Review

The Risky Business of Studying Prognosis

LILY SIOK HOON LIM and BRIAN M. FELDMAN

ABSTRACT. Prognosis studies provide important healthcare information. Clinicians use prognostic factors to predict disease progress, thus allowing individualization of disease management. Prognosis is the issue in many translational studies that aim to identify biomarkers to predict outcomes. In a clinical trial, researchers may use prognostic factors to sort patients into risk groups, to clarify the effects of a new therapeutic agent. Prognosis studies can have significant effects on clinical practice. (First Release Dec 1 2012; J Rheumatol 2013;40:9–15; doi:10.3899/jrheum.121360)

Key Indexing Terms:
PROGNOSIS BIAS EPIDEMIOLOGY ANTIPHOSPHOLIPID SYNDROME
SYSTEMIC LUPUS ERYTHEMATOSUS RISK FACTORS

In this review we will focus on studies of prognosis. The report by Broder and Putterman[1] in this issue of *The Journal* was chosen as an example of a prognosis study. The investigators asked whether patients with systemic lupus erythematosus (SLE) treated with hydroxychloroquine (HCQ) were less likely to develop or maintain persistently active antiphospholipid antibodies (aPL) and/or lupus anti-coagulant (LAC).

Prognosis refers to the possible outcomes of a disease[2]. Prognosis studies either report the frequency of different outcomes or investigate relationships between prognostic factors and the occurrence of outcomes[3].

Prognosis studies provide important information to everyone involved in healthcare[4,5]. These studies suggest answers to patients' questions about outcomes. Clinicians use prognostic factors to predict which patients are going to do well or do badly, and thus to individualize disease management. In fact, prognosis is the question at the heart of many translational studies that aim to identify biomarkers to predict outcomes[6,7]. Researchers may also use prognostic factors to stratify patients into different risk groups, to clarify the effects of a new therapeutic agent in a clinical trial. Prognosis studies thus can have major effects on clinical practice.

Despite the obvious importance of prognosis studies, they are challenging to design. Most prognosis studies are observational: subjects are not randomized as in a randomized controlled trial. Researchers have no control over prognostic factors other than the one(s) being studied. Other prognostic factors may occur concurrently or subsequently. This uncertainty makes these studies especially vulnerable to biases unless careful considerations of methods to deal with them have been incorporated into the study design[8].

A recent editorial listed problems prevalent in prognosis research today[6]. One of the problems raised was poor methodological standards, e.g., underpowered studies, inappropriate statistical handling of data, and a lack of clearly defined primary outcomes. Very few prognosis research studies are reported as being clearly protocol-driven, unlike clinical trials. Without a clear protocol written before the conduct of the study and to guide analysis, the product of a study may be difficult to interpret — only the sensational or "statistically significant" results

From the Division of Rheumatology, The Hospital for Sick Children; the Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada.

L.S.H. Lim, MBBS, MRCPCH, PhD, Clinical Research Fellow and Doctoral Candidate; B.M. Feldman, MD, FRCPC, MSc, Head, Division of Rheumatology, Senior Scientist, The Hospital for Sick Children; Professor, Institute of Health Policy Management and Evaluation, University of Toronto.

Address correspondence to Dr. B.M. Feldman, Hospital for Sick Children, 555 University Avenue, Suite 8253, Toronto, Ontario, Canada.
E-mail: brian.feldman@sickkids.ca

Accepted for publication November 5, 2012.

See HCQ use and odds of aPL in SLE, *page 30*

This is the first in a series of articles that will appear in The Journal of Rheumatology. The editorial board will assign a paper published in the same issue to be reviewed and critiqued by clinical epidemiologists. This series aims to help readers gain a better understanding of the methodological underpinnings of different kinds of studies. We hope that a better understanding of research methods will help readers in selecting studies to inform their practice.

may be presented. A publication bias toward "positive" results is indeed detectable[9].

