The Reporting of Observational Research Studies in Dermatology Journals

A Literature-Based Study

Sinéad Langan, MD, MSc, PhD; Jochen Schmitt, MD, MPH; Pieter-Jan Coenraads, MD, PhD, MPH; Åke Svensson, MD, PhD; Erik von Elm, MD; Hywel Williams, MSc, PhD, FRCP; for the European Dermato-Epidemiology Network (EDEN)

Objective: To assess the quality of reporting in observational studies in dermatology.

Data Sources: Five dermatology journals—the Archives of Dermatology, the British Journal of Dermatology, the Journal of the American Academy of Dermatology, the Journal of Investigative Dermatology, and Acta Dermato-Venereologica.

Study Selection: Cohort, case-control, and cross-sectional studies published as original articles during the period January 2005 through December 2007. Studies were identified with a literature search of PubMed combining the journal title and the term epidemiological studies (free text) and by hand searching all of the issues of each journal to identify relevant articles.

Data Extraction: All articles were extracted by 2 reviewers independently using standardized checklists based on the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations.

Data Synthesis: The number and proportion of reported STROBE items were analyzed for each article. The proportion of studies with good reporting for each item was also assessed.

Results: A total of 138 articles were included and analyzed. Reporting quality was very mixed. Key areas that were infrequently reported included sample size calculations (n=10 [7%]), missing data (n=8 [6%]), losses to follow-up (n=17 [12%]), and statistical methods (n=19 [14%]). Only 13 studies (9%) explained the role of funders in the research. The quality of reporting was similar across study designs for “critical” questions with the exception of reporting of participant details, which was better reported in cohort studies (96%) compared with cross-sectional (80%) and case-control (70%) studies.

Conclusions: It is difficult to judge the quality of dermatological research unless it is reported well. This study has identified a clear need to improve the quality of reporting of observational studies in dermatology. We recommend that dermatology journals adopt the STROBE criteria.

Arch Dermatol. 2010;146(5):534-550

INTERPRETATION OF DATA FROM OBSERVATIONAL STUDIES IS OFTEN LIMITED BY POOR QUALITY OF REPORTING. POOR REPORTING LIMITS THE ASSESSMENT OF A STUDY'S STRENGTHS AND WEAKNESSES AND GENERALIZABILITY. IT ALSO LIMITS THE USE OF OBSERVATIONAL DATA FOR SECONDARY ANALYSES. THE CLINICAL AND SCIENTIFIC UTILITY OF RESEARCH DATA MAY BE LOST IN POORLY REPORTED STUDIES.

THE STRENGTHENING THE REPORTING OF OBSERVATIONAL STUDIES IN EPIDEMIOLOGY (STROBE) CRITERIA HAVE RECENTLY BEEN DEVELOPED TO TRY AND IMPROVE THE QUALITY OF REPORTING OF OBSERVATIONAL STUDIES IN MEDICAL RESEARCH. THE STROBE CHECKLIST AIMS TO ASSESS AND IMPROVE THE REPORTING OF EPIDEMIOLOGICAL STUDIES, BUT IT IS NOT A MEASUREMENT FOR THE QUALITY OF THE RESEARCH ITSELF. POOR REPORTING MAY ALSO CONSTITUTE AN ETHICAL PROBLEM IF THE NECESSARY INFORMATION IS COLLECTED BUT IS NOT DESCRIBED IN RESEARCH ARTICLES, THUS REDUCING ITS USEFULNESS. SIMILAR INITIATIVES...
tives for improving the quality of reporting of particular study types have already shown benefits in terms of the reporting of randomized controlled trials using the Consolidated Standards for Reporting Trials (CONSORT) statement. From our collective experience in performing systematic reviews within the field of dermatology, we hypothesized that the quality of reporting of observational studies in dermatology was “hit and miss,” with many published articles failing to mention key aspects that would allow the reader to judge the validity of the study findings and conclusions. We set out to systematically assess the quality of reporting of observational studies in dermatology using the STROBE checklist as a guideline and to highlight specific areas that could be improved.

