Psoriasis is a chronic skin disorder affecting 1–3% of the general population. It is characterized by multiple erythematous and scaly plaques that may be disfiguring and a course that is typically punctuated by exacerbations and remissions (Stern, 1997). The causes are unknown but genetic–environmental interaction seems to offer a plausible etiological explanation. Keratinocyte hyperproliferation and inflammatory changes are constant pathological features (Barker, 1991). Although the disease does not affect mortality, it may substantially affect the quality of life and in several countries and physicians (Stern et al, 1999; Van de Kerkhof et al, 2000; Zachariae et al, 2001). The EDEN psoriasis project is an ongoing survey of the quality (i.e., internal validity) and generalizability (i.e., external validity) of the therapeutic studies for psoriasis published in a number of dermatological and medical journals since 1977. This paper aims to describe the range of treatment comparisons, study designs and quality of reporting of randomized clinical trials (RCTs) published up to the end of 2000, and to analyze time trends with quality items.

METHODS

Searching strategy for therapeutic studies This study was based on a survey of published reports of original therapeutic studies of psoriasis, i.e., studies evaluating the intended effect of a treatment for psoriasis. Our focus was on quality assessment of original papers reported in full. Abstracts, letters to the Editor, Congress proceedings, review papers, individual case reports, clinical studies involving less than 5 patients, papers mainly focusing on side-effects of treatment, and studies that dealt only with the articular symptoms of psoriatic arthritis were excluded.

This study aims to describe the range of treatment comparisons, study designs and quality of reporting of randomized clinical trials (RCTs) in psoriasis published in a variety of medical and dermatological journals, and to analyze time trends with quality items. Hand-searching of clinical trials of psoriasis published from 1977 to 2000 in 13 medical or dermatological journals, selected as relevant to a European readership, was performed. A total of 249 trials published in 226 papers were classified as RCTs. Of these, 139 (55.8%) employed a parallel control group design, 107 (43.0%) studies adopted a self-control design and 3 (1.2%) a cross-over design. The median number of patients recruited per study was 40 (range 6–699). Overall, 55 different treatment modalities, including topical, ultraviolet-based, systemic, and other miscellaneous therapies were assessed. Only 31 (12.5%) RCTs were comparative studies of treatment modalities in different therapeutic classes. Most of the studies were short-term with a median study duration of 7 weeks (range 1–104), with only 18 studies (7.2%) lasting for more than four months. A variety of outcome measures including 44 different score systems were employed. According to the conclusions of the authors, 196 (78.7%) studies were judged to provide striking or definite observations in favor of one of the treatments examined. No important variations over time were documented for quality items. Based on our survey we have identified an enormous range of treatments that have been evaluated for psoriasis over the examined period. Most studies were short-term, and only a handful compared treatment options in different therapeutic classes. Since we did not examine all the relevant journals, the number of treatment options may be even greater than we have documented. There is an urgent need to reset the research agenda focusing on long-term comparative RCTs. Editors of major medical and dermatological journals are urged to take a role in improving the quality of RCT reporting.

Randomized Clinical Trials for Psoriasis 1977–2000: The EDEN Survey

Luigi Naldi, Ake Svensson, Thomas Diepgen, Peter Elsner, Jean-Jacques Grob, Pieter-Jan Coenraads, Jan Nico Bouwes Bavinck, Hywel Williams, on behalf of The European Dermato-Epidemiology Network (EDEN)

1Department of Dermatology, University of Geneva – Geneva, Switzerland & Department of Dermatology, Bergamo General Hospital – Bergamo, Italy; 2Department of Dermatology, University of Lund – Malmo, Sweden; 3Department of Social Medicine and Occupational & Environmental Dermatology, University of Heidelberg, Heidelberg, Germany; 4Department of Dermatology, University of Jena – Jena, Germany; 5Service de Dermatologie, University of Marseille – Marseille, France; 6Occupational & Environmental Dermatology Unit, Groningen University Hospital – Groningen, the Netherlands; 7Department of Dermatology, Leiden University Medical Center – Leiden, the Netherlands; 8Center of Evidence-Based Dermatology, University of Nottingham – Nottingham, UK

See related Commentaries on pages iv, v and ix
Papers were classified and assessed, employing a predefined checklist (available upon request from the corresponding author), by a pair of investigators independently.

