Bias in Uncontrolled Therapeutic Trials in Rheumatology Due to Selection of Populations with Extreme Characteristics

JOHN P.A. IOANNIDIS and FOTINI B. KARASSA

ABSTRACT. ***Objective.*** To assess the prevalence of biases from selection of patients with extreme characteristics in recent uncontrolled therapeutic studies in rheumatology.

Methods. We hand searched 4 major rheumatology journals for uncontrolled trials published in 1997 or 1998 that measured therapeutic efficacy by comparing one or more variables at followup vs at baseline. We evaluated the susceptibility to bias from random measurement error and natural variability for variables used for defining eligibility that overlap with those used for defining outcomes.

Results. Twenty-five studies were analyzed. In 22 studies, the eligibility criteria were related to the outcome criteria and defined a patient population with extreme characteristics. Only 3 studies clearly reported that they had performed a baseline measurement separate from the screening (eligibility) measurement. The remaining 19 reports (76%) might be susceptible to bias: in 7, identical variables were used for eligibility criteria and outcomes; 3 used outcome variables that were also used for characterizing eligibility along with other criteria; 2 used specific eligibility variables that were part of composite outcome scores; and 7 selected patients on the basis of vague descriptors of disease severity, while disease severity was also the outcome.

Conclusion. Several recent uncontrolled trials of therapeutic interventions in rheumatology are subject to biases stemming from the selection of patients with extreme characteristics. Baseline evaluations separate from the screening measurements should be performed and eligibility criteria and outcomes should be carefully defined. (J Rheumatol 2001;28:1881–7)

Key Indexing Terms:
UNCONTROLLED STUDIES REGRESSION-TO-THE-MEAN SELECTION BIAS

Controlled trials are the gold standard for assessing medical interventions[1,2]. However, controlled designs are not always feasible and uncontrolled designs are often used in evaluating therapeutic interventions. Uncontrolled designs are favored in early development phases, but for several therapeutic questions of interest in rheumatology and in other fields, controlled studies are never performed.

Uncontrolled studies are subject to biases. One particular concern arises when populations are selected for inclusion on the basis of some extreme variable and then the study measures the response to treatment of this same variable. Random measurement error and natural variability in the eligibility/outcomes variables may spuriously inflate the observed treatment effect[3-5], unless special precautions are taken in the study design. We evaluated the prevalence and implications of these problems in the recent rheumatology literature.

From the Clinical Trials and Evidence-Based Medicine Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and the Division of Clinical Care Research, Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts, USA; and the Department of Rheumatology Research, the Windeyer Institute of Medical Sciences, University College, and Middlesex Hospital Medical School, London, UK.

J.P.A. Ioannidis, MD, Associate Professor, Chairman of Hygiene and Epidemiology, University of Ioannina School of Medicine, Associate Professor of Medicine, Tufts University School of Medicine; F.B. Karassa, MD, Rheumatologist, Department of Rheumatology Research, Windeyer Institute of Medical Sciences.

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

Submitted October 18, 2000; revision accepted February 15, 2001.

MATERIALS AND METHODS

Definitions and Theory

Bias for continuous variables. The typical regression-to-the-mean bias occurs when 3 criteria are met: (1) the study population is selected using inclusion/exclusion criteria that set a certain cutoff value in a variable for patients to be eligible; (2) the same variable is used for assessing outcome during followup; and (3) outcome assessment is based on the comparison of a followup measurement against a single screening measurement. Regression-to-the-mean would not be a problem if the selected patients have one or more measurements other than the screening value before they are included and the mean of these extra measurements is used as the baseline value. If only the last qualifying measurement is used as the baseline, regression-to-the-mean is still present, regardless of how many increased values have been recorded previously. Although regression-to-the-mean may also occur in randomized trials when patients are selected with cutoff values, randomization obviates the problem if the comparison is not focused on the before-after change in one arm but in the change between the 2 randomized arms.

Regression-to-the-mean is a consequence of the random error that accompanies any single measurement[3,4]. Problems arise when eligibility criteria are based on screening values above or below a certain cutoff value. Upon remeasurement, the selected patients with extreme values tend, on average, to have less extreme values.

