Design, Quality, and Bias in Randomized Controlled Trials of Systemic Lupus Erythematosus

FOTINI B. KARASSA, ATHINA TATSIONI, and JOHN P.A. IOANNIDIS

ABSTRACT. ***Objective.*** To appraise systematically the study design and quality of reporting of randomized controlled trials (RCT) on systemic lupus erythematosus (SLE) and to identify potential defects and biases.

Methods. RCT with at least 5 patients with SLE were retrieved from MEDLINE, EMBASE, and the Cochrane Library. We analyzed study design, quality of reporting, and trial results.

Results. Ninety-four trial reports (37 on lupus nephritis) were eligible with 2,257 SLE patients (n = 795 in lupus nephritis trials). Median sample size was 28 patients. Fifty-one trials (54.3%) were double blind, but only 31 (33.0%) mentioned the randomization mode, only 19 (20.2%) described allocation concealment, and only 7 (7.5%) were adequately powered. Sixty-three trials (67%) described adequately reasons for withdrawals. Nephritis trials had on average longer followup ($p = 0.001$) and were less likely to be double blind ($p < 0.001$), to describe reasons for withdrawals [both overall ($p = 0.008$) and per arm ($p = 0.009$)] and to involve a comparison against placebo or no treatment ($p < 0.001$). Larger trials scored higher on several quality characteristics. Significant efficacy or trend for efficacy was claimed in 72 reports (76.6%) and this was even more common in trials published in 1999-2002 (89.5%). Significant efficacy was found more frequently in trials that clearly specified withdrawals per arm ($p = 0.001$) and outcomes ($p = 0.001$) and used intention-to-treat analyses ($p = 0.03$). Besides outcome specification, no other quality variables seemed to improve significantly over time.

Conclusion. Several aspects of the design and reporting of RCT on SLE can be improved. Larger, adequately powered, and accurately reported trials are needed. (J Rheumatol 2003;30:979–84)

Key Indexing Terms:
SYSTEMIC LUPUS ERYTHEMATOSUS
RANDOMIZED CONTROLLED TRIALS
BIAS
STUDY DESIGN

For patients with systemic lupus erythematosus (SLE), treatment options for several disease manifestations are limited and new treatments are important to develop. New treatments need to be evaluated with randomized controlled trials (RCT), since other designs may entail biases[1,2]. However, even RCT may not be devoid of potential defects in their design, conduct, analyses, and reporting[3,4]. Investigations in RCT of autoimmune diseases other than SLE suggest that important biases may be prevalent[5,6].

RCT in SLE are challenging. The heterogeneity of disease manifestations, the unpredictable course of the disease, the longterm nature of the major outcomes, and logistic considerations have all posed obstacles to the adoption of large multicenter trials. We systematically appraised RCT that have generated the rational basis of SLE therapeutics over the last 30 years. We evaluated variables that reflect characteristics of the study populations, study design, quality of reporting, and reported significance of the main findings. We assessed whether elements of the study design and its quality relate to finding significant efficacy for the tested interventions, and whether aspects of the study design and reporting are improving over time. Finally, we tried to identify potential deficiencies that may be important to prevent in future trials.

From the Clinical Trials and Evidence-Based Medicine Unit, Department of Hygiene and Epidemiology, the University of Ioannina School of Medicine, Ioannina, Greece and the Division of Clinical Care Research, Tufts-New England Medical Center, Boston, Massachusetts, USA.

F.B. Karassa, MD; A. Tatsioni, MD, Clinical Trials and Evidence-Based Medicine Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine; J.P.A. Ioannidis, MD, Clinical Trials and Evidence-Based Medicine Unit, Chair, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine and the Division of Clinical Care Research, Tufts-New England Medical Center.

Address reprint requests to Dr. J.P.A. Ioannidis, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, 45110, Greece. E-mail: jioannid@cc.uoi.gr

Submitted July 15, 2002; revision accepted October 31, 2002.