Selection of journals The following journals were selected on the basis of a consensus among assessors taking into account journal impact factor, content, publication during the time interval examined (restricted to English, French, German, and Italian), and relevance to an European readership: Acta Dermato-Venereologica, Annales de Dermatologie et de Vénéréologie, Archives of Dermatology, British Journal of Dermatology, British Medical Journal, Clinical Experimental Dermatology, Dermatology, Giornale Italiano di Dermatologia e Venerologia, Journal of the American Academy of Dermatology, Journal of Investigative Dermatology, Journal of the American Medical Association, Lancet, New England Journal of Medicine. Information on the impact factor was derived from data provided by the Journal Citation Reports (JCR), annual publication from the Institute for Scientific Information/Thompson Scientific for the year 1990 (referring to the two previous years 1988–9).

Data abstraction The complete volumes of all journals of the years 1977–2000 (including supplements) were hand-searched by a pair of investigators independently. Each investigator searched between two and four journals. Initially, based on our checklist, papers were classified either as therapeutic studies lacking internal control, including studies without any explicit control group and studies with external historical controls, or as studies with internal control, including parallel concurrent control studies, self-controlled studies (e.g., right/left body comparison studies) and crossover studies (Lavers et al, 1983; Bailer et al, 1984; Louis et al, 1984). For all these studies, the year of publication, intervention(s) examined, country of the study, and total number of patients included were reported. Subsequent evaluation was limited to internally controlled studies. The following items were assessed without making any judgment of the appropriateness of the methods applied in any single paper: randomization, entry criteria, rejection log, distribution of pretreatment variables, description of the blinding procedure(s), duration of experimental treatment(s), dropout(s), efficacy parameters, major end point(s), statistical power calculation, appropriateness of the analysis according to time, intention-to-treat analysis, confidence intervals, subgroup analyses, main message of the study, appropriateness of the inference(s) (Chalmers et al, 1981; DerSimonian et al, 1982). Whether the study reported any sponsorship by a pharmaceutical company was also evaluated. Data were entered independently by two investigators into a computer-based questionnaire.

Statistics These original datasets were assessed for agreement between investigators. Subsequently, disagreements were resolved through discussions between the two data abstractors or a third reviewer if needed. The intrarater agreement, assessed by calculating Cohen’s Kappa statistics, ranged from 0.31 (sponsorship) to 0.76 (patient blinding) with most of the other values exceeding 0.60. For the purpose of this study’s quality analysis, only RCTs were considered, i.e., studies explicitly mentioning randomization as a means of allocating treatment. Descriptive statistics were calculated for all the parameters assessed.

RESULTS

Number of studies per paper A total of 422 papers dealing with the treatment of psoriasis were initially identified. Of all the papers, 354 (83.9%) reported one study, while 68 (16.1%) reported several studies with 49 (11.6%) papers reporting two studies, 13 (3.0%) three and 6 (1.4%) four or more, giving a total of 515 studies to be assessed. Of these, 249 (48.3%) studies published in 226 papers were classified as RCTs. In particular, 194 papers reported one RCT as the only study, 14 papers reported a combination of one RCT and one observational study (usually a follow-up study of patients initially recruited in the RCT), 14 papers reported two different RCTs, 3 papers reported 3 different RCTs and one paper reported 4 different RCTs. Details of the examined studies are reported in a Table published with the web version of this paper (http://www.blackwellpublishing.com/products/journals/suppmat/JID/jidl2145/jidl2145s.htm). The proportion of papers presenting at least one RCT increased from 78 out of 202 (38.6%) during the period 1977–88 to 148 out of 220 (67.3%) during the period 1989–2000. Four journals (British Journal of Dermatology, Journal of the American Academy of Dermatology, Dermatology, and Archives of Dermatology) accounted for 314 (74.4%) of all papers. The same journals accounted for 187 (82.7%) of the papers reporting a RCT. Only 8 papers were published in the four general medical journals. A total of 63 papers reporting a RCT came from the United States, 58 from the United Kingdom, 18 from Germany, 17 from Denmark, 16 from the Netherlands, 11 from Sweden, 9 from Finland, 6 from France, and 28 from other countries or represented international collaboration.