An additional consideration when patients are selected based on an

extreme value is whether there is natural variability over time. When regular (e.g., circadian) or irregular variability exists, patients with extreme values may exhibit less extreme values upon remeasurement. This bias is not always corrected by using a baseline measurement separate from the screening measurement. The ability to correct this bias depends on the timing of measurements, the nature of the variability, and on whether we know when peaks and troughs occur. Unfortunately, for many variables, variability may not fit determinable patterns.

Bias for categorical variables. Similar problems occur when patients are selected for being in the worst possible state as defined by a screening categorical variable. For example, patients with rheumatoid arthritis (RA) may be selected on the basis of having failed completely to respond to previous therapy. With 2 categories conceptualized, response and failure, patients are selected from the worst category. Unless we specify gradations within the unfavorable category, patients entering a study in that category may either continue in that category or may reach the "response" category during followup — they cannot be classified as doing worse. Bias may also occur if patients are selected for being in the best possible state. For example, a study may recruit exclusively patients in "complete remission." Such patients may either stay in remission or flare during followup — they cannot do better.

The reasoning for the bias is the same as for continuous variables. A categorical evaluation is also a measurement that entails random measurement error (wrong categorization) and natural variability (spontaneous changes in category). The extent of natural variability and the likelihood of major transitions in category differ across various diseases.

Related, but not identical, screening and outcome variables. The typical biases we discussed assume that the screening and outcome variables are identical. Sometimes the screening variable may be strongly related, but not identical, to the outcome variable. For example, the outcome may be one component of a disease severity scale and the full severity scale may have been used as the screening variable — or vice versa. Or screening and outcome variables may be composite scores that share in some components. Random error and natural variability may then affect the shared components; however, the overall bias on outcome evaluation is difficult to model.

Often some eligibility and outcome criteria are identical, but there are additional eligibility criteria as well. The additional eligibility criteria supplant in part the role of second measurements, because patients have to pass several criteria to be eligible; this is less likely to happen by chance alone.

Natural course of the disease. Finally, when followup outcomes are assessed, the longterm natural course of the disease must be considered. The effect of the longterm natural course of the disease will not be discussed in detail. Its importance may be suspected from available longitudinal data, but it can never be fully accounted for in uncontrolled studies, since we do not know how the patients would have evolved without the study treatment[6]. It is often impossible to have a control group representing the natural course of the disease, especially when treatments exist that have been proven to be effective.

Thus, the outcome change D in an uncontrolled study may be simplified as: $D = T + R + V + N + e$, where T is the true treatment effect, R is the regression-to-the-mean effect due to random error, V is the effect of short term natural variability, N is the effect of the longterm natural course of the disease, and e is an error term. The assumption $D = T$ is oversimplified.

Database of Studies, Eligibility Criteria, and Data Extraction

We hand searched 4 rheumatology journals (*Annals of the Rheumatic Diseases*, *Arthritis and Rheumatism*, *British Journal of Rheumatology*, and *Journal of Rheumatology*) for articles published in 1997 and 1998. Articles were included if (1) therapeutic interventions were evaluated; (2) no control arm was used; and (3) outcome evaluation was based on the comparison of one or several variables during followup vs the same variables at study entry.

For each eligible report, we extracted the following information: year, journal, authors, inclusion criteria, exclusion criteria, definitions of outcome variables, and whether one or several baseline measurements were obtained. Inclusion and exclusion criteria were examined to assess which ones would result, by definition, in the selection of a patient population with extreme characteristics. We noted also whether screening and outcome variables were related, and if so, to what extent.

Hand searching was performed by a rheumatologist. The selected articles were discussed to ensure eligibility. Data extraction was performed in duplicate by a rheumatologist and a methodologist.

RESULTS

We found 25 eligible studies[7-31] published in 1997 (n = 11) or 1998 (n = 14) (Table 1). More than half of the reports dealt with RA (n = 13) and 5 were on systemic lupus erythematosus (SLE). Diverse therapeutic regimens were tested, with cytotoxic drugs being most prominent. Most studies were small, with 13 reports having 20 or fewer patients, but 5 studies had over 100 patients each. Followup varied from one month to many years (Table 1).

Table 2 shows the inclusion and exclusion criteria and the outcome variables of each of the 25 studies. Eligibility criteria had direct or indirect relationship with outcome variables in all but 3 studies[9,20,24]. Nineteen of the remaining 22 reports studied patients selected with screening determinations suggestive of very active disease. The opposite occurred in one study[25], which included only patients with mild or moderate disease. Finally, 2 studies excluded some patients with mild disease, as well as some patients with active/severe disease[13,14].