MATERIALS AND METHODS

Eligibility criteria and identification of studies. RCT were eligible if they had randomized at least 5 patients with SLE, regardless of whether patients with other diseases were considered as well. Studies with alternate subject allocation (pseudo-randomization) and sub-studies were excluded. Duplicate publications of the same trial (same patients and same followup) were screened and only one report was retained. We considered all separate reports that pertained to distinct temporal phases of the same project, describing longer followup and/or additional arms compared with the first published report. We retained the report with the longest followup and largest number of arms for estimating the cumulative number of patients and outcomes. Searches used MEDLINE (1966-April 2002) and EMBASE (1989-2002) for publications in English, French, German, or Italian. We

used the terms systemic lupus erythematosus, glomerulonephritis, Raynaud's phenomenon, pulmonary hypertension, neuropsychiatric lupus, or corticosteroid AND osteoporosis. RCT were identified using the terms proposed by the Cochrane Collaboration (www.cochrane.org). Second, we screened the Cochrane Library Controlled Trials Register (Update Software, Issue 1, 2002). Finally, references of retrieved articles were screened. We excluded meeting proceedings, since unpublished trials cannot be adequately evaluated for their design and quality through the limited abstract information.

Data extraction and replication. Data extraction followed the layout used in a previous evaluation of RCT in scleroderma[5]. We recorded the following data from each report: journal, authors, total sample size, number of patients with SLE, demographics, prior disease duration, mean (median) followup, deaths during followup, location of enrolment, year of publication, number of arms, parallel or cross-over design, treatment(s), whether the trial was a comparison of a treatment against placebo/no treatment or a comparative trial of different regimens, and funding sources. We recorded targeted disease manifestations and whether nephritis was an eligibility criterion. Quality variables included the description of sample size per arm, randomization mode, allocation concealment, masking, reasons of withdrawals (overall and per arm), percent of withdrawals, performance of power calculations, well-specified study outcomes (identified, defined and/or graded as primary and secondary upfront), and whether intention-to-treat analyses were performed. Finally, the results of each trial were categorized as showing significant efficacy, trend for efficacy, no effect, trend for harm, or significant harm for a tested experimental intervention versus the intervention of the control arm. This categorization was based on overall appraisal of the key outcome results, with significance claimed for $p < 0.05$. For non-significant differences, trends versus no effects were differentiated based on the interpretation of the abstract and the discussion of the report[5].

Two investigators independently extracted data. Differences were solved by consensus. Concordance of categorization of results between the 2 extractors was estimated by the weighted kappa coefficient[7].

Analyses. Trials focusing on nephritis were compared against all other trials. We evaluated whether quality attributes were related to the study sample size and duration of followup and whether sample size, followup and quality characteristics correlated with finding significant efficacy. We also evaluated whether design and quality characteristics changed over time in quartiles defined per publication year.

Analyses used Mann-Whitney U test, t test, and analysis of variance (ANOVA) for continuous variables; chi-square, Fisher's exact test, and trend-adjusted chi-square for categorical data; and Spearman coefficients in SPSS 10 (SPSS, Inc., Chicago, IL, USA) and StatXact 3 (Cytel, Boston, MA, USA). P values are 2 tailed.

RESULTS

Database. We reviewed 174 reports. Eighty-one were excluded [no clear inclusion of SLE patients or < 5 SLE patients (n = 40), non-randomization (n = 16), duplication of data or sub-study status (n = 12), non-eligible language (n = 6), pseudo-randomization (n = 4), and no original data (n = 3)]. One eligible article reported 2 RCT. Thus 94 RCT published between 1971 and 2002 were included (English: n = 92, German: n = 2).

Eleven reports were earlier publications of projects subsequently reanalyzed with longer followup and/or more randomized arms. Seventy-one included only SLE patients; of the remaining 23, 11 had only 5–20% SLE representation. Exclusion of the latter group yielded largely similar results (not shown). Only 12 reports pertained to multicenter trials, 6 of these including exclusively SLE patients.

Thirty-seven of the 94 reports focused on lupus nephritis. Forty-eight of the remaining 57 trials targeted patients on the basis of specific SLE manifestations or disease status: 13 targeted specific SLE manifestations (articular and/or cutaneous manifestations: n = 6; Raynaud's phenomenon: n = 5; and other: n = 2); 12 enrolled patients on longterm corticosteroids with or without osteoporosis; 12 targeted subjects with stable or mild disease; 7 pertained to patients with significant disease activity; 2 addressed lupus pregnancy; and 2 required the presence of specific autoantibodies.