METHODS

A study protocol was devised prior to the study, and all methods were defined and piloted a priori.

ELIGIBLE JOURNALS

We selected the 4 journals with the highest impact factor (IF) in the category “Dermatology” in the Journal Citation Report 2007 that have a section for the reporting of epidemiological studies, namely, the Archives of Dermatology (IF, 2.84), the British Journal of Dermatology (IF, 3.50), the Journal of the American Academy of Dermatology (IF, 2.90), and the Journal of Investigative Dermatology (IF, 4.83). Acta Dermato-Venereologica (IF, 1.93) was also included because this journal publishes a high proportion of epidemiological studies. The journal articles were reviewed over a 3-year period, from January 2005 through December 2007.

ELIGIBLE STUDIES

We included epidemiological studies that corresponded to a relevant STROBE checklist, namely, cohort, case-control, and cross-sectional studies published during the study period. Studies that were labeled as a randomized controlled trial (RCT) or laboratory research that were in fact observational epidemiological studies were included in this study under the appropriate category.

EXCLUSION CRITERIA

Epidemiological studies other than cohort, case-control, or cross-sectional studies such as ecological studies were excluded. Studies in abstract format only were also excluded. Genetic epidemiological studies were not included because this study subtype has its own reporting guideline (Strengthening the Reporting of Genetic Associations [STREGA]).

STUDY SELECTION

Studies were identified with a literature search of PubMed combining the journal title and the term epidemiological studies (free text) and by hand searching all of the issues of each journal to identify relevant articles. This process was carried out independently by 2 reviewers (S.L. and J.S.), and disagreements were resolved by discussion and arbitration by a third person (H.W.) if needed. Articles were then numbered consecutively and randomly allocated to each reviewer using a random numbers table.

DATA ABSTRACTION

To determine whether further clarification was needed to define the questions when scoring items from the STROBE checklist, 3 articles representing each main study design underwent a pilot data abstraction by all reviewers (S.L., J.S., P.-J.C., and Å.S.) and by the European Dermato-Epidemiology Network (EDEN) steering group (10 members including S.L., J.S., P.-J.C., and Å.S.). Disagreements were resolved by discussion with all reviewers. Following the pilot phase, all articles were extracted by 2 reviewers independently. Disagreements were resolved between pairs (S.L. and J.S.; Å.S. and P.-J.C.), and an independent arbitrator (H.W.) was involved for persisting differences.

Relevant data from included articles was summarized in an Excel spreadsheet (Microsoft Corp, Redmond, Washington) containing information on the number and proportion of the items in the checklist that were reported against the STROBE checklist.

REPORTING OF QUALITY OF STUDY REPORTING

The STROBE checklists of 22 items was operationalized into a series of questions for each study design, which can be answered “yes,” “partly,” “no,” “unclear,” or “not applicable.” One author (E.v.E.) was also involved in the development of the STROBE checklist. This guaranteed the consistency of this operationalization and of our research in general with the aims of the STROBE initiative. The checklists were piloted by all investigators on the 3 selected study types as previously described. Each investigator independently submitted their analysis to their paired author. A panel discussion by e-mail followed to resolve the differences in analyses and to obtain a consensus on how the responses are defined by individual investigators.

A glossary of rules was developed throughout the study, which will help inform the STROBE initiative and other researchers. Briefly, a number of key decisions were made to translate the STROBE reporting items into data that could be assessed reliably.

ANALYSIS

The number and proportion of reported items (“yes” responses) and not reported items (all responses except “yes” or “not applicable”) were analyzed for each study. We then examined the median and range of reported items by study design. The proportion of studies with good reporting for each item was also assessed. Items that were not applicable were excluded from the analysis. A sensitivity analysis was carried out, in which “partly” responses were analyzed as “yes” responses to assess the effect on study conclusions.

