomes	6a Defines primary outcome measure(s)	90	0
	6b Defines secondary outcome measure(s)	81	0
	6c When primary outcomes are assessed at several time points after randomization, the prespecified time point of primary interest is indicated	90	26
	6d When secondary outcomes are assessed at several time points after randomization, the prespecified time point of primary interest is indicated	71	31
	6e The provenance and properties of all scales are indicated by a reference	90	0
	6f Defines if any particular steps to increase the reliability of the measurements were given or not (eg, multiple observations, training of assessors)	27	0
7. Sample size	7a Gives sample size calculations	35	0
8. Sequence generation	8a Defines generation of random allocation sequence	16	0
	8b Defines if no restriction was used (simple randomization). Otherwise, the methods used to restrict the randomization, along with the method used for random selection, are specified	29	0
9. Allocation concealment	9a Defines method of allocation concealment (eg, numbered containers or central telephone)	15	0
10. Randomization implementation	10a Defines separation of generation and implementation of assignments	10	0
	10b States who generated the allocation sequence	14	0
	10c States who enrolled participants	0	0
	10d States who assigned participants to their groups	8	0
	10e States location of randomization code	4	0
11. Blinding (masking)	11a States whether participants were blinded or not	94	0
	11b States whether care providers were blinded or not	41	0
	11c States whether evaluators, monitors or analysts were blinded or not blinded	46	0
	11d States mechanism of blinding (eg, states if capsules, tablets or sham procedures were administered)	77	8
	11e States the similarity of placebo (eg, states similar appearance or taste of drugs, if placebo was "matching" or "identical")	44	16
	11f States how the success of blinding was evaluated (eg, evaluate participants' and care providers' opinion about what intervention was administered)	2	11
12. Statistical methods	12a Defines statistical analysis	97	0
	12b Defines, if applicable, which additional analyses, such as subgroup analyses and adjusted analyses, were used and why	55	69
13. Participant flow	13a Gives flowchart	33	0
	13b Reports the number of persons assessed for eligibility	49	0
	13c Reports the number of persons excluded before randomization	50	0
	13d Reports the number of persons randomized	96	0
	13e Reports the numbers of persons allocated to each of the intervention groups	95	0
	13f Reports the numbers of persons that did receive and did not receive the allocated intervention or are lost to follow up	88	0
	13g Reports the reasons for not receiving the allocated intervention or lost to follow up	84	0
	13h Reports the numbers of persons analyzed and excluded from analysis in each of the intervention groups	93	0
14. Dates of recruitment and follow-up	14a Gives recruitment period	39	0
	14b Gives study period	43	0
	14c Gives follow-up period (if applicable)	44	91

(continued)

Table 1 (continued). Checklist Items Based on the CONSORT Statement and Their Reporting in Randomized Controlled Trials on Bipolar Disorder

CONSORT Checklist	Item Number and Description	Bipolar Disorder Trials Reporting, %	
		Yes (if applicable)	Not Applicable
15. Baseline demographics	15a Gives prognostic variables (baseline characteristics) for intervention and control group: age	93	0
	15b Gives prognostic variables: sex	93	0
	15c Gives prognostic variables: education	7	0
	15d Gives prognostic variables: bipolar disorder I, bipolar disorder II	67	7
	15e Gives prognostic variables: duration of illness	42	0
	15f Gives prognostic variables: numbers of previous episodes	32	0
	15g Gives prognostic variables: severity of baseline symptomatology	91	0
16. Number of participants	16a States whether the analysis was by "intention-to-treat"	79	0
	16b Gives results as absolute numbers for primary outcomes	85	0
	16c Gives results as absolute numbers for secondary outcomes	74	8
17. Summary of results	17a Gives results for primary outcomes	98	0
	17b Gives results for secondary outcomes	94	7
	17c States if protocol deviations took place or not	8	0
	17d Confidence intervals/standard deviations are presented for the point estimates for each group	72	0
	17e Confidence intervals are presented for the point estimates/contrast between groups (effect/effect size)	18	0
	17f In trials in which interim analyses were performed, interpretation focuses on the final results at the close of the trial, not the interim results	93	86
	17g For both binary and survival time data, the results are also expressed as the number needed to treat for benefit or number needed to harm	8	0
18. Report other analysis	18b States whether or not adjusted analyses, including the choice of variables to adjust for, were planned	4	0
	18c If applicable, both unadjusted and adjusted analyses should be reported	14	93
19. Adverse events	19a Provide estimates of the frequency of the main adverse events separately for each intervention group	82	0
	19b Authors provide operational definitions for their measures of the severity of adverse events	6	4
20. Interpretation of results	20a States possible sources of bias	62	0
	20b States possible sources of imprecision	44	0
21. Generalizability of results	21a Discusses and interprets generalizability (external validity) of the trial findings	55	0
22. General interpretation	22a Interprets results with regard to former relevant studies	87	0
	22b Interprets results with regard to totality of data, that is in most cases a systematic review	19	0