Types of study design and patient numbers Of the 249 RCTs identified, 139 (55.8%) had a parallel control group, 107 (43.0%) studies adopted a self-control (e.g., right/left body comparison in the same patient) design and 3 (1.2%) a cross-over design. The median number of patients per RCT was 40 (range 6-699) and only 85 (34.1%) of all the RCTs included more than 50 patients. Papers reporting on more than one study did not include more patients than papers reporting on one single RCT. The median number of patients enrolled per RCT was 40 in papers reporting one study, 29 in papers reporting two studies and 28 in papers reporting three or more studies. The median number of patients was 32 (upper limit 354) for studies published up to 1988 and 42 (upper limit 699) for studies published from 1989 to 2000.

Description of participants A total of 138 RCTs (53.4%) made statements about the patient recruitment process reporting information on entry criteria and exclusions while 111 (44.6%) RCTs either failed to mention any entry criteria or provided poor information. As for the clinical variety of the patients actually recruited, 196 (78.7%) RCTs dealt with psoriasis vulgaris, either alone (134 studies) or in association with other varieties (62 studies), a total of 17 studies concentrated on palmoplantar and other varieties of pustular psoriasis, 15 studies on nail/scalp psoriasis, 2 studies on guttate psoriasis, and 19 studies on unspecified varieties or forms of psoriasis. The disease was usually described by topography and type of lesions while disease duration was considered in 21 (8.4%) RCTs and previous treatments in 67 (26.9%) RCTs.

Heterogeneity of treatments A total of 55 different treatment modalities were evaluated in the RCTs with 28 (50.9%) modalities being assessed in no more than one single RCT. This summary statistic concerning treatment modalities does not include dose variations of the same drug, drug combinations or different modalities of application of the same topical agent. The regimens tested could be classified into topical agents, ultraviolet (UV) radiation regimens, including oral psoralens associated with UVA radiation (PUVA) and UVB radiation, systemic treatments, and other miscellaneous treatment modalities. Within each group several treatment classes could be identified, e.g., the topicals group included a range of treatments such as tar, vitamin D derivatives, keratolytics, and topical steroids. Miscellaneous therapies included plasmapheresis, psychotherapy, dietetic measures, radiotherapy, and acupuncture. The large majority of the studies compared the experimental treatment against placebo, or against a drug within the same therapeutic class, e.g., studies of different steroids. There were only 31 (12.4%) comparative studies of treatments classifiable into two or more different therapeutic classes. These included 17 out of 120 (14.1%) RCTs assessing a topical agent, 7 out of 46 (15.2%) RCTs assessing UV light treatment, 6 out of 75 (8.0%) RCTs assessing a systemic treatment, and one of 8 RCTs assessing treatment modalities not otherwise classifiable. We found only two RCTs contrasting two or more different systemic treatments. They compared cyclosporine with etretinate in 86 and 210 people, respectively (Italian Multicenter Study Group on Cyclosporin in Psoriasis, 1993; Mahle et al, 1995). No single RCT included systemic methotrexate, a popular drug in
many countries for severe psoriasis, in one of the study arms. Six RCTs compared UV radiation therapies with a systemic treatment, namely etretinate or acitretin. The median number of patients enrolled per study was 40 (upper limit 699) for RCTs involving topical therapies, 44 (upper limit 224) for RCTs involving UV light therapies, 42 (upper limit 400) for RCTs involving systemic therapies, and 25 (upper limit 145) for RCTs involving miscellaneous therapies.

**Study duration and outcome measures** Most of the studies were short-term, with a median study duration of 7 weeks (range 1–104) and with only 18 studies (7.2%) lasting for more than four months. A total of 39 studies (15.6%) gave no information on the study duration. The median study duration was 6 weeks (upper limit 104) for RCTs involving topical therapies, 8 weeks (upper limit 30) for studies involving UV light treatment, 10 weeks (upper limit 52) for systemic therapies, and 6 weeks (upper limit 28) for miscellaneous treatment modalities. The median study duration was 7 weeks (upper limit 52 weeks) for the studies published in the period 1977–88 and 8 weeks (upper limits 104) for those published in the period 1989–2000.

