

An Audit of Behcet's Syndrome Research: A 10-year Survey

FEHIM ESEN, ELIZABETH K. SCHIMMEL, HASAN YAZICI, and YUSUF YAZICI

ABSTRACT. Objective. Data suggest that the use of disease control groups and proper use of power calculations were neglected in published reports. We surveyed these and other methodological shortcomings in reports published within the last decade about one specific topic, Behcet's syndrome. We reason that recognizing such methodological shortcomings will lead to better quality clinical and basic science articles.

Methods. Articles published in the 15 highest impact factor journals on rheumatology, ophthalmology, dermatology, and general medicine between January 1999 and January 2009 were searched for original reports on Behcet's syndrome. Study designs (study types and time element), control groups, demographic data, use of power calculations, and reporting of negative results were specifically tabulated.

Results. Most studies on Behcet's syndrome were cross-sectional (83%). Prospective longitudinal studies were few (7%). In a considerable proportion of papers (21%), some basic demographic data were missing. Power calculations were rare (3%) even in randomized controlled trials and were not considered at all in clinical hypothesis-testing. Disease control groups were present in slightly over half of clinical and laboratory original research, while just 13% of genetic association studies included disease controls. Only 12% of all reports concerned mainly negative outcomes.

Conclusion. A considerable number of the published research articles have methodological weaknesses. The generalizability of what we observed in Behcet's syndrome to other research topics needs to be formally studied. (First Release Oct 15 2010; J Rheumatol 2011;38:99–103; doi:10.3899/jrheum.100335)

Key Indexing Terms:

BEHCET'S SYNDROME
POWER CALCULATIONS

METHODOLOGY

NEGATIVE TRIALS
BIAS

It has been proposed that certain “thought barriers” are present in our understanding of rheumatic diseases. The usual cross-sectional approach to rheumatological diseases is one important barrier, because these diseases are, for the most part, chronic¹. The lack of attention to the proper control groups, the backbone of specificity, can be another such barrier².

Our purpose was to survey a defined list of methodological issues in a systematic review of publications about Behcet's syndrome (BS) in selected medical journals over a specific time period. It is our hope that a better insight into these methodological problems will help to improve the quality of research in BS and its implementation.

From Cerrahpasa Medical Faculty and Division of Rheumatology, Istanbul University, Istanbul, Turkey; University of Mississippi Medical Center, Jackson, Mississippi; and the New York University (NYU) School of Medicine, NYU Hospital for Joint Diseases, New York, New York, USA.

F. Esen, MD, Cerrahpasa Medical Faculty, Istanbul University; E.K. Schimmel, MD, University of Mississippi Medical Center; H. Yazici, MD, Division of Rheumatology, Cerrahpasa Medical Faculty, Istanbul University; Y. Yazici, MD, NYU School of Medicine, NYU Hospital for Joint Diseases.

Address correspondence to Dr. Y. Yazici, NYU Hospital for Joint Diseases, 246 East 20th Street, New York, NY 10003, USA.

E-mail: yusuf.yazici@nyumc.org

Accepted for publication August 25, 2010.

MATERIALS AND METHODS

We made a PubMed search in the 15 highest impact factor journals of rheumatology, ophthalmology, dermatology, and general medicine (according to 2007 rankings) with the term “Behcet” and the limits “published in the last 10 years” and “English.” Only original research articles in which the main subject of the article was BS were selected. Case reports, letters, and reviews were excluded. Case series (reports of more than 1 case) were included into an appropriate group (Table 1) only if they aimed to lead to a specific clinical, laboratory, or management conclusion(s). Two observers read through each article independently to see whether they met the inclusion/exclusion criteria and to note the publication year, the study type as described in Table 1, the number of patients with BS studied, the gender distribution of patients, the mean age of patients, the makeup of the control groups, the presence of power calculations, and the reporting of negative results. Both observers entered/encoded their data in structured spreadsheets according to a standard data collection key. One researcher put the data side by side in Excel software and compared the data mainly by using Excel filters, and checked for accuracy. When there was disagreement between the observers, the authors came together and resolved the disagreement. When observers could not agree on data classification, one author made the final decision (n = 7).