Seventy different interventions were evaluated. Cytotoxic agents and immunomodulators were tested in 36 reports, 17 of which dealt with cyclophosphamide regimens against steroids, other immunosuppressives or combinations thereof. Bone formation-stimulating agents and antiresorptive drugs were tested in 12 trials (bisphosphonates in half of them). Other frequently tested interventions were hormonal therapies (n = 5), antimalarials (n = 5), different glucocorticoid doses/preparations (n = 5), calcium channel blockers (n = 3), fish oil (n = 3), nonsteroidal antiinflammatory drugs (n = 3), and immunizations (influenza or pneumococcal, n = 3).

Trial characteristics (Table 1). Although 9 reports each had over 100 patients, they were mostly studies that also included non-SLE patients. Only one project (2 trial reports) had over 100 SLE patients. The total number of SLE patients was 2,257 (795 in nephritis trials). A total of 131 deaths were recorded. Women predominated in all trials. Half of the trials had a followup of at least 1 year. Cross-over designs were relatively uncommon and half of the reports involved comparisons against placebo or no treatment.

Several quality variables were neglected in most trial reports, especially specification of the mode of randomization, allocation concealment, and power calculations. About half the trials were double blind and most described the reasons for withdrawals. Withdrawals were considerable, exceeding 20% in 21 of 76 trial reports where information was available. Outcomes were well specified in only slightly over half the trial reports and a multiplicity of outcomes was the rule (median 5 per trial). The main categories of specific outcomes used in the 71 trials that included exclusively SLE patients are shown in Table 2. Intention-to-treat was used in the analysis of about half the trials.

There was high agreement between the 2 independent investigators in terms of the assessment of the trial results (weighted kappa 0.90, 95% CI: 0.83 to 0.96). Slightly over three-quarters of the trial reports described either significant efficacy or efficacy trends for the experimental intervention. Only 2 trials mentioned that harm exceeded benefits from the experimental intervention.

Compared with other trials, nephritis trials included younger patients ($p < 0.001$) with shorter disease duration ($p = 0.04$), had longer followup ($p = 0.001$) and were less

Table 1. Characteristics of eligible trials.