A total of 171 (68.7%) studies summarized treatment responses using a score system. Forty-four different score systems were employed which, in variable ways, took into account clinical features such as extent of skin lesions, degree of erythema and scaling. The most popular scoring system was the so-called Psoriasis Area and Severity Index (PASI) adopted by a total of 83 (33.3%) studies. PASI is a scoring system only partly validated (Kirby et al, 2001), made up of arbitrary numerical scores given to the extension of skin lesions, erythema, scaling and infiltration. A total of 99 (39.7%) studies assessed outcome as the attainment of a clinical category, e.g., clearance. Only 9 (36%) studies considered maintenance of remission over time and relapse rates. A total of 19 (7.6%) studies assessed patient’s preference or satisfaction and only one paper evaluated quality of life. No paper assessed such factors as the impact of treatment on accessory care and hospitalization.

**Issues of blinding and dropouts** A total of 166/249 (66.6%) studies were declared as double blind, but of these only 48 gave a satisfactory description of blinding procedures. Because of the treatment modality or peculiar side-effects, PUVA, UVB and retinoids are particularly difficult to blind. A satisfactory description of blinding procedures with these treatments was given by 4/11, 5/10, and 4/17 double-blinded RCTs, respectively. A satisfactory modality to blind outcome assessment was to have a clinical category, e.g., clearance. Only 9 (36%) studies considered maintenance of remission over time and relapse rates. A total of 19 (7.6%) studies assessed patient’s preference or satisfaction and only one paper evaluated quality of life. No paper assessed such factors as the impact of treatment on accessory care and hospitalization.

**Authors’ interpretation of data and sponsorship issues** According to the conclusion of the authors, 196 (78.7%) studies were judged to provide striking or definite observations in favor of one of the treatments examined while 48 (19.3%) reported negative results; in five studies unclear statements about treatment results were provided.

In 138/226 (60.1%) papers reporting a RCT, sponsoring by a pharmaceutical company was mentioned or made obvious by the affiliation of the authors. Fifty-eight (25.7%) papers were categorized as nonsponsored, while in 30 papers (13.3%) sponsorship remained unclear. A total of 108 (78.2%) out of 138 sponsored papers and 41/58 (70.7%) nonsponsored papers presented at least one study reporting results in favor of one of the treatments examined. The median number of patients enrolled was 40 for sponsored and 38 for nonsponsored studies. The median study duration was 7 weeks for sponsored and 8 weeks for nonsponsored studies. No major difference was observed between sponsored and nonsponsored studies for quality items such as entry criteria, blinding and analysis of drop-outs. The proportion of sponsored papers was 57.7% (45 out of 78) up to 1988 and 62.8% (93 out of 148) from 1989 to 2000.

**DISCUSSION**

**General poor quality and clinical usefulness of psoriasis studies** Based on our survey, published studies on the therapy of psoriasis have been characterized, in the period analyzed and limited to the examined journals, by an enormous range of treatments that have been mainly evaluated in small, short-term RCTs with a very limited number of comparative studies and a heterogeneity of outcome measures. Frequent methodological flaws were also documented, including the lack of presentation of entry criteria, inadequate information on blinding procedures, and failure to consider drop-outs in the analysis. A large proportion of RCTs reported positive results. Our analyses confirm and expand the results of previous surveys concerning the quality of RCTs on psoriasis and other inflammatory skin diseases (Bigby et al, 1985; Marks et al, 1989; Petersen and Kristensen, 1992; Williams and Seed, 1993; Ashcroft et al, 1999).

**Potential limitations of this study** There are aspects of our survey that may limit the generalizability of the results. We dealt only with reports published in full and restricted the survey to 13 journals. The journals were chosen based on a consensus which took into account, among the others, impact factor, likely yield of RCTs based on the work of the Cochrane Skin Group specialized trials register, and relevance to a European readership. Even if our study missed a small subset of published psoriasis trials, it is not expected that important RCTs were overlooked (Moher et al, 2000). Moreover, no substantial variations over time were observed to suggest a modifying trend. Most of our main results concerned objective aspects of RCTs such as type of treatment arms, study size and duration. Since we did not examine all the relevant journals, the number of treatment options may be even greater than we have documented. Our analysis is not a systematic review trying to summarize treatment efficacy but rather a broad overview of methods used in psoriasis trials with emphasis on study design and quality of reporting. As such it differs from previous studies summarizing the evidence for a given variety of psoriasis, e.g., severe psoriasis, or a single drug, e.g., calcipotriol (Ashcroft et al, 2000; Griffiths et al, 2000).