Hayden, *et al* identified 6 domains at risk of bias in the design of prognosis studies: study population, study attrition, outcome, prognostic factor, confounder, and statistical analysis[4]. They suggested that these 6 domains be assessed to judge the quality of prognosis studies when performing evidence synthesis. Although systematic reviews and metaanalyses can help to accelerate the incorporation of new therapies into patient management algorithms[10], systematic reviews of prognosis questions are relatively uncommon. This is due in part to the great heterogeneity of study designs and sample populations, and the highly variable mix of prognostic and confounding factors. Further, the prevalent low quality of prognosis studies means that the evidence will not be strong[11] even if a synthesis can be performed.

No observational study will ever be completely free of bias. Bias does not, however, exist to the same extent in the different domains (methodological areas) within a prognosis study[4]. Only by understanding the possible issues arising in different domains within a study can readers know how to use the information in the study and how much trust they can invest in the information presented. Readers of prognosis studies can also use the framework suggested by Hayden, *et al* to evaluate the quality of information in any prognosis paper. We will thus use Hayden's framework to evaluate the Broder and Putterman study in the 6 study domains common to prognosis studies[4].

Study population. To assess the study population, we should ask the questions listed in Table 1. This study included patients from a single center in Bronx, New York. We need to know whether this tertiary center is the only one in the Bronx seeing patients with SLE, and whether it sees the least severe cases, the most severe ones, or the whole range of severity. Information about the source population was limited in the report. This information is crucial for readers to understand the kinds of patients being studied and to assess whether the results can be applied to the patients under their own care.

How was the study sample assembled? To answer this, we need to know the inclusion and exclusion criteria, as well as the sampling strategy. Inclusion and exclusion criteria are used to homogenize the population being studied, but can run the risk of overselection so that the study sample is very different from the source population. The Broder and Putterman study included only those patients with at least 2 aPL measured. One might postulate a potential for an underestimation of the risk reduction; if aPL were measured repeatedly in patients who were at increased risk of clotting or who had a history of clotting, and if HCQ were prescribed more commonly in this group of patients (because of known beneficial effects of HCQ in reducing the risk for thrombotic events), this might result in an underestimation of the protective effect of HCQ. No information was reported regarding the sampling strategy, i.e., how the study participants were identified. Information about sampling strategy is important because different strategies to identify eligible patients, e.g., from a registry, laboratory records, or physicians' recall, are associated with different possibilities for bias. In an extreme example, if physician recall were used and physicians tended to recall those with the mildest (or the worst) disease manifestations, then the study population would be skewed to extremes of the eligible population. If only the mildest cases were included, for example, the effect of HCQ may be overestimated.

The characteristics of the study sample, including demographics and relevant clinical characteristics, should be described to an extent that the reader can understand the kind of patients being studied. Some of these characteristics were reported in Table 2 of Broder and Putterman. This information helps the reader to understand the context of this study, i.e., whether HCQ is protective from the time of diagnosis, or at any time in the disease course.

Prognosis studies should ideally be performed using inception cohorts, i.e., patients included at an early and similar point in the course of disease[12]. Some epidemiologists say that when choosing prognosis studies to inform practice, if an inception cohort was not assembled, move on to the next article[13]. Without an inception cohort, the conclusions drawn may be biased in unpredictable ways. If, for instance, only current patients are studied for prognosis, because those who had the most severe disease had already died, the observed outcomes will be overly optimistic. An inception cohort includes every patient with the disease at a uniform time (e.g., from the time of diagnosis), thus avoiding this problem.

Study attrition. To evaluate attrition, we first need to discern the kind of study design, because attrition is irrelevant for some. In cohort studies, the subjects are assembled based on their exposure status; therefore, if patients are chosen based on whether they received HCQ and are followed for their aPL/LAC status, that would be a cohort study[14]. Cross-sectional studies measure outcomes and exposures at the same time (e.g., a survey), and so really cannot identify predictive risks[14]. In the case-control design, researchers sample subjects with and without the outcome of interest (in this case, aPL/LAC status), and then identify predictors (i.e., HCQ exposure) retrospectively[14,15].

The research question in the Broder and Putterman study[1] suggests a cohort design. The study sample seems, however, to have been assembled based on outcome (aPL/LAC status), and the results are also presented according to the outcome. Although the authors called their study cross-sectional, the study may have really been a case-control design; more details would be necessary to be sure. Attrition is not relevant in either case — attrition is important when evaluating cohort studies.

