SciMedCentral Journal of Translational Medicine & Epidemiology

Editorial

Lesser Known Facts about Nested Case-Control Designs

Ryung S. Kim*

Department of Epidemiology and Population Health, Albert Einstein College of Medicine, USA

Corresponding author

Ryung S. Kim, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York 10461, USA, Email: ryung.kim@einstein.yu.edu

Submitted: 21 September 2013

Accepted: 21 September 2013

Published: 23 September 2013

Nested case-control designs (or equivalently, incidence density sampling designs) are are a common approach for reducing the costs the costs of exposure assessment in a prospective epidemiologic study. Exposure data are obtained from all cases and a pre-specified number of controls randomly chosen at each failure time from all subjects who had entered the study but had not failed or left it yet [1]. The subjects selected across failure times are confusingly referred as 'controls'. However, unlike traditional matched case-control designs, these 'controls' may include failures and some controls may be selected multiple times across different failure times. There can be a large redundancy between the controls.

It is well documented that a naïve Cox regression analysis yields biased results when it is applied to nested case-control studies. Traditionally, the conditional logistic approach [1] were used to analyze these studies by 'tricking' statistical software written for conditional logistic regression by including multiple inputs for subjects who are selected multiple times, converting all randomly selected failures to non-failures. The resulting estimates of the conditional logistic approach are that of log of hazard ratios and not odds ratios. However, as Langholtz [2] commented, the 'fixation' on odds ratios is still pervasive. For example, among sixteen nested case-control studies published in the American Journal of Epidemiology between 2009 and 2011, only one interpreted its estimates as hazard ratios.

When the retrospective access to the outcome and the matching variables of the full cohort is available, several methods can be used that are more powerful than the conditional logistic approach. For example, Samuelsen [3] proposed a method in which the individual log-likelihood contributions are weighted by the inverse of the inclusion probabilities of ever being included in the nested case-control study. This method was shown to be more efficient than the conditional logistic regression approach. Chen [4,5] considered the same form of the likelihood, but refined the weights by averaging the observed covariates from subjects with similar failure times to estimate the contribution from unselected controls. This method was also shown to be more efficient than the conditional logistic approach. In addition, one may model exposures non-parametrically conditioned on other 'always-observed' covariates. In some situations, these maximum semiparametric likelihood has shown to increase the efficiency [6,7]. However, it is surprising that these methods are rarely used in practice. For example, all aforementioned sixteen nested case-control studies were analyzed by the conditional or unconditional logistic approaches. None used Samuelsen or Chen's methods.

Finally, one of the main perceived limitations of nested case-control designs in epidemiologic studies had been their inability to re-evaluate the data to study the association between a new outcome and the exposure data of the original study. While this limitation existed a decade ago, valid inferences about secondary outcomes can be made in nested case-control studies using the inverse probability weighting approach [8-10] or the likelihood approach [11]. To my knowledge, however, these analytical methods are still rarely used to analyze secondary outcomes of nested case-control studies. In addition, the inverse probability weighting method can be used to merge data sets across varying study designs. Large cohort studies often have numerous sub-studies that are nested case-control designs, case-cohort-designs, or matched case-control designs. If eligibility criteria are similar across these studies, the existing data sets can be merged and hypotheses can be generated and tested. It is time to let the words get out.

REFERENCES

1. Thomas D. Addendum to 'Methods of cohort analysis: Appraisal by application to asbestos mining' by Liddell FDK, McDonald JC, Thomas DC. Journal of the Royal Statistical Society. 1977; 140: 469-491.
2. Langholz B. Case-Control Studies = Odds Ratios: Blame the Retrospective Model. Epidemiology. 2010; 21: 10-12.
3. Samuelsen S. A pseudo-likelihood approach to analysis of nested case-control studies. Biometrika. 1997; 84: 379–394.
4. Chen KN. Generalized case-cohort sampling. J R Statist Soc B. 2001; 63: 791-809.
5. Chen KN. Statistical estimation in the proportional hazards model with risk set sampling. Annals of Statistics. 2004; 32: 1513-1532.
6. Chen HY. Double-Semiparametric Method for Missing Covariates in Cox Regression Models. Journal of American Statistical Association. 2002; 97: 565-576.
7. Scheike TH, Juul A. Maximum likelihood estimation for Cox's regression model under nested case-control sampling. Biostatistics. 2004; 5: 193-206.
8. Salim A, Yanga Q, Reilly M. The value of reusing prior nested case-control data in new studies with different outcome. Statistics in Medicine. 2012; 31: 1291-1302.

Cite this article: *Kim RS (2013) Lesser Known Facts about Nested Case-Control Designs. J Transl Med Epidemiol 1(1): 1007.*

9. Kim RS. Analysis of Secondary Outcomes in Nested Case-Control Study Designs. Division of Biostaistics, Albert Einstein College of Medicine: City, 2013.

10. Støer N, Samuelsen SO. Comparison of estimators in nested case-control studies with multiple outcomes. Lifetime Data Anal. 2012; 18: 261-283.

11. Saarela O, Kulathinal S, Arjas E, Läärä E. Nested case-control data utilized for multiple outcomes: a likelihood approach and alternatives. Stat Med. 2008; 27: 5991-6008.

Cite this article

Kim RS (2013) Lesser Known Facts about Nested Case-Control Designs. J Transl Med Epidemiol 1(1): 1007.