**How is the agenda of psoriasis trials driven?** It appears that the development of new treatments for psoriasis follows a rather repetitive pattern where important questions for clinicians and their patients like the comparative value of different treatment options and the long-term impact of the treatment on the disease are scarcely considered. A large proportion of RCTs (59%) are sponsored by pharmaceutical companies. New drugs are being developed that include those that target memory-effector T cells, e.g., efalizumab (Ellis and Krueger, 2000), and cytokine blocking agents, e.g., etanercept (Mcarte et al, 2000).

**What is needed to address the problem** Based on our survey a number of suggestions can be made for future research.

1. Placebo use. There is a need to establish criteria for the use of placebo and comparative treatments in this area. They should be developed with the active and informed participation of the public and should be considered by review boards and regulatory agencies. In principle, placebo use should be restricted to the early phase of development of a new treatment. Subsequently, comparative studies are needed. The comparator treatment should reflect the usual clinical practice. Factorial design can be considered for treatment combinations.

2. Study duration. Long-term results are simply not predictable from short-term studies. Long-term “pragmatic” RCTs conducted under conditions close to clinical practice are needed.
Consensus should be reached on study duration to document the effects of suppressive and remitting therapies. Several treatment cycles can be assessed as a way of maintaining remission.

3 Outcome measures. An array of clinical activity scores have been developed, the most popular one being the PASI. There is no documented evidence that such indices are a reliable surrogate for outcomes that matter to the patient, like disease suppression and duration of remission, patient satisfaction and autonomy and disease-related quality of life. In the long term, the duration of remission, the way the disease is controlled and the treatment side-effects are vitally important, and simple outcome measures applicable in all patients seem to be preferable (Wright, 2000). These may include the number of patients in remission, the number of hospital admissions or ambulatory consultations, or major disease flare-ups. Clearly, remission or recurrence are not as frequent as less dramatic variations in disease activity measured by highly sensitive expanded complex clinical scales (Al-Suwaidan and Feldman, 2000). This in turn affects the sample size calculation.

4 Handling of drop-outs. Drop-outs merit special attention since they may strongly reflect dissatisfaction with treatment (Schiffner et al, 2001). They cannot be simply ignored. Both intention-to-treat and per-protocol analyses should be presented.

5 Systematic reviews. The large number of RCTs in this area may point to systematic reviews as a way to clarify uncertainty where there are several small conflicting studies. However, it is not expected that systematic reviews alone could overcome the limitations we have pointed to concerning the study duration and the lack of comparative studies (Griffiths et al, 2000). Quite paradoxically, if systematic reviews concentrate on one of the many treatments that have been tested for psoriasis, they may risk giving undue emphasis to such treatments. Systematic reviews alone are not expected to fill the gap, and primary research and high quality relevant RCTs are urgently needed.

CONCLUSION

In conclusion, we have identified an enormous range of treatments that have been evaluated for psoriasis during the study period. Most studies were short-term and only a few compared treatment options in different therapeutic classes. A heterogeneity of outcome measures were also employed. Faced with this situation, it is hardly surprising that marked variations between centers and countries have been documented in the management of psoriasis with many treatment decisions being based on such issues as local traditions, individual preference and effective marketing.

Our main focus was on reporting not on the ways the studies were actually conducted. However, previous studies have documented that poor quality of reporting is strongly related to poor trial quality (Juni et al, 2001). The four most important recommendations when researchers and editors plan and publish future psoriasis trials are (i) to select for comparison an active intervention restricting the use of placebo; (ii) to consider longer term studies, e.g., at least six month duration; (iii) to adopt clinically relevant outcomes such as duration of remission, major disease flare-ups, patient satisfaction; and (iv) to consider drop-outs in the analysis by presenting both intention-to-treat and per-protocol analyses.

REFERENCES

Wright JG: Evaluating the outcome of treatment: Shouldn’t we be asking patients if they are better? J Clin Epidemiol 53:549–553, 2000