Although we cannot assess attrition in the Broder and Putterman study, we will discuss how to assess attrition in prognosis studies where this is relevant. The questions to ask for assessing attrition are listed in Table 1. The proportion of patients left in followup at the end of a cohort study should always be reported. This is a matter of relative proportions and not a fixed number. Even 5% attrition can be potentially serious if the prevalence of outcome in the study population is low, e.g., 1%[2], but all 5% of those who were lost experienced the outcome of interest.

Reasons for attrition should be reported[16]. The distribution of prognostic factors and outcomes of those lost should be compared to those remaining in the study[4,17,18]. Observed outcomes will be biased if patients who do particularly well or badly preferentially leave a cohort[19,20]; reporting the outcomes of those who remained within the cohort will then be overly pessimistic (if only those who were very sick stayed) or optimistic (if only those who were very well stayed). This type of exploration for systematic differences is only rarely performed. There may be no information available to the researchers about those lost, or investigators may fear that revealing that they have a biased population will damage the credibility of their study. As a community, we can improve the quality of evidence by encouraging transparency in reporting potential biases in observational studies. This transparency to bias will help us use prognostic information more intelligently, by understanding the limitations of our evidence.

Outcome. We ask similar questions when assessing the study domains of outcome, prognostic factor, and confounder[4]. Table 1 lists the assessment questions.

We should look for a clear definition of the outcome of interest. In this case, the outcome is "persistently positive aPL/LAC." Persistence of antibodies was defined as 2 positive results at least 12 weeks apart. A moderate-high titer of ≥ 40 units was used to dichotomize aPL status as being positive or negative. Although the istotypes of aPL antibodies were defined (i.e., IgG, IgM, IgA), the kinds of aPL antibodies were not defined in the methods; this information was listed in Table 1 of the report. The timing of outcome measurement was unclear, i.e., whether the first aPL/LAC was performed within a certain period after diagnosis or during any assessment for any reason or at the time of a new referral.

We then ask whether the outcomes were measured in a valid and reliable way to limit misclassification[4]. Because this report sought to answer whether patients treated with HCQ were less likely to develop or maintain persistently positive aPL/LAC, the timing of outcome measurement is critical. From the research question, we must suppose that the outcome occurred after HCQ exposure and subjects either developed *new* aPL/LAC (from a previously negative or transiently positive state) or maintained the same positivity as prior to HCQ (for most subsequent measurements). The exact method by which outcomes were classified was unclear. Were patients classified based on 2 positive results *after* HCQ exposure? If there were 6 measurements of aPL/LAC with varying positivity over time (at least 2 positive), and varying HCQ exposure during that period among the patients, the number of possible combinations might be large (Figure 1). There is, therefore, a possibility of misclassification. This is always a problem when a time-varying outcome is treated as a single cumulative outcome, without clearly specifying the classification of

Table 1. Domains for assessment of the risk of bias in prognosis studies, adapted with permission from Hayden, *et al*[4].

Study Domains	Assessment Questions
Study population	1. Was the source population (from which the study sample was drawn) adequately defined?
	2. How was the sample assembled?
	3. Were the inclusion and exclusion criteria adequately described?
	4. Was there adequate participation by the eligible population?
	5. Was the baseline study sample adequately described in terms of key characteristics?
Study attrition	1. Was the response rate of followup adequate?
	2. Were the reasons for loss to followup reported?
	3. Were the participants lost to followup adequately described for key characteristics?
	4. Were there any attempts to collect information on those lost to followup?
Outcome prognostic factor	1. Was the outcome/prognostic factor well-defined?
	2. Was the outcome/prognostic factor measured in a valid and reliable manner (to limit misclassification)?
	3. Was the outcome/prognostic factor/confounder measured in similar settings and by similar methods?
Confounder	1. Was the confounder well-defined?
	2. Was the confounder measured in a valid and reliable manner (to limit misclassification)?
	3. Was the confounder measured in similar settings and by similar methods?
	4. Were all important confounders measured?
	5. Were important confounders accounted for in study design?
	6. Were important confounders accounted for in analysis?
Statistical analysis	1. Was there sufficient presentation of data to assess adequacy of analysis?
	2. Was the analysis appropriate (selected method/model and strategy of model-building)?
	3. Was there any risk of selective reporting?