SELECTION OF THE MOST CRITICAL STROBE ITEMS

To inform future studies, we believed it would be useful to identify a priori which questions within the STROBE checklist should be critical to complete when reporting observational studies. To achieve this, each author independently identified what they perceived to be 10 essential questions, and responses were then pooled to determine a critical list of items (Table 1). An item was considered essential if at least 4 of 5 reviewers (S.L., J.S., P.-J.C., Å.S., and H.W.) agreed on the item for inclusion, and these are given in Table 2. Of the 22 items in the STROBE checklist, 4 were considered essential by the panel of authors prior to completion of the study. The
<table>
<thead>
<tr>
<th>Original STROBE Item No.</th>
<th>STROBE Description</th>
<th>Title and Abstract</th>
<th>Rules Developed During the Study</th>
<th>No. of Reviewers Who Considered Item Critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Is the design described adequately in the title or abstract? Does the abstract provide an informative summary of what was done and found?</td>
<td></td>
<td>If the study design was not specifically stated, this was recorded as not being complete. For incorrectly labeled studies, the appropriate category of study design was identified and used for further analyses.</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Is the scientific background and rationale for the investigation reported? Are the objectives reported?</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Are any prespecified hypotheses reported? Are the objectives reported?</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Are the key elements (ie, retrospective/prospective, cohort/cross-sectional) of the study design presented?</td>
<td>Methods</td>
<td>Studies that did not report the setting or locations but referred readers to a previous publication were considered as inconsistent with complete reporting. If the dates of recruitment were recorded anywhere in the article (not necessarily in the “Methods” section), the corresponding item was rated as complete.</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Are the settings reported? Are the locations reported? Are relevant dates including periods of recruitment reported?a Are relevant dates including periods of exposure reported? Are relevant dates including periods of follow-up reported? Are relevant dates including periods of data collection reported?</td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>Are the eligibility criteria for participants described? Are the sources of participants described? Are the methods of selection described? Are the methods of follow-up described? (only used for n=101) If it is a matched study, are the matching criteria and the numbers of exposed and unexposed described? (only used for n=30)</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>Are all outcomes described if applicable? Are all exposures described if applicable? Are all predictors described if applicable? Are potential confounders described? Are all effect modifiers described? Are diagnostic criteria described if applicable?</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>Are the sources of data and details of methods of measurement given for each variable of interest? If there is more than 1 group, are the measurement methods comparable?</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>Was there any effort to address potential sources of bias?</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>Did they describe how the study size was determined?</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>Did they describe how quantitative variables were handled in the analysis? Did they describe which groupings were chosen for quantitative variables? Did they describe why quantitative groups were chosen?</td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>Did they describe all statistical methods including those to deal with confounding? Did they describe methods to examine subgroups and interactions? Did they explain how missing data was addressed? Did they explain if applicable how losses to follow-up were addressed? Did they describe any sensitivity analysis?</td>
<td></td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

(continued)
The search strategy identified 291 studies, of which 138 were relevant for this study (a list is provided in an online repository available at: http://eden.dermis.net/content/e02eden/e02projects/e88/index_ger.html). The figure shows the flow diagram of included studies: 53 articles (38%) were cohort studies, 44 (32%) were case-control studies, and 41 (30%) were cross-sectional studies.

PROPORTION OF REPORTED ITEMS PER ARTICLE

The median (range) number of reported items per article was comparable across the different study designs. For cohort studies, the median number of reported items per ar-
The median was 12 (7-17) for cross-sectional designs and 12 (4-18) for case-control studies.

**ITEMS USUALLY REPORTED**

Of the 138 studies, 137 (99%) reported key results in relation to study objectives (Table 2). Similarly, 134 studies (97%) reported the scientific background and rationale for carrying out the study. Most studies (n=129 [93%]) provided a reasonable summary of the research in the abstract. The majority of studies reported the study objectives (n=123 [89%]), design (n=120 [87%]), and outcomes (n=122 [88%]).

**ITEMS NOT USUALLY REPORTED**

Sample size calculations were reported by only 10 (7%) of included studies. Other areas that were very poorly described relate to the issues of management of missing data and sample size calculations.

