Improving reporting for observational studies: STROBE statement

Technical meeting on the reporting of human studies submitted for the scientific substantiation of health claims
EFSA Parma – 20 November 2013

Erik von Elm, MD MSc FMH
Institute for Social and Preventive Medicine (IUMSP)
Cochrane Switzerland
University Hospital (CHUV)
Lausanne / Switzerland
Outline

- Need for better reporting of observational research
- STROBE statement
- Case study: a health claim
Outline

 Need for better reporting of observational research
 STROBE statement
 Case study: a health claim
An early call for complete reporting

“A basic principle can be set up that … it is at least as important to describe the techniques employed and the conditions in which the experiment was conducted, as to give the detailed statistical analysis of results.”

Anectodal evidence of poor reporting

“Our readers would be amazed to learn how often we have to remind authors to simply mention where and when their study was conducted.”

Alfredo Morabia, Editor Preventive Medicine
Evidence of poor reporting

- Most empirical evidence on reporting is from randomised trials

- But similar concerns apply to other types of studies:
  - Diagnostic accuracy studies
  - Observational studies
    - case-control / cohort / cross-sectional studies
    - studies based on routine databases
  - Prognostic studies
  - Qualitative studies
  - Systematic reviews
Reporting of case-control studies

Lee *Br J Psychiatry* 2007

- “The reporting of methods in the 408 identified papers was generally poor, with basic information about recruitment of participants often absent …”

- “Poor reporting of recruitment strategies threatens the validity of reported results and reduces the generalisability of studies.”
Survey of observational studies

Pocock *BMJ* 2004

- Examined 73 published in Jan 2001 in general medical & specialist journals, mostly case-control or cohort studies
- Rationale behind choice of confounders usually unclear
- Extent of adjustment varied greatly
- Many exposures, outcomes, subgroups in same study
  - Multiple statistical tests of hypotheses
  - High probability of spurious findings for associations
  - Risk of overinterpretation
Clustering of p-values around 0.05

Distribution of P values for first primary result in each article and corresponding absolute values of standardised normal deviates z (two sided P=0.05, 0.01, 0.001, and 0.0001 correspond to z=1.96, 2.58, 3.29, and 3.89, respectively)

Pocock *BMJ* 2004
Credibility of epidemiology is at stake

“The credibility of risk factor epidemiology can withstand only a limited number of false alarms”

Alvan Feinstein 1981
The epidemiologists

Have they got scares for you!

John Brignell
Eat to beat the big C

There may not be a cure for cancer yet, but research shows how even small changes in diet can dramatically reduce your risk of getting it.

**BREAST**
Three glasses of milk a day can cut the risk by 50 per cent

**OVARIAN**
Five carrots a week can cut the risk by 50 per cent

**COLON**
Eating fish twice a week can cut the risk by 50 per cent

**STOMACH**
A few radishes a week can cut the risk by 35 per cent

**LUNG**
Eating tomato ketchup every day can cut the risk by 25 per cent

**LIP**
Wearing lipstick can cut the risk by 50 per cent

**MOUTH**
Six sweet potatoes a week can reverse cancer of the mouth

**SKIN**
Lemon tea can cut the risk by 70 per cent
The scandal of poor epidemiological research

Reporting guidelines are needed for observational epidemiology

- Compared to CONSORT additional complexity with reporting guideline for observational studies
- Several main types of observational studies: cohort, case-control, cross-sectional
- Even well-conducted observational studies might still be misleading if important confounders are missed, not measured precisely, not known...
- In articles, additional importance of interpretation & discussion because of many choices made (e.g. for adjustment for confounding)

von Elm Egger BMJ 2004
Impact of adjustment for SES

Not adjusted for socioeconomic status
Pfeffer et al 1978
Hernandez Avila et al 1990
Mann et al 1994
Heckbert et al 1997
Grodstein et al 2000
Varas-Lorenzo et al 2000
Combined

Adjusted for socioeconomic status
Rosenberg et al 1993
Sidney et al 1997
Sourander et al 1998
Combined

Meta-analysis of cohort studies and case-control studies of hormone replacement therapy and coronary heart disease. There is little evidence for a protective effect when analyses are adjusted for, in contrast to studies not adjusted for, socioeconomic status. Adapted from Humphrey et al, reference 7 (Ann Int Med 2002)
Consequences of poor reporting