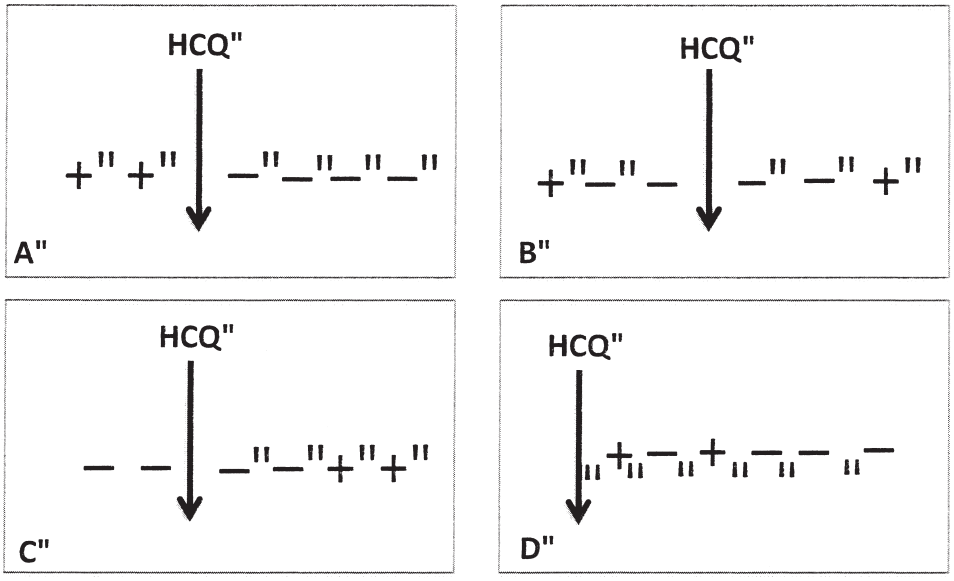

Figure 1. Different classification systems for outcomes as measured by antiphospholipid antibodies (aPL) and lupus anticoagulant (LAC). "+" denotes positive aPL/LAC measurement and "–" negative. HCQ denotes a patient's first known exposure to hydroxycholorquine. Potential classifications of outcomes could be highly variable in Broder and Putterman because no clear temporal relationship was defined for HCQ and aPL/LAC measurements. For example, (A) could be persistently positive aPL/LAC (if the first 2 positive aPL/LAC were defined as persistence), or loss of positive aPL/LAC (if temporal relationship to HCQ was considered). (B) could be persistently positive (any 2 positive) or transiently positive. (C) could mean development of new aPL after HCQ exposure but it is not constantly positive (under the common definition of requiring at least two-thirds of measurements)[37]. (D) could be persistently positive but not constantly positive. This could also be nonclassifiable because the status of aPL/LAC before HCQ was unknown.

various possible combinations. Depending on how the outcomes were classified, the protective effect of HCQ could be overestimated or underestimated.

Finally, we ask whether the setting and method of outcome measurement was similar for all subjects in the study. All the aPL and LAC measures were performed at the same institutional laboratory. The assay used for the aPL antibodies was reported but similar information was not reported for the LAC. Using the same method(s) and having the same setting of outcome measurement is important to ensure comparability of the results among all the subjects within the study. This information is needed to determine whether the study results could be applied directly to the reader's own practice.

Prognostic factors. The Broder and Putterman study[1] clearly named HCQ use as the primary prognostic factor of interest. To fully define this prognostic factor, the reader should be informed about the dose and duration of HCQ exposure. In this study, any history of exposure was taken to indicate the presence of the prognostic factor.