In the data analysis, the time element (cross-sectional, prospective longitudinal, retrospective longitudinal), control groups, power calculations, and presence of negative results were considered only in the article types suitable for inclusion of such elements in the study design. Thus, the survey for the time element was not considered suitable for controlled drug trials, “drug use, other,” and genetic association studies. Because we were aware that none of the epidemiology studies were true incidence studies, the time element was also not studied in the related reports. Similar adjustments were made for other item inquiries (Table 2).

Table 1. Description of the study.

Study Types	
1.	Clinical descriptive: a study in which the aim is to define a certain clinical feature such as a symptom, a laboratory or a clinical finding, or the disease course under different conditions
2.	Clinical hypothesis-testing: any human study in which the aim is testing a clinical hypothesis
3.	Pure epidemiology: any field survey in which the main aim is to search for the prevalence or incidence of the disease by formal, population-based methods
4.	Epidemiology—other: any epidemiologic study that is not a true population-based field survey. This group includes, among others, attempts at diagnostic and classification criteria
5.	Controlled drug trial: any controlled trial of a drug
6.	Drug use— other: any human drug study other than a controlled drug trial
7.	Mainly laboratory: any experimental study in humans that is not a genetic association study
8.	Genetic association: any study in which the association between a disease or disease phenotype and a genetic polymorphism or mutation was studied
9.	Animal studies: any study based on animal experimentation
10.	Other: any study that does not meet any of the descriptions above
Time Elements	
1.	Cross-sectional study: in which the patients are seen/evaluated only once during the study period
2.	Longitudinal study (prospective or retrospective): in which a patient was followed by at least 2 patient visits during the study period
3.	Retrospective longitudinal study: surveying the fate of a cohort at the present time with a historical inception point in the past

Table 2. Adjustments in the analyses.

Time Element	Control Groups	Power Calculations	Negative Results
Article types included			
Clinical descriptive	Clinical hypothesis-testing	Clinical hypothesis-testing	Clinical hypothesis-testing
Clinical hypothesis-testing	Mainly laboratory studies	Controlled drug trial	Controlled drug trial
Mainly laboratory studies	Genetic association	Drug use, other	Drug use, other
Animal studies	Other	Mainly laboratory studies	Mainly laboratory studies
Other		Genetic association	Genetic association
		Animal studies	Animal studies
		Other	Other
Article types excluded			
Pure epidemiology	Clinical descriptive	Clinical descriptive	Clinical descriptive
Epidemiology, other	Pure epidemiology	Pure epidemiology	Pure epidemiology
Controlled drug trial	Epidemiology, other	Epidemiology, other	Epidemiology, other
Drug use, other	Controlled drug trial		
Genetic association	Drug use, other		
	Animal studies		

RESULTS

Initially, 602 articles were identified, most with BS as the main subject matter. Of those, 280 were full articles reporting original work. Most of the articles were published in rheumatology journals (61%), followed by ophthalmology (23%), dermatology (14%), and general medicine (2%). The number of published articles did not show significant changes year to year, with a range between 20 and 40 articles per year.

Table 3 summarizes the study types. It is to be noted that mainly laboratory studies and the genetic association studies — separately tabulated in this audit — added up to just over

half. On the other hand, there were very few animal studies and controlled clinical trials.

Table 4 shows that most of the BS work (83%) was cross-sectional. It also shows that 9.3% of the studies did not have any control groups, while 42% had a disease control group. The use of disease controls was highest (57.4%) in the laboratory studies, followed by the articles testing clinical hypotheses (52.6%). The use of disease control groups in genetic association studies was low (12.8%).

Some basic demographic data, crucial for correct interpretation of the results, was missing in 23/280 (8.2%) of the articles. In 55/280 articles (19.6%) the mean ages of the

Table 3. Study types and demographic data.

Study Type	% (n)	Mean No. of Patients with BD	Women, %	Mean Age, Yrs, of Patients with BD
Clinical, descriptive	18.9 (53)	174.4	43.7	36.6
Clinical, hypothesis-testing	6.8 (19)	59.4	38.4	35.1
Pure epidemiology	1.1 (3)	2.3/5708.6*	61.8	38.4
Epidemiology, other	3.2 (9)	144.2	42.4	37.3
Controlled drug trial	2.1 (6)	60.2	37.4	31.3
Drug use, other	11.8 (33)	19.7	41.7	33
Mainly laboratory	34.3 (96)	37.6	41.6	36
Genetic association	16.8 (47)	100	43.4	39.2
Animal studies	1.4 (4)	147	NA	NA
Other	3.6 (10)	81.1	33.7	37.4
Total	100 (280)	79.0	41.7	36.2

* Mean number of patients with BD identified/mean number of population surveyed. BD: Behçet's disease; NA: not applicable.