Based on the 2001 Consolidated Standards of Reporting Trials (CONSORT) Checklist.[12]

covers, for example, randomization and allocation concealment (8a, 9a, 10), blinding (11a is absolutely essential; 11b and 11c are essential; 11d and 11e are needed to assess correctness; 11f would be good), and sample for statistical analyses, eg, intention to treat (16a). Clinically relevant items cover information on, for example, the participants (inclusion and exclusion criteria [3a, 3b], including validation of diagnosis and eligibility criteria [3c]), the intervention (4), baseline demographics (15a, 15b, 15e, 15f, 15g), and outcomes (6a, 16b, 17a, and 17d are all sufficient if given for primary efficacy outcome; 19a is absolutely essential; 19b is essential).

Data Extraction

As a first step, D.S., B.S., B.W., and A.P. all rated the same 6 articles. All raters were trained in the critical appraisal of randomized controlled trials. At this stage, any differences of opinion between investigators were due to an oversight or misunderstanding and were easily rectified. After discussion, there was total interobserver concordance on rating results, with no differences in opinion found. As a second step, D.S., B.S., B.W., and A.P. each rated a subset of the relevant articles.

Data Analysis

We did not award an aggregate score of the checklist items to the individual trial, as the items were not felt to be of equal weighting. Descriptive and analytic statistics (χ^2 test) were used to assess the quality of reporting with regard to time periods of publication and whether journals endorse or do not endorse the URM.

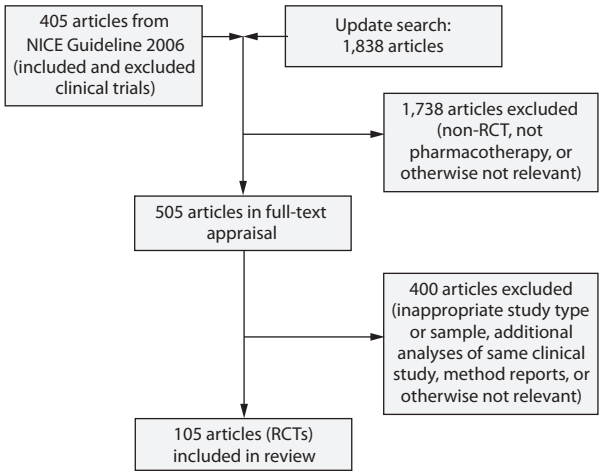
RESULTS

Number and Type of Trials

A total of 105 randomized controlled trials published in 27 journals were included in the analysis. Thirty of those were published in the period from January 2000 through June 2004 and 75 between July 2004 and December 2008. See Figure 1 for details on literature search and assessment.

Four journals, representing 42% of the included trials, followed the URM (*The New England Journal of Medicine*, *Archives of General Psychiatry*, *The American Journal of Psychiatry*, and *The Journal of Clinical Psychiatry*). More than two-thirds of the trials (68%) were published in the following 6 journals: *The Journal of Clinical Psychiatry* (no. = 20), *The American Journal of Psychiatry* (no. = 16), *Journal of Affective*

Figure 1. Flowchart Regarding Literature Search and Assessment



Abbreviations: NICE = National Institute for Health and Clinical Excellence, RCT = randomized controlled trial.

Disorders (no. = 10), *Bipolar Disorders* (no. = 9), *Journal of Clinical Psychopharmacology* (no. = 9), and *Archives of General Psychiatry* (no. = 7).

Assessment of Reporting Quality

As explained in the Method section, we did not award an aggregate score of the checklist items to the individual trial.