A separate baseline evaluation was explicitly mentioned to have been performed in 3 of the 22 reports where eligibility criteria and outcomes overlapped[12,29,31]. Of the remaining 19 studies, three[7,13,27] used outcome variables that were also used for characterizing eligibility along with other criteria; bias is difficult to decipher. Two studies[14,25] selected patients based on the presence or absence of specific clinical manifestations that were part of the outcome scale used for evaluation of response; bias should be small, since the composite score is influenced by many other components. Seven studies[8,10,16,19,21,26,28] selected patients based on nonspecific descriptors of disease severity (e.g., "unsatisfactory response to treatment," "active inflammation," "unremitting disease") and then used outcome variables that similarly reflected disease severity. It is unknown whether patient selection was based on more detailed, unreported rules or no rules were set. Nevertheless, the lack of specific, reproducible eligibility rules casts doubt on the selection process for study entry. Selection bias may also have affected data analyses, if vague eligibility criteria were also applied retrospectively for selecting patients with more favorable responses. Finally, in 7 studies some eligibility criteria and outcomes were identical variables. We will describe these cases in more detail.

Martin-Suarez, *et al*[15] evaluated the effect of low dose cyclophosphamide in severe connective tissue diseases. Per

Table 1. Characteristics of studies included in the analysis.

Author	Year	Therapy	Disease	Sample Size	Followup, mo
Kavanaugh[7]	1997	Murine Mab to ICAM-1	RA	8	4
McGonagle[8]	1997	G-CSF + methylprednisolone	RA	5	1 day
Youssef[9]	1997	Pulse methylprednisolone	RA	18	1/30
Caccavo[10]	1997	Cyclosporin A	SLE	30	24
Wallace[11]	1997	IV cyclophosphamide + IV methylprednisolone	JRA	4	31–35
Menon[12]	1998	IV prostacyclin	SSc/pulmonary hypertension	7	1 day
Davis[13]	1998	2-chloro-2′deoxyadenosine	SLE/nephritis	12	12
Gansauge[14]	1997	MTX	SLE	22	6
Martin-Suarez[15]	1997	IV cyclophosphamide	Various	90	5–213 (median 56)
Munro[16]	1998	IM gold	RA	440	60
Ruperto[17]	1998	MTX	JRA (chronic)	111	6
Rau[18]	1997	MTX	RA	271	12–108
Marchesoni[19]	1997	Cyclosporin A	Adult Still's disease	6	8–60
Reynoso-von Drateln[20]	1997	IV cyclophosphamide	Pyoderma gangrenosum	9	6–36
Ravelli[21]	1998	MTX	Pediatric onset SLE	11	7–23
Davis[22]	1998	Fludarabine	RA	26	12
Weinblatt[23]	1998	MTX	RA	26	132
Flipo[24]	1998	Cyclosporin A Neoral	RA	28	4.5
Van Vollenhoven[25]	1998	Dehydroepiandrosterone	SLE	50	6–12
Maksymowych[26]	1998	Pamidronate	Ankylosing spondylitis	16	9
Varai[27]	1998	IV cyclophosphamide	SSc/pulmonary fibrosis	5	12
Bologna[28]	1997	MTX	RA	453	3–106 (mean 35.2)
Kiely[29]	1998	Oxpentifylline	RA	20	3
Fairney[30]	1998	Cyclical etidronate	Osteoporosis	115	48
Wassenberg[31]	1998	MTX	RA/Felty's syndrome	7	12

ICAM-1: intercellular adhesion molecule 1; IM: intramuscular; IV: intravenous; JRA: juvenile rheumatoid arthritis; Mab: monoclonal antibody; RA: rheumatoid arthritis; SLE: systemic lupus erythematosus; SSc: systemic sclerosis; MTX: methotrexate.

entry criteria, patients were not in remission; the outcome was similarly clinical remission. In addition to the tested regimen, most patients also received other treatment modalities during followup. The authors conclude that the proposed regimen "compares in efficacy with the higher monthly doses previously advocated,"[15] apparently based on historical data.

Ruperto, *et al*[17] used the number of swollen joints with active arthritis as both an eligibility criterion and outcome criterion. They found significant improvements in patients with juvenile chronic arthritis treated with methotrexate (n = 111). Data are presented without formal statistical testing. Five other outcome variables were used, also linked indirectly to eligibility criteria.