The method of prognostic factor measurement should be evaluated for validity and reliability to limit the possibility of misclassification. The validity of this prognostic factor (any exposure to HCQ) is challenging. This implies that exposure to any dose of HCQ, for any time period (whether a day or 5 years), at any time during the disease course, is considered similar in its effect on the outcome. The proportions of patients who could be classified as having the prognostic factor of interest increase when it is defined this way, but this approach may possibly underestimate the effect of HCQ on the aPL/LAC status.

The method and setting of prognostic factor measurement were not reported in the study. Readers would benefit from knowing how these data were collected to be satisfied about validity, e.g., through a chart review or from a research database, and whether by the same person.

Study confounding. Confounding may be a major problem in any non-experimental research. A confounder is a factor associated with the distribution of the prognostic factor within the sample. By itself, a confounder is a cause of the outcome but does not lie on the causal pathway between the prognostic factor and the outcome[8,14]. If confounding is present and not accounted for, any conclusion will be misleading. Confounders can be dealt with in the design and/or analysis phase. In the design phase, subjects can be matched or stratified by potential confounders (so that the confounders are no longer associated with the distribution of the prognostic factor). In the analysis stage, estimates of the

prognostic factor may be adjusted by including multiple confounders in multiple regression. Unfortunately, studies have shown that confounding is often not reported adequately[4,21].

Table 1 tells how to assess a study for confounding. In the Broder and Putterman study, no confounder was considered in the methods; we cannot answer the questions.

The authors did attempt to address confounding in the results when they explored their sample for confounding by indication[22]. Confounding by indication is present if HCQ is selectively prescribed to those with milder (or conversely, worse) disease[22]. To examine for the presence of this, the authors compared those who were treated with HCQ to those not treated with HCQ by using subject demographics, disease duration, "medications," and Charlson comorbidity index. In SLE, severity is likely related to disease activity or damage accrual. This concept of disease severity was not defined in the study. It is thus challenging to understand how severity has been accounted for; medication use likely indicates concurrent disease activity and the Charlson comorbidity score may be a surrogate for damage; demographics and disease duration do not perfectly correlate with either concept. In general, if confounding by indication were present, the effects of the prognostic factor cannot be accurately interpreted in isolation. In this instance, if HCQ were given to those with milder disease, the effects of HCQ could be overestimated: those with low disease activity may lose antibody positivity over time anyway, regardless of HCQ history.

The effects of confounders can be assessed by stratified analysis. Patients are stratified according to the confounder into 2 groups and then the relationship of the prognostic factor to the outcome is plotted in 2 contingency tables (by groups) for computing separate odds ratios. If the effect of HCQ disappears or changes significantly but similarly in the 2 stratified groups, as determined by the OR, then the confounding effect (say, of disease severity) is proven[8]. The authors performed a subgroup analysis on those who did not receive any immunosuppressant and who therefore were presumed to have mild disease. They concluded that HCQ was associated with an "independent effect on aPL/LAC positivity." This analysis would be more meaningful if the whole sample were used (n = 90; i.e., if a separate analysis was done using just those who were taking immunosuppressants, and then comparing the 2 analyses).

Study analysis. The answers readers seek to the questions regarding quality of analysis (Table 1) are found in clear descriptions of the analysis. The International Committee of Medical Journal Editors has advised that statistical methods should be reported in sufficient detail "to enable a knowledgeable reader with access to the original data to verify the reported results"[23]. Statisticians have long taught that specifics about variable selection should be presented and not only results of the final model[24,25]. The rationale for testing certain factors and including confounders should be discussed. This is because different strategies of variable selection may result in different analytic models[26]. Sometimes, there are several possible models. In such cases, the decision process resulting in the choice of the final model should be reported as well[16]. Different statistical models may be chosen that either overestimate or under-estimate the effect of a prognostic factor; it is important for readers to be clear about the choice of modeling strategy.

In the Broder and Putterman study, the authors informed us that they did not perform adjustments for multiple comparisons. They mentioned adjustment for logistic regression models (in the results) but did not provide information about model-building or reduction strategy[1].