Characteristics	All Trials n = 94	Nephritis Trials n = 37	Other Trials n = 57
Design			
Sample size, median (IQR)	28 (19, 50)	20 (13, 42)	30 (21, 55)
Sample size [SLE only], median (IQR)	21 (12, 40)	20 (13, 42)	21 (11, 40)
Proportion of women, median (IQR)	90 (81, 97.3)	86 (81, 89.8)	93.3 (81.5, 100)
Proportion of Caucasian or Hispanic patients, median (IQR)	71 (45, 89)	67 (23.3, 80)	73 (45, 94.5)
Mean age, median (IQR) yrs	35.9 (31, 43)	30.3 (27.1, 34)*	38.7 (33.5, 50)*
Mean disease duration, median (IQR) mos	57.6 (36, 92.3)	42.7 (33, 56.4)**	77.2 (54, 102.5)**
Mean followup, median (IQR) mos	12 (5.8, 24)	18 (6, 42.5)*	9.3 (5, 15.7)*
Enrolment in the USA, n (%)	46 (48.9)	22 (59.5)	24 (42.1)
Year of publication, median (IQR)	1993 (1984, 1998)	1991 (1982, 1998)	1995 (1985, 1998)
Cross-over design, n (%)	16 (17)	6 (16.2)	10 (17.5)
More than 2 arms, n (%)	18 (19.1)	12 (32.4)	6 (10.5)
Comparison against placebo/no treatment, n (%)	47 (50)	9 (24.3)*	38 (66.7)*
Reported funding			
Industry, n (%)	16 (17)	3 (8.1)	13 (22.8)
Government, n (%)	29 (30.9)	9 (24.3)	20 (35.1)
Private, n (%)	29 (30.9)	13 (35.1)	16 (28.1)
Quality			
Sample size per arm specified, n (%)	78 (83)	30 (81.1)	48 (84.2)
Randomization mode specified, n (%)	31 (33)	11 (29.7)	20 (35.1)
Allocation concealment specified, n (%)	19 (20.2)	7 (18.9)	12 (21.1)
Masking			
Double blind, n (%)	51 (54.3)	10 (27)*	41 (71.9)*
Single blind, n (%)	7 (7.5)	3 (8.1)	4 (7)
Open label, n (%)	36 (38.3)	24 (64.9)*	12 (21.1)*
Withdrawals described, n (%)	76 (80.9)	25 (67.6)*	51 (89.5)*
Withdrawals described per arm, n (%)	63 (67)	19 (51.4)*	44 (77.2)*
Percent of withdrawals, median (IQR)	9.9 (4, 21.4)	7 (4.4, 18)	10 (4, 25.4)
Power calculations mentioned, (%)	22 (23.4)	8 (21.6)	14 (24.6)
Power 80% at $\alpha = 0.05$, n (%)	7 (7.5)	0 (0)**	7 (12.3)**
Outcomes well specified, n (%)	55 (58.5)	20 (54.1)	35 (61.4)
Multiple outcomes, n (%)	88 (93.6)	35 (94.6)	53 (93)
Multiple comparison adjustment used, n (%)	9 (10.2)	4 (11.4)	5 (9.4)
Number of outcomes, median (IQR)	5 (4, 6)	5 (3, 6)	5 (4, 6)
Intention-to-treat analysis used, n (%)	50 (53.2)	23 (62.2)	27 (47.4)
Results			
Results for the experimental intervention			
Significant efficacy, n (%)	50 (53.2)	17 (45.9)	33 (57.9)
Trend for efficacy, n (%)	22 (23.4)	12 (32.4)	10 (17.5)
No effect, n (%)	20 (21.3)	8 (21.6)	12 (21.1)
Trend of harm, n (%)	1 (1.1)	0	1 (1.8)
Significant harm, n (%)	1 (1.1)	0	1 (1.8)

IQR: interquartile range, SLE: systemic lupus erythematosus. * $p < 0.01$ for nephritis trials vs other trials; ** $0.01 < p < 0.05$ for nephritis trials vs other trials.

likely to be double blind ($p < 0.001$), to involve a comparison against placebo/no treatment ($p < 0.001$), and to describe reasons for withdrawals overall ($p = 0.008$) and per study arm ($p = 0.009$); none was adequately powered ($p = 0.02$) (Table 1). As expected, renal outcomes were infrequent in non-nephritis trials; conversely, patient and physician ratings were rare in nephritis trials (Table 2).

Correlations between quality characteristics and study size and duration (Table 3). Larger trials were more likely than smaller trials to specify clearly the sample size per arm, the randomization mode, withdrawals (both overall and per arm), and study outcomes. Power calculations were also somehow more common in larger trials.

Conversely, trials with longer duration of followup did not seem to have improved quality characteristics compared with trials with shorter followup with the exception of more frequent clarification of the sample size per arm. Trials with longer followup were less likely to be double blind and had more withdrawals.

Correlates of significant efficacy of the experimental intervention. Significant efficacy was found more frequently in trials that specified withdrawals per arm (odds ratio 4.56,

Table 2. Main categories of outcomes used in 71 SLE trials. Only outcomes used by at least 8 trial reports are shown.

Outcome n (%)	All SLE Trials n = 71	Nephritis Trials n = 37	Other SLE Trials n = 34
Mortality	11 (15.5)	9 (24.3)	2 (5.9)
Flare or relapse	15 (21.1)	7 (18.9)	8 (23.5)
Remission or response outcome	15 (21.1)	11 (29.7)	4 (11.8)
Safety, toxicity, adverse effects	31 (43.7)	19 (51.4)	12 (35.3)
Dose of medication or requirement for medication	22 (30.9)	9 (24.3)	13 (38.2)
Immunological measurements	30 (42.3)	14 (37.8)	16 (47.1)
SLAM score	8 (11.3)	4 (10.8)	4 (11.8)
Physician rating of disease activity or clinical state	23 (32.4)	4 (10.8)*	19 (55.9)*
Patient rating on disease activity or clinical state	8 (11.3)	0*	8 (23.5)*
End-stage renal disease	9 (12.7)	9 (24.3)*	0*
Renal function tests	33 (46.5)	26 (70.3)*	7 (20.6)*
Proteinuria and/or findings on urinary sediment	10 (14.1)	10 (27)*	0*
Composite outcomes	16 (22.5)	6 (16.2)	10 (29.4)

* $p < 0.01$ for the comparison of nephritis vs other SLE trials.