- Reliability of individual studies cannot be assessed
  - If methods not described in detail, weaknesses may not be apparent

- A body of evidence cannot be used for further decision making e.g. policy regarding a health claim

- Consequences for
  - Other researchers
  - Professional users: clinicians, policy makers, regulators
  - Lay users: patients, consumers
Outline

- Need for better reporting of observational research
- STROBE statement
- Case study: a health claim
Reporting guidelines

- Established by international collaborative groups incl. researchers and editors

- RG specify a minimum set of items required for a clear and transparent account of **what was done** and **what was found** in a study

- Usually checklist, flow diagram, explicit text

- They focus on issues that might introduce bias into health research

- Should be based on evidence if available. If not, consensus opinion.
Knowledge translation

Practice Policy -> Research -> Publication -> Utilization -> Dissemination

Gap between what is done and what is reported
STROBE Statement

- Collaborative effort of working group since 2004
- Checklist of 22 essential items that should be reported for a cohort, case-control / cross-sectional study
- Published 2007 in several journals (open access)
- Translations available
- Comprehensive explanatory paper (E&E) with examples of good reporting
- Several extensions: STREGA, STROBE-ME, RECORD

www.strobe-statement.org
<table>
<thead>
<tr>
<th>Item number</th>
<th>Recommendation</th>
</tr>
</thead>
</table>
| **TITLE and ABSTRACT** | (a) Indicate the study’s design with a commonly used term in the title or the abstract  
(b) Provide in the abstract an informative and balanced summary of what was done and what was found |
| **INTRODUCTION** | Explain the scientific background and rationale for the investigation being reported |
| **METHODS** | |
| Study design | Present key elements of study design early in the paper |
| Setting | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection |
| Participants | (a) **Cohort study**—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up  
Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls  
Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants  
(b) **Cohort study**—For matched studies, give matching criteria and number of exposed and unexposed  
Case-control study—For matched studies, give matching criteria and the number of controls per case |
| Variables | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable |
| Data sources/measurement | For each variable of interest, give sources of data and details of methods of assessment (measurement).  
Describe comparability of assessment methods if there is more than one group |
| Bias | Describe any efforts to address potential sources of bias |
| Study size | Explain how the study size was arrived at |
| Quantitative variables | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why |
| Statistical methods | (a) Describe all statistical methods, including those used to control for confounding  
(b) Describe any methods used to examine subgroups and interactions  
(c) Explain how missing data were addressed  
(d) **Cohort study**—If applicable, explain how loss to follow-up was addressed  
Case-control study—If applicable, explain how matching of cases and controls was addressed  
Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy  
(e) Describe any sensitivity analyses |
## RESULTS

| Participants | 13* | (a) Report the numbers of individuals at each stage of the study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed  
(b) Give reasons for non-participation at each stage  
(c) Consider use of a flow diagram  
| Descriptive data | 14* | (a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders  
(b) Indicate the number of participants with missing data for each variable of interest  
(c) Cohort study—Summarise follow-up time (e.g., average and total amount)  
| Outcome data | 15* | Cohort study—Report numbers of outcome events or summary measures over time  
Case-control study—Report numbers in each exposure category, or summary measures of exposure  
Cross-sectional study—Report numbers of outcome events or summary measures  
| Main results | 16 | (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included  
(b) Report category boundaries when continuous variables were categorized  
(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period  
| Other analyses | 17 | Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses  

## DISCUSSION

| Key results | 18 | Summarise key results with reference to study objectives  
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias  
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence  
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results  

## OTHER INFORMATION

| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based  

---

*Give such information separately for cases and controls in case-control studies, and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of *PloS Medicine* at http://www.plosmedicine.org/, *Annals of Internal Medicine* at http://www.annals.org/, and *Epidemiology* at http://www.epidem.com/). Separate versions of the checklist for cohort, case-control, and cross-sectional studies are available on the STROBE Web site at http://www.strobe-statement.org/.

doi:10.1371/journal.pmed.0040297.t001
Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration

Jan P. Vandenbroucke¹, Erik von Elm²-³, Douglas G. Altman⁴, Peter C. Gøtzsche⁵, Cynthia D. Mulrow⁶, Stuart J. Pocock⁷, Charles Poole⁸, James J. Schlesselman⁹, Matthias Egger²-¹⁰* for the STROBE Initiative

1 Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, 2 Institute of Social & Preventive Medicine (ISPM), University of Bern, Bern, Switzerland, 3 Department of Medical Biometry and Medical Informatics, University Medical Centre, Freiburg, Germany, 4 Cancer Research UK/NHS Centre for Statistics in Medicine, Oxford, United Kingdom, 5 Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark, 6 University of Texas Health Science Center, San Antonio, United States of America, 7 Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, United Kingdom, 8 Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, United States of America, 9 Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, and University of Pittsburgh Cancer Institute, Pittsburgh, United States of America, 10 Department of Social Medicine, University of Bristol, Bristol, United Kingdom

- Examples of good reporting
- Explanatory text
- Key concepts in text boxes (definitions / study conduct)
Box 6. Missing data: problems and possible solutions

A common approach to dealing with missing data is to restrict analyses to individuals with complete data on all variables required for a particular analysis. Although such ‘complete-case’ analyses are unbiased in many circumstances, they can be biased and are always inefficient [108]. Bias arises if individuals with missing data are not typical of the whole sample. Inefficiency arises because of the reduced sample size for analysis.

Using the last observation carried forward for repeated measures can distort trends over time if persons who experience a foreshadowing of the outcome selectively drop out [109]. Inserting a missing category indicator for a confounder may increase residual confounding [107]. Imputation, in which each missing value is replaced with an assumed or estimated value, may lead to attenuation or exaggeration of the association of interest, and without the use of sophisticated methods...
Example of good reporting: item 8

8 Data Sources/Measurement: For Each Variable of Interest Give Sources of Data and Details of Methods of Assessment (Measurement). Describe Comparability of Assessment Methods If There is More Than One Group

Example 1

“Total caffeine intake was calculated primarily using US Department of Agriculture food composition sources. In these calculations, it was assumed that the content of caffeine was 137 mg per cup of coffee, 47 mg per cup of tea, 46 mg per can or bottle of cola beverage, and 7 mg per serving of chocolate candy. This method of measuring (caffeine) intake was shown to be valid in both the NHS I cohort and a similar cohort study of male health professionals (...). Self-reported diagnosis of hypertension was found to be reliable in the NHS I cohort.”

---

Institut universitaire de médecine sociale et préventive, Lausanne
Has it improved reporting?

- We don’t know yet
  - More evidence available for RCTs
  - Difficult to identify specific impact of a guideline

- Some evidence from experimental studies using RGs during peer review (Cobo BMJ 2011)

- Any effect depends on endorsement / enforcement of journals

- Empirical studies keep identifying deficiencies in reporting
Outline

- Need for better reporting of observational research
- STROBE statement
- Case study: a health claim
Health claim from an EFSA application

„...supports the development of healthy and strong bone in children“

Target population: „infants & young children (<3 yrs.)“

ESFA Journal 2013, 11(7) 3331
A Randomized Controlled Study of Effects of Dietary Magnesium Oxide Supplementation on Bone Mineral Content in Healthy Girls

Thomas O. Carpenter, Maria C. DeLucia, Jane Hongyuan Zhang, Gina Bejnerowicz, Lisa Tartamella, James Dziura, Kitt Falk Petersen, Douglas Befroy, and Dorothy Cohen

Context: The role of magnesium (Mg) as a determinant of bone mass has not been extensively explored. Limited studies suggest that dietary Mg intake and bone mineral density are correlated in adults, but no data from interventional studies in children and adolescents are available.

Objective: We sought to determine whether Mg supplementation in periadolescent girls enhances accrual of bone mass.

Design: We carried out a prospective, placebo-controlled, randomized, one-year double-blind trial of Mg supplementation.

Setting: The study was conducted in the Clinical Research Centers at Yale University School of Medicine.

Patients or Other Participants: Healthy 8- to 14-yr-old Caucasian girls were recruited from community pediatricians' offices. Dietary diaries from over 120 volunteers were analyzed, and those with dietary Mg intake of less than 220 mg/d were invited to participate in the intervention.