It is important to identify the study design because the type of risk estimate reported and model used are determined by study design[27]. In a case-control study, where the outcome is binary (i.e., having the outcome of interest or not), logistic regression is commonly used. If there is matching, then conditional logistic regression should be used, because unconditional logistic regression may inappropriately inflate the OR[28]. In the case of the Broder and Putterman study, the choice of logistic regression is appropriate if, indeed, the study is a case-control study.

Finally, we assess whether the data were presented sufficiently and whether there was any risk of selective reporting. Univariable analysis was presented in Table 2 of Broder and Putterman and final multivariable models were reported in Table 3. Without any information about the modeling strategy, readers may question selective reporting.

We note that the authors have taken on a very difficult task with this research question, and many of the design and analysis decisions most likely reflected making the best of available data. We now proceed to suggest methods that future researchers may consider for similar kinds of questions.

In our discussion of outcome assessment, we alluded to the problem of simplifying repeated measured outcomes into a cumulative outcome. By doing this, researchers lose rich information in repeatedly measured data. This is a complex area; exposure to HCQ and other therapeutic agents in a relapsing disease such as SLE will vary over time. A longitudinal design in which outcomes and prognostic factors are repeatedly measured may be better at answering the question posed by this study. In this way, the aPL/LAC profiles need not be forced into a binary outcome on a single occasion. The temporal relationship of HCQ (prognostic factor) and aPL/LAC status (outcome) are clearly specified in a repeated-measures design. A longitudinal modeling method, such as the generalized estimating equation[29] or a binary mixed random-effects model[30], should be used for statistical inference where appropriate. The Broder and Putterman study would likely have benefited from this approach.

The authors have taken on the very challenging task of trying to answer therapeutic-type questions in an observational setting. The authors were rightly concerned about confounding by indication affecting the use of HCQ[22]. When the kinds of patients exposed to particular treatments are systematically different from those who are not exposed, it is impossible to comment on relative efficacy of the treatments. Stratified analysis (stratified by a single potential confounder) is not possible if there are several confounders. Propensity score techniques[31,32] have been successfully used in observational settings to address the problem of matching for many confounders[33,34], but these kinds of studies are still rare. Expert statistical consultation is advisable.

Broder and Putterman have tackled an interesting research question[1]. It is, indeed, hard to answer complicated questions using the methods commonly described in the prognosis literature today. We have learned that HCQ may possibly have a role in reducing the odds of persistently positive aPL/LAC. The true effect is hard to judge; there are several aspects of the study design that may lead to an overestimation or underestimation of the effects of HCQ. aPL/LAC is a known significant factor predicting thrombotic events[35,36,37]. HCQ has been shown to be protective against thrombotic events in several studies in which more sophisticated analyses were performed[35,36]. It is intuitive that reducing aPL/LAC may translate into reduced thrombosis. This lends support to the findings from the Broder and Putterman study.

We have demonstrated how the reader may systematically assess a prognosis study. In designing any prognosis study, researchers should seek to decrease the risk of bias that may result from various sources: study population, attrition, measurement of the prognostic factor, measurement of the confounder, measurement of the outcome, and statistical analysis. Readers should evaluate each study of prognosis rigorously to decide how, or even whether, to use the information. Journals can help improve the overall standards of reporting in observational studies by promoting the Strengthening The Reporting of OBservational studies in Epidemiology (STROBE) standards[16].