Table 3. Quality characteristics, study size and duration: correlation coefficients (p values).

Quality Variables	Sample Size	Trial Duration
Sample size per arm specified	0.42 (< 0.001)	0.34 (0.001)
Randomization mode specified	0.42 (< 0.001)	0.16 (0.11)
Allocation concealment specified	0.00 (0.99)	0.01 (0.94)
Masking: double blind	–0.06 (0.57)	–0.43 (< 0.001)
Withdrawals described	0.28 (0.006)	–0.08 (0.43)
Withdrawals described per arm	0.27 (0.009)	–0.16 (0.13)
Percent of withdrawals	0.08 (0.50)	0.20 (0.09)
Power calculations mentioned	0.21 (0.046)	0.01 (0.93)
Outcomes well specified	0.43 (< 0.001)	0.02 (0.84)
Intention-to-treat analysis used	0.11 (0.31)	0.06 (0.54)

p = 0.001) and outcomes (odds ratio 4.11, p = 0.001) and in those that used intention-to-treat analyses (odds ratio 2.57, p = 0.03). Other quality variables were not strongly associated with finding significant efficacy. Significant efficacy was also correlated with sample size (r = 0.27, p = 0.008), but not with duration (r = –0.04, p = 0.69).

Changes over time. Recent trial reports were more likely to specify outcomes adequately and to show significant efficacy or at least a trend of efficacy. Nine of 10 trials published after 1993 claimed that the experimental intervention was effective or potentially effective (Table 4). Besides outcome specification, no other quality variables seemed to improve significantly over time. The sample size of the trials did not increase significantly over time (r = 0.19, p = 0.07).

DISCUSSION

There is a relative dearth of randomized evidence for the management of SLE. Despite almost 100 trials performed over 30 years, most have been very small and account for only slightly over 2,000 SLE patients. Given the very diverse medical problems that SLE patients face, the evidence is hardly adequate for any major disease manifestation. Even for lupus nephritis, the most common serious manifestation of SLE, fewer than 800 patients have been randomized.

Besides a limited sample size, several quality variables have been relatively neglected in the reporting of most SLE trials to date. Randomization mode and allocation concealment are recorded infrequently and power calculations are uncommon. Similar problems have been described in the appraisal of randomized trials for scleroderma[5] and in unselected samples of rheumatology-related trials[6], thus suggesting a problem that pertains to the broader field of rheumatology. Similar deficiencies have been reported also across other medical fields[3,8,9]. Actually, for some other quality variables, such as the reporting of withdrawals, SLE trials tend to be better than other rheumatology trials[6] and trials in other medical domains[9]. The wider dissemination of the CONSORT recommendations[10,11] should help improve the standardized reporting of RCT in the medical literature in the near future.

One variable that is already improving in recent trials is the specification of study outcomes. SLE is characterized by

Table 4. Changes in trial design, quality characteristics and outcomes over time.