With respect to all measured checklist items (no. = 72), 42% (no. = 30) were generally reported adequately (reported adequately in more than 75% of all trials). Items generally reported adequately included essential methodological items such as whether participants were blinded or not (11a) and the mechanism of blinding (11d). Also, several items of information with essential clinical relevance were generally reported adequately, such as the inclusion and exclusion criteria (3a, 3b), the thorough definition of each intervention (4a), and the timing of each intervention (4c).

Twenty-five percent of all items (no. = 18) were generally reported inadequately (reported inadequately in less than 25% of all trials). Among the items generally reported inadequately, we also found essential methodological items such as generation of random allocation sequence (8a), method of allocation concealment (9a), and all items relevant for randomization implementation (10a–10e). Also, information with essential clinical relevance was generally reported inadequately, such as the effect size (17e) and the number needed to treat (17g).

In the publications from July 2004 through December 2008, in comparison with the earlier publication period, the quality of reporting was better for several essential methodological items (covering randomization and allocation concealment and blinding) and clinical relevance items (covering the description of participants, the intervention, and baseline demographics). For details on individual items, see Table 2. As can be seen, even with improvement, several aspects are still clearly suboptimally reported, eg, randomization and blinding procedures (8, 11) and baseline demographics (15).

The calculation of separate mean scores for all CONSORT items showed no consistent trend to differential reporting of essential methodological items in journals that endorse the URM (see Table 1 for a detailed report of results for all checklist items). The reporting of information on clinical relevance (items 17a, 17b, 17e, 17g, 19a) and generalizability (items 3a, 3b, 15a, 15b, 21a) of results, however, showed a consistent trend toward better reporting in journals endorsing the URM, with significant differences for the reporting of secondary outcomes (100% vs 89.9%; $P = .03$) and adverse events (93.2% vs 73.8%; $P = .011$) and interpretation of results with regard to totality of data (30.2% vs 11.5%; $P = .029$).

DISCUSSION

We analyzed the reporting quality of all randomized controlled trials of pharmacologic treatment of bipolar disorders published between 2000 and 2008. Of the applicable items on our checklist, 42% were generally reported adequately and 25% inadequately. Among both categories—adequately and inadequately reported items—we found essential methodological items and items with essential clinical relevance.

Reporting was especially poor for randomization procedures. Good randomization decreases the potential for bias to influence final results. Successful randomization hinges on 2 steps: generation of an unpredictable allocation sequence and concealment of this sequence from the investigator enrolling the participant (allocation concealment). The latter helps to prevent selection bias, protects the randomization sequence before and until the interventions are given to study participants, and can always be implemented. Trials that have not reported adequate allocation concealment have been found to be associated with exaggerated treatment effects.[3,4] Only 16% of bipolar disorder trials defined the generation of the random allocation sequence, 15% defined the method of allocation concealment, and 10% defined how the generation and implementation of assignments were separated. The reporting on blinding measures for participants, care providers, and evaluators (monitors, analysts) was more adequate (94%, 41%, and 46%, respectively) but also leaves considerable room for improvement. Despite the well-known phenomenon that psychopharmacologic side effects risk unmasking the blinding of participants, investigators, and assessors, only 2% of trials stated how the success of blinding was evaluated.

It is important to acknowledge that criteria such as blinding, randomization, and allocation concealment clearly are relevant for assessing the quality of study methodology. However, our study investigated this issue only through assessing the reporting. Accordingly, Huwiler-Müntener et al[29] state that a clear distinction should be made between the reporting quality and the methodological quality of the trials because well-conducted trials may be reported badly.

Table 2. Improvement of Reporting on Essential Methodological and Clinical Relevance Items Over Time and With Respect to Endorsement or No Endorsement of the Uniform Requirements for Manuscripts Established by the International Committee of Medical Journal Editors