---

**Table 2. Checklist of 22 Questions With Overall Scoring for Included Studies and Percent Complete**

<table>
<thead>
<tr>
<th>Original Item No.</th>
<th>STROBE Description</th>
<th>Yes</th>
<th>Partly</th>
<th>No</th>
<th>Unclear</th>
<th>Not Applicable</th>
<th>Individual Item Adequate</th>
<th>Overall Item Adequate</th>
<th>Results of Sensitivity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title and Abstract</td>
<td>120 (87) 13 (9) 3 (2) 2 (1)</td>
<td>120 (87) 135 (98) 137 (99)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Methods</td>
<td>134 (97) 4 (3)</td>
<td>134 (97) 138 (100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Are the objectives reported?</td>
<td>12 (9) 4 (3) 121 (88) 1 (1)</td>
<td>12 (9) 123 (89) 133 (96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Are the key elements (ie, retrospective/ prospective, cohort/cross-sectional) of the study design presented?</td>
<td>123 (89) 10 (7) 4 (3) 1 (1)</td>
<td>123 (89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Are the settings reported?</td>
<td>112 (81) 6 (4) 13 (9) 4 (3) 3 (2)</td>
<td>112 (83) 108 (78) 122 (88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Are the eligibility criteria for participants described?</td>
<td>116 (84) 13 (9) 6 (4) 2 (1) 1 (1)</td>
<td>116 (85) 115 (83) 129 (93)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Are all outcomes described if applicable?</td>
<td>122 (88) 7 (5) 1 (1) 4 (3) 4 (3)</td>
<td>122 (91) 73 (53) 114 (83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Are the sources of data and details of methods of measurement given for each variable of interest?</td>
<td>105 (76) 23 (17) 9 (7) 1 (1)</td>
<td>105 (77) 118 (86) 132 (96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Was there any effort to address potential sources of bias?</td>
<td>43 (31) 1 (1)</td>
<td>2 (1) 5 (4)</td>
<td>87 (63) 43 (84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Did they describe how the study size was determined?</td>
<td>12 (7) 125 (91)</td>
<td>3 (2) 10 (7) 10 (7) 10 (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Did they describe how quantitative variables were handled in the analysis?</td>
<td>43 (31) 13 (9) 52 (38) 4 (3)</td>
<td>43 (38) 40 (36) 50 (45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Did they describe all statistical methods including those to deal with confounding?</td>
<td>37 (27) 6 (4) 51 (37) 4 (3) 40 (29) 37 (38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Did they describe which groupings were chosen for quantitative variables?</td>
<td>37 (27) 6 (4) 51 (37) 4 (3) 40 (29) 37 (38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Did they describe why quantitative variables were chosen?</td>
<td>11 (8) 2 (1) 74 (54) 5 (4) 46 (33) 11 (12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Did they describe methods to examine subgroups and interactions?</td>
<td>20 (14) 5 (4) 67 (49) 1 (1) 44 (32) 20 (22)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Did they explain how missing data was addressed?</td>
<td>8 (6) 2 (1) 124 (90) 1 (1) 3 (2) 8 (6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Did they explain if applicable how losses to follow-up were addressed?</td>
<td>17 (12) 7 (5) 74 (54) 1 (1) 39 (28) 17 (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Did they describe any sensitivity analysis?</td>
<td>8 (6) 125 (91) 1 (1) 4 (3) 8 (6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
and losses to follow-up (n=17 [12%]). In addition, infrequently addressed areas included the number of individuals at each stage of the study (n=54 [39%]) and the statistical methods (n=19 [14%]). Thirteen studies (9%) explained the role of funders in the research.

REPORTING OF CRITICAL ITEMS

Of those questions considered essential (Table 2), questions 1 (98%) and 6 (83%) were frequently reported, whereas questions 7 (53%) and 12 (14%) were infrequently described. Examining the critical items by study design, item 1 was adequately completed by at least 93% of studies in all categories. Item 6 was reported in 96% of cohort studies compared with 80% of cross-sectional studies and 70% of case-control studies. Completion rates were similar for items 7 (52%-54%) and 12 (10%-18%) across all study designs.