Table 4. Time element and control groups in the study designs of the articles surveyed. The survey of the time element was not considered relevant for studies having only a single possible study design (pure epidemiology, epidemiology–other, controlled drug trials, drug use–other, and genetic association). The study types clinical descriptive, pure epidemiology, epidemiology–other, controlled drug trials, drug use–other, and animal studies were not considered eligible for the survey of control groups and were excluded from this analysis.

Time Element	% (n)	Control Groups	% (n)	Use of Disease Control Groups	% (n)
Cross-sectional	83.0 (151)	No control groups	9.3 (16)	Clinical hypothesis testing	52.6 (10)
Prospective longitudinal	7.1 (13)	Healthy controls	82.0 (141)	Mainly laboratory	57.4 (55)
Retrospective longitudinal	9.9 (18)	Disease controls	41.9 (72)	Genetic association	12.8 (6)
		Intradisease controls	11.0 (19)		
		Family-based controls	1.7 (3)		

patients were not given, and 23/280 articles (8.2%) did not mention the gender ratio of the study population.

Power calculations were reported in only 3% of the studies and were not reported in any of the clinical hypothesis-testing and mainly laboratory studies. Only 1 in 6 (16.7%) of the controlled drug trials had power calculations. Similarly, power calculations were also scarce among the genetic association studies (4/47; 8.5%).

Studies with negative results were also rare (12%) and none of the 19 clinical hypothesis-testing studies had a negative result, while 2/6 drug trial reports were about the inefficacy of the drug studied. In genetic association studies 4/47 (8.5%) of the articles reported negative results.

We stratified the articles according to the journal specialty and the impact factor (Table 5). When we compared the 2 subspecialty journal groups with the highest number of articles (the rheumatology and ophthalmology journals), the only statistically significant difference was that the use of disease control groups was more frequent in the studies published in rheumatology journals ($p = 0.017$, chi-square = 5.69). We also compared the articles published in the journals with the top 5 impact factors and in those with the lowest 5 impact factors (ranking 11–15). There were fewer missing demographic data ($p = 0.037$, chi-square = 4.34) and the use of disease control groups was more common ($p =$

0.037, chi-square = 4.33) in the journals with the lowest impact factors, while the use of power calculations was more frequent in the 5 journals with the highest impact factors ($p = 0.039$, chi-square = 4.24). Other differences shown in Table 5 were not statistically significant.

DISCUSSION

There are many published reports of quality assessment in medical research across a wide range of disciplines. Much of this work focuses on therapeutic trials^{3,4,5,6,7} and usually on a single or a few issues such as funding sources⁸, study types⁹, or the correctness of references¹⁰. We are unaware of previous studies similar to ours either in BS or in other rheumatologic diseases.

We are also unaware of established guidelines to survey the methodological quality of research in a single disease, as we aimed to do. Well known research quality checklists such as CONSORT (Consolidated Standards of Reporting Trials)¹¹ or STROBE (Strengthening the Reporting of Observational studies in Epidemiology)¹² aim at specific types of research, randomized controlled trials for the former and epidemiologic work for the latter, rather than the whole spectrum of research on a certain condition.

This lack of established guidelines made it rather difficult to allocate many articles to some of the categories listed in

Table 5. Methodological issues stratified according to specialty and impact factor of the journals.

No. of Articles from Each Journal Type	Missing Demographic Data, %	Diseased Control Groups, %	Power Calculation, %	Negative Results, %
Rheumatology (n = 172)	20.4	49.6	4.6	16.7
Ophthalmology (n = 64)	25	25	0	4.4
Dermatology (n = 39)	23.1	28	0	6.1
Medicine (n = 5)	20	50	0	0
Top 5 journals in rheumatology (n = 67)*	28.4	42.9	8.8	12.3
Lowest 5 journals in rheumatology (n = 66)	13.6	65	0	19.6

* Top 5 and lowest 5 refer to journals with the highest and lowest impact factors.

Table 1. This was especially true in designating the most suitable allocation for case series. There is a need for widely acceptable quality assessment guidelines for similar work in the future.