Intervention: Magnesium (300 mg elemental Mg per day in two divided doses) or placebo was given orally for 12 months.

Main Outcome Measure: The primary outcome measure was interval change in bone mineral content (BMC) of the total hip, femoral neck, Ward's area, and lumbar spine (L1–L4) after 12 months of Mg supplementation.

Results: Significantly increased accrual (P = 0.05) in integrated hip BMC occurred in the Mg-supplemented vs. placebo group. Trends for a positive Mg effect were evident in the pre- and early puberty and in mid-late puberty. Lumbar spinal BMC accrual was slightly (but not significantly) greater in the Mg-treated group. Compliance was excellent; 73% of capsules were ingested as inferred by pill counts. Serum mineral levels, calcitropic hormones, and bone markers were similar between groups.

Conclusions: Oral Mg oxide capsules are safe and well tolerated. A positive effect of Mg supplementation on integrated hip BMC was evident in this small cohort. (J Clin Endocrinol Metab 91: 4866–4872, 2006)
Results

Study population

A total of 122 subjects were screened, 50 subjects enrolled, and 44 completed the study. Dropout rate was 37% (15%) for the placebo and two supplemented group. Reasons given are:

- 8-14 yr. old girls with low Mg intake
- N=50 (not 120)

Primary outcome:
- p=0.053 (not sign.)
2 observational studies

Current Research

The Relationship of Dietary and Lifestyle Factors to Bone Mineral Indexes in Children

WENDY BOUNDS, PhD, RD; JEAN SKINNER, PhD, RD; BETTY RUTH CARRUTH, PhD, RD; PAULA ZIEGLER, PhD, RD

ABSTRACT
Objective To identify factors related to children’s bone mineral indexes at age 8 years, and to assess bone mineral indexes in the same children at ages 6 and 8 years.

Conclusions Because many nutrients are related to bone health, children should consume a varied and nutrient-dense diet.


Osteoporosis
International

Influence of Pre-adolescent Diet on Quantitative Ultrasound Measurements of the Calcaneus in Young Adult Women

M. C. Wang¹,², E. C. Moore¹, P. B. Crawford³, M. Hudes³, Z. I. Sabry³, R. Marcus⁴,⁵ and L. K. Bachrach¹

Departments of ¹Pediatrics, ²Health Research and Policy, and ⁴Medicine, Stanford University School of Medicine, Stanford, California; ³School of Public Health, University of California, Berkeley, California; and ⁵Musculoskeletal Research Laboratory, Geriatric Research, Education, and Clinical Center, Veterans Affairs Medical Center, Palo Alto, California, USA

IUMSP
Institut universitaire de médecine sociale et préventive, Lausanne
**Observational study: Bounds 2005**

**Current Research**

The Relationship of Dietary and Lifestyle Factors to Bone Mineral Indexes in Children

WENDY BOUNDS, PhD, RD; JEAN SKINNER, PhD, RD; BETTY RUTH CARRUTH, PhD, RD; PAULA ZIEG

---

**ABSTRACT**

**Objective** To identify factors related to children’s bone mineral indexes at age 8 years, and to assess bone mineral indexes in the same children at ages 6 and 8 years.