In a study of methotrexate for severe RA[18], joint counts and erythrocyte sedimentation rate were used as variables for both defining eligibility cutoffs and outcomes. The first followup measurements show considerable improvement compared with baseline; subsequent measurements show no further change. This pattern may well represent an effective treatment that produces all its benefit up to the first followup measurement. However, bias from extreme selection would cause exactly the same pattern, even for a totally ineffective treatment: a spuriously improved first followup measurement and no change in subsequent measurements. Given the study design, the relative contribution of true efficacy and bias cannot be deciphered. Another concern is losses to followup: only 48 of 271 patients were evaluated at 48 months. Weinblatt, *et al*[23] (longterm extension of a randomized trial) may be susceptible to similar biases.

Davis, *et al*[22] used swollen joint count as both an eligibility variable and an outcome measure for evaluating fludarabine in severe refractory RA; other outcomes also overlapped with the eligibility variables. The study also compared 2 different doses, and the authors concluded that "fludarabine resulted in clinical improvement in the high-dose group," although the comparison of high vs low dose was nonsignificant (p = 0.28). Inference was based on the comparison of each "arm" against baseline values, an approach subject to the aforementioned biases.

Fairney, *et al*[30] used cyclical etidronate in patients with radiological evidence of osteoporosis (compression fractures) and in patients with low lumbar spine bone mineral density (BMD). The evaluation of spine BMD in the second group is susceptible to bias. Interestingly, the authors reported a significant effect in the lumbar spine BMD, but no effect at an independent site (femoral neck). Although the difference may reflect chance or a preferential spinal effect, bias could have caused this divergence.

Finally, Wallace and Sherry[11] report on 4 patients given

Table 2. Eligibility criteria and outcomes in 25 analyzed uncontrolled rheumatology studies.*