SENSITIVITY ANALYSIS

Recoding partly responses as yes responses altered the proportion of completely reported items as given in Table 2. Specifically, in relation to question 7 (study variables), this increased completion rates from 53% (n=73) to 83% (n=114). Similarly for question 14 (de-
talents of study participants), completion rate increased from 43% (n=60) to 67% (n=92). Significant increases were also observed for the questions 16 (reporting of crude and confounder adjusted estimates), 19 (study limitations), 20 (cautious interpretation), and 21 (generalizability of the findings).

**SUMMARY OF FINDINGS**

This study has highlighted important areas requiring improvement in the reporting of observational studies in dermatology. On the basis of these findings, we believe it would be useful for authors if dermatology journals adopted the use of the STROBE criteria to guide authors. Authors of epidemiological studies might also start using the guidance regardless of which dermatology journals they submit their article to. This study has defined a number of key rules that will be useful for researchers. These include the conclusion that settings and locations should always be defined even in the setting of referring readers to previous publications. Another important conclusion was that for studies based on disease registries or databases, a number of the checklist items are not applicable, for example, the dates of recruitment, numbers eligible at each stage of the study, reasons for nonparticipation, or flow diagrams.

**STRENGTHS AND LIMITATIONS OF THIS STUDY**

This is a novel study that has systematically highlighted important areas requiring improvement in the reporting of observational studies in dermatology. The results have direct relevance for readers and authors of epidemiological studies. The use of pilot methods to devise the study checklists, independent data extraction, and arbitration has helped improve the methodological rigor of our study. One of the key limitations is that it was not possible to formally assess interrater agreement because discussion between authors and the creation of rules were required to resolve these issues as the study identified new problems as it progressed. A degree of subjectivity is reflected by the findings of the sensitivity analysis.

**FINDINGS IN CONTEXT**

The quality of reporting of observational studies has been previously assessed in other fields of medical research but not, to our knowledge, in dermatology. We only identified studies assessing the quality of reporting of observational studies using the STROBE criteria in ophthalmology and sexual health. Both specialties have identified similar problems with incomplete reporting, with particularly poor reporting of management of missing data and confounding.

Similar findings have also been shown in the dermatology literature in relation to the reporting of RCTs and the CONSORT guidelines. In 1985, Bigby et al reported (n=61 trials) similar findings in relation to reporting of RCTs, with only 3% of studies reporting power and 76% addressing losses to follow-up. The situation for dermatology RCTs appears to have improved following the introduction of CONSORT guidelines.

**CLINICAL AND RESEARCH IMPLICATIONS**

The quality of reporting of observational studies has been previously assessed in other fields of medical research but not, to our knowledge, in dermatology. We only identified studies assessing the quality of reporting of observational studies using the STROBE criteria in ophthalmology and sexual health. Both specialties have identified similar problems with incomplete reporting, with particularly poor reporting of management of missing data and confounding.
Bleach Baths to Reduce Severity of Atopic Dermatitis Colonized by Staphylococcus

Fiona E. Craig, MBBS, MRCP(UK); Emma V. Smith, MBChB, MRCP; Hywel C. Williams, MSc, PhD, FRCP; Departments of Dermatology, Aberdeen Royal Infirmary, Aberdeen, Scotland (Dr Craig), and Singleton Hospital, Swansea, Wales (Dr Smith); and Centre of Evidence-Based Dermatology, University of Nottingham, King’s Meadow Campus, Nottingham, England (Dr Williams)

**Commentary on:** Treatment of Staphylococcus aureus colonization in atopic dermatitis decreases disease severity

Huang JT, Abrams M, Tiougan B, et al


**Question:** Does suppression of Staphylococcus aureus growth with sodium hypochlorite (bleach) baths and intranasal mupirocin improve atopic dermatitis (AD) severity?

**Design:** A 3-month, investigator-blinded, placebo-controlled, randomized controlled trial (RCT).

**Setting:** Pediatric dermatology clinic at the Children’s Memorial Hospital, Chicago, Illinois.

**Participants:** Thirty-one participants aged 6 months to 17 years with moderate to severe AD and clinical signs of bacterial infection were randomized. Fifteen were assigned to receive active treatment and 16 to receive placebo. At baseline, the treatment group had higher mean Eczema

---