REFERENCES

1. Broder A, Putterman C. Hydroxychloroquine use is associated with lower odds of persistently positive antiphospholipid antibodies and/or lupus anticoagulant in systemic lupus erythematosus. J Rheumatol 2013;40:30-3.
2. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. JAMA 1994;272:234-7.
3. Altman DG. Systematic reviews of evaluations of prognostic variables. BMJ 2001;323:224-8.
4. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006;144:427-37.
5. Hemingway H. Prognosis research: Why is Dr. Lydgate still waiting? J Clin Epidemiol 2006;59:1229-38.
6. Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. BMJ 2009;339:b4184.
7. Hemingway H, Henriksson M, Chen R, Damant J, Fitzpatrick N, Abrams K, et al. The effectiveness and cost-effectiveness of biomarkers for the prioritisation of patients awaiting coronary revascularisation: A systematic review and decision model. Health Technol Assess 2010;14:1-151, iii-iv.
8. Rothman KJ, Greenland S, Lash TL, editors. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
9. Matthews GA, Dumville JC, Hewitt CE, Torgerson DJ. Retrospective cohort study highlighted outcome reporting bias in UK publicly funded trials. J Clin Epidemiol 2011;64:1317-24.
10. Berlin JA, Colditz GA. The role of meta-analysis in the regulatory process for foods, drugs, and devices. JAMA 1999;281:830-4.
11. Hayden JA, Chou R, Hogg-Johnson S, Bombardier C. Systematic reviews of low back pain prognosis had variable methods and results: Guidance for future prognosis reviews. J Clin Epidemiol 2009;62:781-96 e1.
12. Ales KL, Charlson ME. In search of the true inception cohort. J Chronic Dis 1987;40:881-5.
13. How to read clinical journals: III. To learn the clinical course and prognosis of disease. Can Med Assoc J 1981;124:869-72.
14. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing clinical research. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2007.
15. Dekkers OM, Egger M, Altman DG, Vandenbroucke JP. Distinguishing case series from cohort studies. Ann Intern Med 2012;156:37-40.
16. Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. PLoS Med 2007;4:e297.
17. Ware JH. Interpreting incomplete data in studies of diet and weight loss. N Engl J Med 2003;348:2136-7.
18. Little RJ, Rubin DB. Statistical analysis with missing data. 2nd ed. New York: Wiley; 2002.
19. Rubin DB. Inference and missing data. Biometrika 1976;63:581-92.
20. Johnson ES. Bias on withdrawing lost subjects from the analysis at the time of loss, in cohort mortality studies, and in follow-up methods. J Occup Med 1990;32:250-4.
21. Mullner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: A cross-sectional survey. Ann Intern Med 2002;136:122-6.
22. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol 1999;149:981-3.
23. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: manuscript preparation and submission: preparing a manuscript for submission to a biomedical journal. 2012. [Internet. Accessed Nov 5, 2012.] Available from: http://www.icmje.org/manuscript_1prepare.html
24. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. BMJ 1983;286:1489-93.
25. Clayton D, Hills M. Statistical models in epidemiology. Oxford: Oxford University Press; 1993.
26. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modelling with logistic regression analysis: A comparison of selection and estimation methods in small data sets. Stat Med 2000;19:1059-79.
27. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. PLoS Med

2007;4:e296.
28. Holford TR, White C, Kelsey JL. Multivariate analysis for matched case-control studies. Am J Epidemiol 1978;107:245-56.
29. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: A generalized estimating equation approach. Biometrics 1988;44:1049-60.
30. Stiratelli R, Laird N, Ware JH. Random-effects models for serial observations with binary response. Biometrics 1984;40:961-71.
31. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998;17:2265-81.
32. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984;79:516-24.
33. Benseler SM, Bargman JM, Feldman BM, Tyrrell PN, Harvey E, Hebert D, et al. Acute renal failure in paediatric systemic lupus erythematosus: Treatment and outcome. Rheumatology 2009;48:176-82.
34. Lam CG, Manlhiot C, Pullenayegum EM, Feldman BM. Efficacy of intravenous Ig therapy in juvenile dermatomyositis. Ann Rheum Dis 2011;70:2089-94.
35. Jung H, Bobba R, Su J, Shariati-Sarabi Z, Gladman DD, Urowitz M, et al. The protective effect of antimalarial drugs on thrombovascular events in systemic lupus erythematosus. Arthritis Rheum 2010;62:863-8.
36. Kaiser R, Cleveland CM, Criswell LA. Risk and protective factors for thrombosis in systemic lupus erythematosus: Results from a large, multi-ethnic cohort. Ann Rheum Dis 2009;68:238-41.
37. Martinez-Berriotxoa A, Ruiz-Irastorza G, Egurbide MV, Garmendia M, Gabriel Erdozain J, Villar I, et al. Transiently positive anticardiolipin antibodies and risk of thrombosis in patients with systemic lupus erythematosus. Lupus 2007;16:810-6.