First author	Inclusion criteria	Exclusion criteria	Outcomes and results
Kavanaugh	1. Criteria for RA, 2. *Active disease (SJC ≥6 plus 2 of: TJC ≥9, EMS ≥45 min, ESR ≥28mm/h)*	None	*Clinical response (>20% improvement in ≥4/6 parameters:TJC, SJC, EMS,* patient's and physician's assessement of disease activity, *ESR): 4/8 achieved in second course of treatment*
McGonagle	1. Criteria for RA, 2. *Ongoing active inflammation (failure in ≥ 3 DMARD), RF (+), erosive arthropathy*	None	*CRP: E, RF: E, EMS: ND, SJC: ND, TJC: ND, VAS score: ND,* White blood cell count: E, neutrophil count: E, CD34+ cell count: E
Youssef	1. Criteria for RA, 2. At least 1 inflamed knee joint with an effusion	None	VAS score: SE, TJC: SE, HAQ score: SE, CRP: SE, various cytokines in serum, synovial fluid/membrane: SE
Caccavo	1. Criteria for SLE, 2. *Poor response to previous therapy* 3. *≥2 organ systems,* 4. Stable corticosteroid dose	1. *Creatinine > 1.5 mg/dl,* 2. *Hypertension,* 3. Cancer, 4. Previous therapy with cyclosporine	*SLAM score: SE*
Wallace	1. Criteria for systemic-onset JRA, 2. *Unremmitting polyarthritis despite maximum therapy,* 3. Joint destruction, 4. Corticosteroid-dependence, 5. Severe growth retardation	None	*Clinical remission (EMS <15 min, no fatigue, no joint swelling or pain for 2 consecutive months): achieved in 3/4,* linear growth velocity: E, corticosteroid dose: E
Menon	1. Criteria for SSc, 2. Dyspnea, 3. Reduced CO diffusing capacity (<50% of predicted), 4. Elevated pulmonary artery systolic pressure (>25 mmHg) on Doppler echo	None	Pulmonary artery pressure by catheterization: SE, Pulmonary vascular resistance: SE, Cardiac output: SE
Davis	1. Criteria for SLE, 2. Age>18y, 3. Contraception, 4. Nephritis (type III, IV or mixed IV+V) *with proteinuria > 3gr/24h or 10 RBC/hpf and cellular casts, 5. Serologically active SLE (anti-dsDNA≥1/160 or C3<43 or C4<14)*	1. Infection, 2. Pregnancy, lactation, 3. Malignancy, 4. Cardiac disease, 5. Other cytotoxic agents, 6. Medical illness, 7. *Serious active or chronic SLE (various criteria listed)* 8. Splenectomy	*SLAM score: ND, Renal response (<1mg/protein/24h, absence of cellular casts and <10 RBC/hpf): E, White blood cells: SE, platelets: ND, anti-dsDNA:ND, C3: ND, C4: ND,* CD3+: SE, CD20+: ND, CD5+: ND, CD16+:ND
Gansauge	1. Criteria for SLE, 2. *Refractory cutaneous rashes or active dermal vasculitis or active pleurisy or active arthritis*	1. *Active nephritis or CNS disease,* 2. Impaired renal function or abnormal liver enzymes	*SLEDAI score: SE,* ESR: SE, *C3: ND, C4: ND, anti-dsDNA: TE*
Martin-Suarez	1. Diagnosis of CTD, 2. Severe, progressive disease (relapse or new organ involvement despite previous therapy)	None	Complete/partial disease remission: E, *ESR: SE, Hb/PLT/CRP (only in vasculitides): SE, anti-dsDNA: SE, C3: SE, C4: SE, Cr: ND, proteinuria (in SLE): SE*
Munro	1.Criteria for RA, 2. *Active, uncontrolled with NSAID or alternative DMARD, inflammatory polyarthropathy*	None	*HAQ score: SE (ND for groups 2,3), EMS: SE, ESR: SE, CRP: SE, Ritchie index: SE, VAS score: SE (ND for group 2, 3)*
Ruperto	1. Criteria for JCA, 2. Disease duration ≥6 months, 3. ≥ 5 joints with active arthritis, 4. *Not adequately controlled with NSAID or DMARD*	None	*Clinical response: E, Physician global assessment: E, Parent global assessment: E,* Functional ability: ND, *number of joints with limited motion: E,* number of joints with active arthritis: E, *ESR: E*
Rau	1. Criteria for RA, 2. Active disease: SJC > 6 and TJC > 9, ESR > 20mm/h (men) or > 30 mm/h (women)	1. Malignancy, chronic liver disease, alcohol abuse, or significant renal impairment, 2. Low WBC 3. Low PLT, 4. Peptic ulcer	SJC: SE, Grip strength: SE, Patient global assessment of pain: SE, Patient global assessment of mobility: SE, ESR: SE
Marchesoni	1. Diagnosis of adult Still's disease, 2. *Treatment with NSAIDs and/or corticosteroids deemed unsatisfactory (unclear on which criteria)*	1. Adverse reaction to CSA, 2. Uncontrolled hypertension, 3. Serious cardiac, hepatic or renal disease, 4. Other second line agents	*4/6 achieved complete disease remission*
Reynoso-von Drateln	1. Clinical diagnosis of pyoderma gangrenosum	1. Comorbidities, 2. Bacteria or fungus, 3. Other immunosuppressives, 4. Pregnancy, 5. Cardiac/hepatic disease, 6. Malignancy, 7. Infection, 8. Low blood counts, 9. High creatinine	Assigned diameter value: E (complete/partial response in 8/9 patients)
Ravelli	1. Criteria for SLE, 2. *Active disease: inadequately controlled by steroids for ≥3 months, or flare*	None	*SLAM score: ND, SLEDAI score: ND,* ESR: ND, *C3: ND, C4: ND, anti-dsDNA: ND,* Proteinuria: ND, Hematuria: ND
Davis	1. Criteria for RA, 2. Age ≥18 y, 3. Duration of RA > 6 months, 4. Active RA with SJC≥ 6 and at least 2 of the following: TJC≥ 9, EMS≥ 45 min, ESR≥ 28mm/h, 5. *Failed other treatments,* 6. Contraception	1. Infection, 2. Hematologic disorder, 3. Significant renal or cardiac disease, 4. Malignancy, 5. Neurologic, pulmonary or other confounding illness, 6. Low gamma globulins	*2/12 in the low dose and 7/14 in the high dose groups had 50% or greater reduction in TJC and/or SJC (after 6months of therapy) compared to baseline: E, TJC: SE,* SJC: SE, *HAQ: E, CRP: E, RF: SE, physician global assessment: E*

Table 2. Eligibility criteria and outcomes in 25 analyzed uncontrolled rheumatology studies.* *(cont.)*