Trial Characteristics	1971–1984 n = 24 (%)	1985–1993 n = 25 (%)	1994–1998 n = 26 (%)	1999–2002 n = 19 (%)	p value
Focus on lupus nephritis	12 (50)	12 (48)	6 (23.1)	7 (36.8)	0.19
Sample size per arm specified	22 (91.7)	19 (76)	20 (76.9)	17 (89.5)	0.35
Mean disease duration < 5 yrs	9 (37.5)	3 (12)	6 (23.1)	2 (10.5)	0.10
Followup > 1 yr	11 (45.8)	9 (36)	9 (34.6)	8 (42.1)	0.84
Cross-over design	3 (12.5)	8 (32)	2 (7.7)	3 (15.8)	0.12
Randomization mode specified	6 (25)	11 (44)	11 (42.3)	3 (15.8)	0.13
Allocation concealment specified	3 (12.5)	8 (32)	4 (15.4)	4 (21.1)	0.34
Masking: double blind	13 (54.2)	15 (60)	12 (46.2)	11 (57.9)	0.78
Withdrawals described	17 (70.8)	18 (72)	25 (96.2)	16 (84.2)	0.07
Withdrawals described per arm	14 (58.3)	14 (56)	23 (88.5)	12 (63.2)	0.05
Power calculations specified	3 (12.5)	4 (16)	9 (34.6)	6 (31.6)	0.19
Outcome well specified	7 (29.2)	13 (52)	19 (73.1)	16 (84.2)	0.0007
Results					
Showing significant efficacy	5 (20.8)	16 (64)	16 (61.5)	13 (68.4)	0.003
Trend for efficacy	13 (54.2)	19 (78)	23 (88.5)	17 (89.5)	0.03
Funding from industry	0	5 (20)	6 (23.1)	5 (26.3)	0.07
Intention-to-treat analysis used	12 (50)	14 (56)	12 (46.2)	12 (63.2)	0.70
Comparison with placebo/no treatment	9 (37.5)	14 (56)	14 (53.8)	10 (52.6)	0.58

a remarkable diversity of clinical and serological features, and the development of uniform, standardized outcome measures and responder indices should be encouraged[12,13]. Besides outcome specification, we found no improvements over time in other quality variables, in contrast to other domains[5,6]. Quality variables correlated with sample size. Aiming for larger trials should parallel efforts aiming for better quality and more accurate reporting. Although current SLE trials are getting larger, most SLE trials continue to be small and underpowered.

Despite the lack of statistical power, SLE trials have exhibited very high rates of results that show significant efficacy for the experimental intervention. The rates of positive trials have reached up to 90% in the last decade, i.e. almost all published SLE trials confer the message that the tested treatment works. While this may well be true, such success rates among small studies raise the possibility of bias, especially publication bias, the lack of publication of trials with negative results[14]. Negative-outcome trials may either disappear completely or may be published with considerable delays (time lag bias)[15]. Another possibility is that the observed quality deficiencies are associated with spuriously inflated treatment effects. Other investigators[3,14,15] have shown that lack of allocation concealment and lack of double blinding tend to inflate the magnitude of the observed treatment effects. However, these associations are not observed consistently[16], and even if they exist in SLE trials, one cannot tell how the results should be corrected to account for quality defects[17,18].

Multiplicity of outcomes may further foster spurious positive findings, especially if outcomes are selected *post hoc*[19]. Furthermore, for some SLE trials analyses were conducted at several time points during followup. While longterm followup is desirable, repeated analyses affect the type I error, and typically much smaller p values than the conventional p = 0.05 are required to claim efficacy in such sequential analyses[20,21]. Finally, imbalance in concomitant treatments (e.g., steroids) in the compared arms may also introduce bias[22].

Much of what we have learned for the treatment of SLE comes from randomized trials, including the major advances for the management of lupus nephritis[23-25]. SLE is not very common on a population basis, thus very large trials[26] may not be feasible. Nevertheless, recent epidemiologic studies in SLE have recruited several hundred or even several thousand patients[27,28]. Although enrolling patients for RCT is much more difficult, there is considerable room for launching multicenter trials with at least a few hundred SLE patients. Non-randomized studies, even with large sample sizes, may still introduce biases in assessing therapeutic interventions[2,29]. A change towards larger RCT requires appropriate resources and organization. Combined evaluation of trials by metaanalysis[30-32] is an alternative, but it would be more successful with larger trials and standardized methodology. Furthermore, longterm followup will continue to be needed. Such trials may be able to use adequately powered hard endpoints (e.g., death or end-stage renal disease) to complement our insight from laboratory or soft clinical outcomes[33].

REFERENCES

1. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286:821-30.
2. Ioannidis JP, Karassa FB. Bias in uncontrolled therapeutic trials in rheumatology due to selection of populations with extreme characteristics. J Rheumatol 2001;28:1881-7.

3. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408-12.
4. Jadad AR, Rennie D. The randomized controlled trial gets a middle-aged checkup. JAMA 1998;279:319-20.
5. Kyriakidi M, Ioannidis JP. Design and quality considerations for randomized controlled trials in systemic sclerosis. Arthritis Rheum 2002;47:73-81.
6. Hill CL, LaValley MP, Felson DT. Secular changes in the quality of published randomized clinical trials in rheumatology. Arthritis Rheum 2002;46:779-84.
7. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1:307-10.
8. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994;272:122-4.
9. Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. JAMA 2001;285:437-43.
10. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637-9.
11. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663-94.
12. Strand V, Gladman D, Isenberg D, Petri M, Smolen J, Tugwell P. Endpoints: consensus recommendations from OMERACT IV. Outcome Measures in Rheumatology. Lupus 2000;9:322-7.
13. Isenberg D, Ramsey-Goldman R. Assessing patients with lupus: towards a drug responder index. Rheumatology Oxford. 1999;38:1045-9.
14. Easterbrook P, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991;337:867-72.
15. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998;279:281-6.
16. Balk EM, Bonis PAL, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized trials. JAMA 2002;287:2973-82.
17. Ioannidis JP, Lau J. Can quality of clinical trials and meta-analyses be quantified? Lancet 1998;352:590-1.
18. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609-13.
19. Ioannidis JPA, Lau J. The impact of high risk patients on the results of clinical trials. J Clin Epidemiol 1997;50:1089-98.
20. Fleming TR, DeMets DL. Monitoring of clinical trials: issues and recommendations. Control Clin Trials 1993;14:183-97.
21. Lan KK, Rosenberger WF, Lachin JM. Use of spending functions for occasional or continuous monitoring of data in clinical trials. Stat Med 1993;12:2219-31.
22. Corzillius M, Bae SC. Methodological issues of corticosteroid use in SLE clinical trials. Lupus 1999;8:692-7.
23. Austin HA 3rd, Klippel JH, Balow JE, et al. Therapy of lupus nephritis. Controlled trial of prednisone and cytotoxic drugs. N Engl J Med 1986;314:614-9.
24. Boumpas DT, Austin HA 3rd, Vaughn EM, et al. Controlled trial of pulse methylprednisolone versus two regimens of pulse cyclophosphamide in severe lupus nephritis. Lancet 1992; 340:741-5.
25. Lewis EJ, Hunsicker LG, Lan SP, Rohde RD, Lachin JM. A controlled trial of plasmapheresis therapy in severe lupus nephritis. The Lupus Nephritis Collaborative Study Group. N Engl J Med 1992;326:1373-9.
26. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med 1984;3:409-22.
27. Cervera R, Khamashta MA, Font J, et al. Morbidity and mortality in systemic lupus erythematosus during a 5-year period. A multicenter prospective study of 1,000 patients. European Working Party on Systemic Lupus Erythematosus. Medicine 1999; 78:167-75.
28. Ward MM. Hospital experience and mortality in patients with systemic lupus erythematosus. Arthritis Rheum 1999;42:891-8.
29. Ioannidis JPA, Haidich A-B, Lau J. Any casualties in the clash of randomised and observational evidence? BMJ 2001;322: 879-80.
30. Illei GG, Takada K, Parkin D, et al. Renal flares are common in patients with severe proliferative lupus nephritis treated with pulse immunosuppressive therapy: Long-term follow-up of a cohort of 145 patients participating in randomized controlled studies. Arthritis Rheum 2002;46:995-1002.
31. Felson DT, Anderson J. Evidence for the superiority of immunosuppressive drugs and prednisone over prednisone alone in lupus nephritis. Results of a pooled analysis. N Engl J Med 1984;311:1528-33.
32. Bansal VK, Beto JA. Treatment of lupus nephritis: a meta-analysis of clinical trials. Am J Kidney Dis 1997;29:193-9.
33. Boumpas DT, Balow JE. Outcome criteria for lupus nephritis trials: a critical overview. Lupus 1998;7:622-9.