**Conclusions** Because many health, children should consume more vitamin D and vitamin D intake in the diet.

J Am Diet Assoc. 2005;10

---

**Exposure:**
diet 2-8 yrs; 11 components incl. Mg

**Outcome:**
BMC & BMD at 8 yrs.

**Adjustment:**
sex, height, weight, BMI, age

Small coeff. & p=0.05

---

**Table 3. Multivariate regression models predicting children’s total BMC$^a$ and BMD$^b$ at 8 years old**

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>$\beta$</th>
<th>Partial $R^2$</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BMC Model 1$^d$ ($R^2=0.69, F=20.71, P&lt;.0001$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein intake (g)$^e$</td>
<td>(+) 2.40</td>
<td>0.08</td>
<td>.008</td>
</tr>
<tr>
<td>Height (cm)$^f$</td>
<td>(+) 9.49</td>
<td>0.50</td>
<td>.0005</td>
</tr>
<tr>
<td>Weight (kg)$^f$</td>
<td>(+) 7.22</td>
<td>0.05</td>
<td>.01</td>
</tr>
<tr>
<td>Age (y)$^f$</td>
<td>(+) 3.97</td>
<td>0.02</td>
<td>.01</td>
</tr>
<tr>
<td>Sex$^g$</td>
<td>(−) 61.33</td>
<td>0.04</td>
<td>.02</td>
</tr>
<tr>
<td><strong>BMC Model 2$^d$ ($R^2=0.69, F=20.08, P&lt;.0001$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphorus intake (mg)$^e$</td>
<td>(+) .11</td>
<td>0.05</td>
<td>.01</td>
</tr>
<tr>
<td>Height (cm)$^f$</td>
<td>(+) 8.28</td>
<td>0.50</td>
<td>.002</td>
</tr>
<tr>
<td>Weight (kg)$^f$</td>
<td>(+) 8.36</td>
<td>0.07</td>
<td>.003</td>
</tr>
<tr>
<td>Age (y)$^f$</td>
<td>(+) 39.82</td>
<td>0.03</td>
<td>.007</td>
</tr>
<tr>
<td>Sex$^g$</td>
<td>(−) 68.04</td>
<td>0.03</td>
<td>.008</td>
</tr>
<tr>
<td><strong>BMD Model 1 ($R^2=0.19, F=5.88, P=.005$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein intake (g)$^e$</td>
<td>(+) .01</td>
<td>0.07</td>
<td>.04</td>
</tr>
<tr>
<td>Sex$^g$</td>
<td>(−) .02</td>
<td>0.12</td>
<td>.03</td>
</tr>
<tr>
<td><strong>BMD Model 2 ($R^2=0.19, F=5.68, P=.006$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnesium intake (mg)$^e$</td>
<td>(+) .0002</td>
<td>0.07</td>
<td>.05</td>
</tr>
<tr>
<td>Sex$^g$</td>
<td>(−) .02</td>
<td>0.12</td>
<td>.03</td>
</tr>
</tbody>
</table>

$^a$n=52 children at 8 years of age.

$^b$BMC = bone mineral content (g).

$^c$BMD = bone mineral density; calculated as g/cm$^2$.
ESFA Guidance document 2007
Appendix I:
- references some reporting guidelines
- mirrors some STROBE items
- more specific for exposure information
- no item on methods for bias & multiple testing
Improving reporting of observational studies - some key aspects

- STROBE is designed as a tool for authors, editors, reviewers, and readers.

- Not a tool to assess methodological quality, but can be used to assess completeness & accuracy of reporting.

- Adherence does not guarantee a high-quality study but more transparency about what was done.
  - Users can judge themselves whether they trust in study results or not.
"Accurate and transparent reporting is like turning the light on before you clean up a room: It doesn’t clean it for you but does tell you where the problems are.”
Davidoff, Ann Intern Med 2000

- Weak studies can be reported well.
  Well-conducted studies can be reported weakly.
# Reporting guidelines initiatives

<table>
<thead>
<tr>
<th>Year</th>
<th>Initiative</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>CONSORT</td>
<td>RCTs (revised 2001 &amp; 2010)</td>
</tr>
<tr>
<td>1999</td>
<td>QUOROM</td>
<td>Meta-analyses of RCTs</td>
</tr>
<tr>
<td>2000</td>
<td>MOOSE</td>
<td>Meta-analyses of obs. studies</td>
</tr>
<tr>
<td>2003</td>
<td>STARD</td>
<td>Diagnostic studies</td>
</tr>
<tr>
<td>2004</td>
<td>TREND</td>
<td>Non-randomised studies</td>
</tr>
<tr>
<td>2007</td>
<td>STROBE</td>
<td>Case-control / Cross-sectional / Cohort studies</td>
</tr>
<tr>
<td>2007</td>
<td>COREQ</td>
<td>Qualitative studies</td>
</tr>
<tr>
<td>2008</td>
<td>SQUIRE</td>
<td>Quality improvement studies</td>
</tr>
<tr>
<td>2009</td>
<td>PRISMA</td>
<td>Syst. reviews &amp; meta-analyses (replacing QUOROM)</td>
</tr>
<tr>
<td>2013</td>
<td>SPIRIT</td>
<td>Protocols of RCTs</td>
</tr>
</tbody>
</table>

See: EQUATOR Library for Health Research Reporting