First author	Inclusion criteria	Exclusion criteria	Outcomes and results
Weinblatt	1. Criteria for RA (long-term follow-up of 26 patients who completed a 24-week randomized study. Initial inclusion criteria included *active disease:* SJC≥3, TJC≥6, EMS>45min, ESR>28mm/h)	None	TJC: SE, SJC: SE, ESR: E
Flipo	1. Criteria for RA, 2. Treatment with stable Sandimmune doses	1. Renal, liver or hemopoietic disease, 2. Hypertension	SJC: ND, Ritchie index: ND, patient global assessment: ND, physician global assessment: ND, CRP, ESR, fibrinogen: ND, HAQ score: ND
Van Vollenhoven	1. Criteria for SLE, 2. Age > 18 y, 3. *Mild or moderate disease (no severe nephritis or CNS)*	None	*SLEDAI score: SE*, patient global assessment: SE, physician global assessment: SE, ESR/complete blood count: ND
Maksymowych	1. Criteria for ankylosing spondylitis, 2. *Active inflammation despite maximum tolerated doses of NSAID*	None	Group I: *BASDAI score: SE, BASFI index: ND, BASMI index: SE*, ESR: SE, Hb: ND, Group II: *BASDAI, BASMI, BASFI*, ESR, Hb: ND
Varai	1. Patients with SSc and pulmonary interstitial involvement (*FVC <=80% of predicted value, or ILD*)	None	Dyspnea score: SE, *FVC: ND, modified Muller score of parenchymal involvement: ND*
Bologna	1. Patients with RA *who were treated with MTX* (apparently had severe disease)	None	*SJC: SE, Ritchie index: SE, ESR: SE, Hb: SE, Platelets: SE, CRP: ND, Larsen's score: SE*
Kiely	1. Criteria for RA, 2. ≥3 criteria: (2 consecutive occasions, ≥1 week apart): *EMS> 45min, TJC or SJC≥ 3, ESR >25mm/h, CRP > 20mg/l*	None	Paulus criteria *(20-50% improvement): 50%, SJC, TJC: ND, EMS: SE*, Pain-Grip: ND, *HAQ: SE*, Patient-physician global assessment: ND, *ESR/CRP: ND*
Fairney	1. Patients with radiological evidence of osteoporosis and fractures or 2. Patients with lumbar spine bone density < 2 SD below the young adult reference	None	Lumbar spine bone density: SE, Femoral neck bone density: ND
Wassenberg	1. Criteria for RA, 2. Felty's syndrome (splenomegaly, granulocytopenia<2000/mm^3 at any time), 3. No other causes for splenomagaly or neutropenia	None	Granulocytes: SE, ESR: ND, SJC: SE

*Criteria and outcomes in bold type are practically identical variables; criteria and outcomes in italics are overlapping variables. E: effective (no statistical testing), SE: significantly effective, TE: trend for effectiveness, ND: no difference, BAL: bronchoalveolar lavage, CNS: central nervous system, CRP: C-reactive protein, CTD: connective tissue disease, DMARD: disease modifying antirheumatic drugs, EMS: early morning stiffness, Echo: echocardiography, ESR: erythrocyte sedimentation rate, FVC: forced vital capacity, Hb: hemoglobin, ILD: interstitial lung disease, JCA: juvenile chronic arthritis, JRA: juvenile rheumatoid arthritis, MTX: methotrexate, NSAID: nonsteroidal antiinflammatory drug, RA: rheumatoid arthritis, RBC/hpf: red blood cells per high power field, RF: rheumatoid factor, SD: standard deviation, SJC: swollen joint count, SLE: systemic lupus erythematosus, SSc: systemic sclerosis, TJC: total joint count, VAS: visual analog scale, PLT: platelets, CSA: cyclosporine A, WBC: white blood cell count.

intravenous pulse cyclophosphamide and methylprednisolone for severe systemic onset juvenile RA. Linear growth rate and corticosteroid dose were used both as eligibility variables and as outcomes. The impressive results are unlikely to be accounted for only by bias, but the small sample size suggests cautious interpretation.

DISCUSSION

Evaluation of recent therapeutic literature from leading rheumatology journals shows that uncontrolled studies in the field are often susceptible to biases that stem from the selection of patients with extreme characteristics. Eligibility criteria were totally independent of outcomes in only 3 of 25 reports. Separate baseline and screening measurements were explicitly reported in only 3 cases. Thus 19 of 25 reports (76%) might be susceptible to bias. Finally, eligibility criteria were often poorly defined.