		Time Period		Statistical	Uniform Requirements for Manuscripts		Statistical
Content	Item	01/2000–06/2004, % Positive	07/2004–12/2008, % Positive	Difference (χ^2 test)	Endorsement, % Positive	No Endorsement, % Positive	Difference (χ^2 test)
Methodological relevance							
Sequence generation	8a	16.7	16.0		11.4	19.7	
	8b	16.7	33.3	*	29.6	27.9	
Allocation concealment	9a	16.7	14.7		15.9	14.8	
Randomization implementation	10a	6.7	10.7		9.1	9.8	
	10b	13.3	14.7		6.8	19.7	*
	10c	NA	NA		0.0	0.0	
	10d	13.3	5.3		9.1	6.6	
	10e	0.0	5.3		2.3	4.9	
Blinding	11a	93.3	94.7		100.0	90.2	**
	11b	43.3	40.0		45.5	37.7	
	11c	40.0	48.0		47.7	44.3	
	11d	64.3	82.4	*	82.5	73.2	
	11e	32.1	50.0	*	40.0	47.9	
	11f	3.5	1.6		0.0	3.9	
Participants in analysis	16a	83.3	77.3		86.4	73.8	
Clinical relevance							
Participants	3a	100.0	98.7		100.0	98.4	
	3b	83.3	93.3		90.9	90.2	
	3c	6.7	26.7	**	13.6	26.2	
Interventions	4a	96.7	98.7		97.7	98.4	
	4b	6.7	8.0		15.9	1.6	**
	4c	76.7	96.0	**	88.6	91.8	
Baseline demographics	15a	93.3	93.3		100.0	88.5	**
	15b	93.3	93.3		100.0	88.5	**
	15c	3.3	8.0		11.4	3.3	
	15d	53.6	72.9	*	80.0	58.6	**
	15e	30.0	46.7		38.6	44.3	
	15f	26.7	34.7		31.8	32.8	
	15g	86.7	93.3		88.6	93.4	
Results	17a	96.7	98.7		100.0	96.7	
	17b	93.1	94.1		100.0	89.3	**
	17c	10.0	6.8		11.4	5.0	
	17d	73.3	72.0		75.0	70.5	
	17e	6.7	22.7	*	18.2	18.0	
	17f	100.0	88.9		100.0	90.9	
	17g	6.7	8.0		9.1	6.6	
Adverse events	19a	86.7	80.0		93.2	73.8	**
	19b	11.1	4.1		9.3	3.5	

Based on the 2001 Consolidated Standards of Reporting Trials (CONSORT) Checklist.[12]
*$P < .10$, **$P < .05$.
Abbreviation: NA = not applicable.

However, according to our results, many essential methodological aspects concerning the study design and analysis of the results were not reported adequately, a fact that limits considerably the quality assessment of these trials.

Authors of evidence-based guidelines or health technology assessments are forced to exclude or downgrade trials with inadequate reporting on study design.[11] In the German evidence-based and consensus-based S3 Guideline for Diagnosis and Therapy of Bipolar Disorders,[27] for example, all studies were included as long as randomization and the percentage of nonbipolar patients in a mixed sample were reported (and were ≤ 10%); randomized controlled trials with inadequate reporting of other criteria of study design were downgraded but still included. In all, only 6 studies were not downgraded from the highest level (1++) that could be achieved; 21 studies were downgraded to 1+, and all others were downgraded to 1– or were even ranked only as uncontrolled studies because the comparability of baseline characteristics of patients in the study arms could not be assessed. Exclusion or downgrading of trials reduces the value of the original research.[11] A high-quality reporting of study design and results is a necessary precondition to these trials' receiving any acknowledgment, without which there is no justification for research with human subjects.[30–32]

Of all analyzed trials, only 32% included a description of prestudy sample size calculations, even though it is considered essential for both scientific and ethical reasons that calculations be based on a clearly defined outcome.[33] Studies with small sample sizes often lack sufficient power to detect small but clinically significant differences between interventions and therefore cannot be considered a valid refutation of the usefulness of the new treatments.

Not only those who are interested in study methodology but also those who are interested in clinical relevance and generalizability of trial findings might have major difficulties in extracting important information in the field of bipolar disorder. For example, clinicians trained in evidence-based medicine and clinical epidemiology[34,35] are more interested in an effect measure, such as number needed to treat or number needed to harm, and in the effect size of new treatments (that is, the contrast between groups) than in individual point estimates for each group. However, only 6% of trials also expressed results as number needed to treat or number needed to harm, and only 15% reported confidence intervals for contrast between groups.

With respect to the specific field of chronic psychiatric disorders such as bipolar disorder, reporting of baseline demographics' serving as prognostic variables was very good for age, sex, and severity of baseline symptomatology (all >90%) but poor for other relevant variables such as education (7%), duration of illness (32%), and number of previous episodes (42%).