There were further limitations to our survey. The survey included only journals published in English, and this can be considered a drawback. On the other hand, the 15 journals with the highest impact are all published in English (*Review of Rheumatism* has both French and English editions), which lessens the importance of this potential limitation. Another drawback was that animal studies and randomized controlled trials were too few to make meaningful methodological criticisms about these types of articles. It is to be noted, however, that the mere paucity of randomized clinical trials and animal studies are 2 important shortcomings of BS research in general.

Our survey brought up several important methodological issues. Although a chronic condition, most original work in BS is limited to cross-sectional studies. Only 31/182 (17%) of the relevant studies were longitudinal and of those, only 13 were both prospective and longitudinal. The remainder were retrospective longitudinal, i.e., they assessed groups of patients from a defined point in the past to the present time, a methodology surely more powerful than a cross-sectional survey but inferior to a true prospective work.

We observed disease control groups in 44% of the studies. This means that a substantial portion of original BS research still does not include disease control groups, a requirement for assessing the specificity of any biological finding. This methodological shortcoming is especially pronounced in genetic association studies. It must be emphasized that the need for specificity is as important in non-diagnostic studies addressing disease mechanisms as it is in clinical diagnostic work.

Our survey indicates that the lack of reporting of power calculations is a major methodological issue. Only in 1/6 of the controlled clinical trials, in which power calculations are most commonly used, was a power calculation present, while none of the clinical hypothesis-testing and laboratory investigations reported a power calculation. We suspect that what we observed in BS research on this issue is quite common in other areas of medical research. There is the preva-

lent and mistaken notion in the medical research community that power calculations are needed mainly in drug trials.

The course of BS, as in many other rheumatologic diseases, shows marked differences according to the sex and age of the patient¹³. Therefore, considering the age and sex of the study population is crucial when interpreting the outcome of a study. In 1/5 of the studies in our survey, one or both of these characteristics for the study population were missing.

The publication bias is probably the most frequently addressed methodological issue, along with less frequent publication of mainly negative data. On the other hand, and perhaps similar to our findings about power calculations, the main debate about publication bias concerns drug trials. It was interesting that none of the clinical hypothesis-testing and only 1/10 of the laboratory investigations was about negative results.

Our survey, for the first time, formally addressed a list of important methodological issues in BS research, the recognition and correction of which will be instrumental for more fruitful future investigations. We suggest that similar surveys on other more common diseases such as rheumatoid arthritis and systemic lupus erythematosus are also needed, and may give more robust results, with better external validity, because of the greater number of articles on these conditions.

REFERENCES

- Holman HR. Thought barriers to understand rheumatic diseases. *Arthritis Rheum* 1994;11:1565-72.
- Esen F, Celik A, Yazici H. The use of diseased control groups in genetic association studies. *Clin Exp Rheum* 2009;27 Suppl 53:S4-5.
- Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Med* 2008;5:e217.
- Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
- Singh JA, Murphy S, Bhandari M. Assessment of the methodologic quality of medical and surgical clinical trials in patients with arthroplasty. *J Rheumatol* 2009;36:2642-54.
- Abrahamyan L, Johnson SR, Beyene J, Shah PS, Feldman BM. Quality of randomized clinical trials in juvenile idiopathic arthritis.

- Rheumatology 2008;47:640-5.
7. Yazici Y, Adler NM, Yazici H. Most tumour necrosis factor inhibitor trials in rheumatology are undeservedly called 'efficacy and safety' trials: a survey of power considerations. *Rheumatology* 2008;47:1054-7.
 8. Davis JM, Chen N, Glick ID. Issues that may determine the outcome of antipsychotic trials: industry sponsorship and extrapyramidal side effect. *Neuropsychopharmacology* 2008;33:971-5.
 9. Funai EF, Rosenbush EJ, Lee MJ, Del Priore G. Distribution of study designs in four major US journals of obstetrics and gynecology. *Gynecol Obstet Invest* 2001;51:8-11.
 10. Reddy MS, Srinivas S, Sabanayagam N, Balasubramanian SP. Accuracy of references in general surgical journals — an old problem revisited. *Surgeon* 2008;6:71-5.
 11. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-9.
 12. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806-8.
 13. Yurdakul S, Yazici H. Behçet's syndrome. *Best Pract Res Clin Rheumatol* 2008;22:793-809.