Selecting patients with extreme characteristics is not a fault per se. Such patients may indeed be the legitimate target of new therapies. New regimens are needed most for patients with severe disease and where existing therapies fail. However, it is important that such populations are carefully defined so that the derived information can be translated to medical practice. Seven studies in our analysis (28%) used very vague descriptors of disease severity that may easily be misinterpreted during the study conduct and analysis. Selection of extreme cases already compromises the external validity of any study, since results pertain only to a few patients. A poor definition of disease severity also compromises the internal validity.

Careful definition of eligibility criteria should consider whether these overlap with criteria used for outcome assessment. If this is unavoidable, the screening measurement should not be used as the baseline measurement for the eval-

uation of response during followup. A separate baseline measurement is mandatory and will alleviate at least the bias due to random measurement error.

Several therapeutic studies performed in rheumatology are still uncontrolled. The evocation of historical controls is problematic, since historical controls may spuriously inflate the magnitude of the treatment effect[6]. Uncontrolled studies are unable to account for the natural course of the disease. The problem is magnified when selected patients are used, and further exaggerated when there are also losses of patients during followup, since poor responders are often selectively lost to followup[32].

The increased technical, organizational, and financial requirements of randomized trials, especially longterm trials, and the belief that new treatments should not be spared from study participants, especially those with severe or terminal disease, sometimes pose difficulties for the implementation of randomized trials. Moreover, many important advances in rheumatology have relied mostly on well designed uncontrolled studies or nonrandomized controlled studies, and randomized studies themselves are not devoid of selection biases. Uncontrolled studies will unavoidably continue to be performed in the field. Such studies may still give us very useful information. It is thus important to give due attention to the design of these observational studies by avoiding the biases we have described.

ACKNOWLEDGMENT

We are thankful to Dr. P.G. Vlachoyiannopoulos and Prof. H.M. Moutsopoulos for insightful discussions.