Limitations

We acknowledge that journal restrictions regarding, for example, word count would have been a limiting factor for the inclusion of all information that we have used to assess reporting quality. However, critical information with a bearing on scientific validity and clinical relevance (see Table 2) should always be prioritized.[36]

After submission of our study, the CONSORT 2010 statement was published.[13] For a review about the reporting of randomized controlled trials between 2000 and 2008, it remains reasonable to refer to the CONSORT 2001 statement.[12] Future reviews about the quality of reporting from 2010 onward, however, should refer to the CONSORT 2010 statement. Changes in the CONSORT 2010 statement are related mainly to wording and specification of appraisal criteria. New items in CONSORT 2010 require reporting about trial registration, availability of trial protocol, and funding sources. Empirical evidence, especially in the field of psychiatry, supports the need for better reporting on these issues.[8,9,36–38]

CONCLUSION

Our findings suggest that, while some trial-related information is well reported, a good part of the reporting quality of randomized controlled trials in bipolar disorder falls well below the required and also practically feasible level for many aspects essential for adequate interpretation of methodological quality and clinical relevance. As results of randomized controlled trials have a significant impact on clinical decision-making, guideline development, and health technology assessment, vigilance is required when relying on the results of these trials for recommending and prescribing new treatments. Authors should be further encouraged to follow the CONSORT criteria when reporting the results of a trial. While the responsibility for improvement of reporting should primarily lie with the investigators, reviewers and editors of psychiatric or general journals could facilitate the process by endorsing and effectively considering guidelines such as the CONSORT statement.

Drug names: fluoxetine (Prozac and others).

Author affiliations: Institute for History, Ethics, and Philosophy of Medicine, CELLS–Centre for Ethics and Law in the Life Sciences, Hannover Medical School, Hannover, Germany (Dr Strech); and Department of Public Health (Dr Soltmann) and Department of Psychiatry and Psychotherapy (Drs Weikert, Bauer, and Pfennig), University Hospital Carl Gustav Carus, Technische Universität, Dresden, Germany. Dr Weikert is now affiliated with the German Agency for Quality in Medicine, Berlin, Germany.
Author contribution: Dr Soltmann provided equal contribution as first author of this work.

Potential conflicts of interest: **Dr Strech** has received speaker honoraria from Roche and the German Association of Research-Based Pharmaceutical Companies. **Dr Bauer** has received grant/research support from The Stanley Medical Research Institute, NARSAD, and Eli Lilly; has been a consultant for AstraZeneca, Eli Lilly, Servier, Janssen-Cilag, Lundbeck, Bristol-Myers Squibb, and Otsuka; and has received speaker honoraria from AstraZeneca, Eli Lilly, GlaxoSmithKline, Lundbeck, Bristol-Myers Squibb, Otsuka, and Pfizer. **Dr Pfennig** has received a stipend from GlaxoSmithKline; has received research support from GlaxoSmithKline and AstraZeneca; and has received speaker honoraria from AstraZeneca and Eli Lilly. **Drs Soltmann** and **Weikert** report no competing interests.
Funding/support: Financial support for the development of the S3-Guideline for Diagnosis and Treatment of Bipolar Disorders was provided by the German Society of Bipolar Disorder (Deutsche Gesellschaft für Bipolare Störungen [DGBS]) and the German Society of Psychiatry, Psychotherapy, and Neurology (Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde [DGPPN]).

REFERENCES

1. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.1. 2008. www.cochrane-handbook.org/. Verified November 11, 2010.
2. Altman DG, Schulz KF, Moher D, et al; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134(8):663–694.
3. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ*. 2001;323(7303):42–46.
4. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273(5):408–412.
5. Dans AL, Dans LF, Guyatt GH, et al; Evidence-Based Medicine Working Group. Users' guides to the medical literature, XIV: how to decide on the applicability of clinical trial results to your patient. *JAMA*. 1998;279(7):545–549.
6. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318(26):1728–1733.
7. McHenry L. Ethical issues in psychopharmacology. *J Med Ethics*. 2006;32(7):405–410.
8. Whittington CJ, Kendall T, Fonagy P, et al. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet*. 2004;363(9418):1341–1345.
9. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358(3):252–260.
10. Melander H, Ahlqvist-Rastad J, Meijer G, et al. Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ*. 2003;326(7400):1171–1173.
11. Atkins D, Best D, Briss PA, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490.
12. Moher D, Schulz KF, Altman DG; CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med*. 2001;134(8):657–662.
13. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010

statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med*. 2010;7(3):e1000251.
14. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publications, 2008. www.icmje.org. Verified November 11, 2010.
15. Falagas ME, Grigori T, Ioannidou E. A systematic review of trends in the methodological quality of randomized controlled trials in various research fields. *J Clin Epidemiol*. 2009;62(3):227–231.
16. CONSORT Group. CONSORT: Transparent Reporting of Trials, 2009; www.consort-statement.org. Verified November 11, 2010.
17. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ*. 1998;317(7167):1181–1184.
18. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996;17(1):1–12.
19. Cipriani A, Malvini L, Furukawa TA, et al. Relationship between quality of reports of antidepressant randomized controlled trials and treatment estimates: systematic review, meta-analysis, and meta-regression analysis. *J Clin Psychopharmacol*. 2007;27(4):352–356.
20. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet*. 2005;365(9467):1348–1353.
21. Han C, Kwak KP, Marks DM, et al. The impact of the CONSORT statement on reporting of randomized clinical trials in psychiatry. *Contemp Clin Trials*. 2009;30(2):116–122.
22. Merikangas KR, Akiskal HS, Angst J, et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch Gen Psychiatry*. 2007;64(5):543–552.
23. Pini S, de Queiroz V, Pagnin D, et al. Prevalence and burden of bipolar disorders in European countries. *Eur Neuropsychopharmacol*. 2005;15(4):425–434.
24. Waraich P, Goldner EM, Somers JM, et al. Prevalence and incidence studies of mood disorders: a systematic review of the literature. *Can J Psychiatry*. 2004;49(2):124–138.
25. Kupfer DJ, Frank E, Grochocinski VJ, et al. Demographic and clinical characteristics of individuals in a bipolar disorder case registry. *J Clin Psychiatry*. 2002;63(2):120–125.
26. Fajutrao L, Locklear J, Priaulx J, et al. A systematic review of the evidence of the burden of bipolar disorder in Europe. *Clin Pract Epidemol Ment Health*. 2009;5(1):3.
27. Pfennig A, Weikert B, Falkai P, et al. [Development of the evidence-based S3 Guideline for Diagnosis and Therapy of Bipolar Disorders] [article in German]. *Nervenarzt*. 2008;79(4):500–504.
28. National Institute for Health and Clinical Excellence. *Bipolar Disorder: The Management of Bipolar Disorder in Adults, Children and Adolescents, in Primary and Secondary Care. National Clinical Practice Guideline Number 38. National Collaborating Centre for Mental Health, Commissioned by the National Institute for Health and Clinical Excellence (NICE)*. London, England: The British Psychological Society and Gaskell; 2006.
29. Huwiler-Müntener K, Jüni P, Junker C, et al. Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA*. 2002; 287(21):2801–2804.
30. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? *JAMA*. 2000;283(20):2701–2711.
31. World Medical Association. *Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*. Seoul, South Korea: World Medical Association; 2008.
32. Council for International Organizations of Medical Sciences. *International Ethical Guidelines for Biomedical Research Involving Human Subjects*. Geneva, Switzerland: Council for International Organizations of Medical Sciences; 2002.
33. Emanuel EJ, Wendler D, Grady C. An ethical framework for biomedical research. In: Emanuel EJ, Grady C, Crouch RA, et al, eds. *The Oxford Textbook of Clinical Research Ethics*. Oxford, England: Oxford University Press; 2008:123–135.
34. Sackett DL, Straus S, Scott Richardson W. *Evidence-Based Medicine: How to Practice and Teach EBM*. Philadelphia, PA: Churchill Livingstone; 2000.
35. Fletcher RW, Fletcher SW. *Clinical Epidemiology: The Essentials*. 4th ed. Baltimore, MD: Lippincott Williams & Wilkins; 2005.
36. Chan AW, Hróbjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004;291(20):2457–2465.
37. Lexchin J, Bero LA, Djulbegovic B, et al. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ*. 2003; 326(7400):1167–1170.
38. Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE*. 2008;3(8):e3081.

eAppendix 1 is available at PSYCHIATRIST.COM.