REFERENCES

1. Byar DP. Why data bases should not replace randomized clinical trials. Biometrics 1980;36:337-42.
2. Byar DP. Problems with using observational databases to compare treatments. Stat Med 1991;10:663-6.
3. Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. Am J Epidemiol 1976;104:493-8.
4. Yudkin PL, Stratton IM. How to deal with regression to the mean in intervention studies. Lancet 1996;347:241-3.
5. Stigler SM. Regression towards the mean, historically considered. Stat Meth Med Res 1997;6:103-14.
6. Sacks H, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. Am J Med 1982;72:233-40.
7. Kavanaugh AF, Schulze-Koops H, Davis LS, Lipsky PE. Repeat treatment of rheumatoid arthritis with a murine anti-intercellular adhesion molecule 1 monoclonal antibody. Arthritis Rheum 1997;40:849-53.
8. McGonagle D, Rawstron A, Richards S, et al. A phase 1 study of safety and efficacy of granulocyte colony-stimulating factor for the mobilization of hematopoietic progenitor cells in active rheumatoid arthritis. Arthritis Rheum 1997;40:1838-42.
9. Youssef PP, Haynes DR, Triantafillou S, et al. Effects of pulse methylprednisolone on inflammatory mediators in peripheral blood, synovial fluid, and synovial membrane in rheumatoid arthritis. Arthritis Rheum 1997;40:1400-8.
10. Caccavo D, Lagana B, Mitterhofer AP, et al. Long-term treatment of systemic lupus erythematosus with cyclosporin A. Arthritis Rheum 1997;40:27-35.
11. Wallace CA, Sherry DD. Trial of intravenous pulse cyclophosphamide and methylprednisolone in the treatment of severe systemic-onset juvenile rheumatoid arthritis. Arthritis Rheum 1997;40:1852-5.
12. Menon N, McAlpine L, Peacock AJ, Madhok R. The acute effects of prostacyclin on pulmonary hemodynamics in patients with pulmonary hypertension secondary to systemic sclerosis. Arthritis Rheum 1998;41:466-9.
13. Davis JC Jr, Austin H III, Boumpas D, et al. A pilot study of 2-chloro-2'-deoxyadenosine in the treatment of systemic lupus erythematosus-associated glomerulonephritis. Arthritis Rheum 1998;41:335-43.
14. Gansauge S, Breitbart A, Rinaldi N, Schwarz-Eywill M. Methotrexate in patients with moderate systemic lupus erythematosus (exclusion of renal and central nervous system disease). Ann Rheum Dis 1997;56:382-5.
15. Martin-Suarez I, D'Cruz D, Mansoor M, Fernades AP, Khamashta MA, Hughes GRV. Immunosuppressive treatment in severe connective tissue diseases: effects of low dose intravenous cyclophosphamide. Ann Rheum Dis 1997;56:481-7.
16. Munro R, Hampson R, McEntegart A, Thomson EA, Madhok R, Capell H. Improved functional outcome in patients with early rheumatoid arthritis treated with intramuscular gold: results of a five year prospective study. Ann Rheum Dis 1998;57:88-93.
17. Ruperto N, Ravelli A, Falcini F, et al, for the Italian Pediatric Rheumatology Study Group. Performance of the preliminary definition of improvement in juvenile chronic arthritis patients treated with methotrexate. Ann Rheum Dis 1998;57:38-41.
18. Rau R, Schleusser B, Herborn G, Karger T. Longterm treatment of destructive rheumatoid arthritis with methotrexate. J Rheumatol 1997;24:1881-9.
19. Marchesoni A, Ceravolo GP, Battafarano N, Rossetti A, Tosi S, Fantini F. Cyclosporin A in the treatment of adult onset Still's disease. J Rheumatol 1997;24:1582-7.
20. Reynoso-von Drateln C, Perla-Navarro P, Gamez-Nava JI, Gonzalez-Lopez L, Galvan-Villegas F, Ramos-Remus C. Intravenous cyclophosphamide pulses in pyoderma gangrenosum: an open trial. J Rheumatol 1997;24:689-93.
21. Ravelli A, Ballardini G, Viola S, Villa I, Ruperto N, Martini A. Methotrexate therapy in pediatric onset systemic lupus erythematosus. J Rheumatol 1998;25:572-5.
22. Davis JC Jr, Fessler BJ, Tassiulas IO, et al. High dose versus low dose fludarabine in the treatment of patients with severe refractory rheumatoid arthritis. J Rheumatol 1998;25:1694-704.
23. Weinblatt ME, Maier AL, Fraser PA, Coblyn JS. Longterm prospective study of methotrexate in rheumatoid arthritis: conclusion after 132 months of therapy. J Rheumatol 1998;25:238-42.
24. Flipo RM, Emery P, Scott DGI, et al. Safety and tolerability of conversion from stable Sandimmun maintenance treatment to Sandimmun Neoral in patients with rheumatoid arthritis. J Rheumatol 1998;25:1263-9.
25. Van Vollenhoven RF, Morabito LM, Engleman EG, McGuire JL. Treatment of systemic lupus erythematosus with dehydroepiandrosterone: 50 patients treated up to 12 months. J Rheumatol 1998;25:285-9.
26. Maksymowych WP, Jhangri GS, Leclercq S, Skeith K, Yan A, Russell AS. An open study of pamidronate in the treatment of refractory ankylosing spondylitis. J Rheumatol 1998;25:714-7.
27. Varai G, Earle L, Jimenez SA, Steiner RM, Varga J. A pilot study of intermittent intravenous cyclophosphamide for the treatment of systemic sclerosis associated lung disease. J Rheumatol 1998;25:1325-9.
28. Bologna C, Viu P, Picot MC, Jorgensen C, Sany J. Long-term follow-up of 453 rheumatoid arthritis patients treated with

methotrexate: an open, retrospective, observational study. Br J Rheumatol 1997;36:535-40.
29. Kiely PDW, Johnson D, Bourke BE. An open study of oxpentifylline in early rheumatoid arthritis. Br J Rheumatol 1998;37:1033-5.
30. Fairney A, Kyd P, Thomas E, Wilson J. The use of cyclical etidronate in osteoporosis: changes after completion of 3 years treatment. Br J Rheumatol 1998;37:51-6.
31. Wassenberg S, Herborn G, Rau R. Methotrexate treatment in Felty's syndrome. Br J Rheumatol 1998;37:908-11.
32. Ioannidis JPA, Bassett R, Hughes MD, Volberding P, Sacks HS, Lau J. Predictors and impact of patients lost to follow-up in a randomized trial of early versus deferred antiretroviral treatment. J Acq Immune Defic Syndr 1997;16:22-30.