# Semiparametric and Nonparametric Modeling for Matched Studies

Inyoung Kim *and Noah Cohen †


## Abstract

This study describes a new graphical method for assessing and characterizing effect modification by a matching covariate in matched case-control studies. This method to understand effect modification is based on a semiparametric model using a varying coefficient model. The method allows for nonparametric relationships between effect modification and other covariates, or can be useful in suggesting parametric models. This method can be applied to examining effect modification by any ordered categorical or continuous covariates for which cases have been matched with controls. The method applies to effect modification when causality might be reasonably assumed. An example from veterinary medicine is used to demonstrate our approach. The simulation results show that this method, when based on linear, quadratic and nonparametric effect modification, can be more powerful than both a parametric multiplicative model fit and a fully nonparametric generalized additive model fit.


**Keywords**: Conditional Logistic Regression; Effect Heterogeneity; Effect Modification; Local Likelihood Regression; Matched Case-Control Studies; Varying Coefficient Model.

**Running Title**: Modeling for Effect Modification in Matched Studies:

*Inyoung Kim (kiy@yumc.yonsei.ac.kr). Cancer Metastasis Research Center, Yonsei University, 134Sinchon-dong, Seodaemun-gu, Seoul 120-749, Korea

†Noah Cohen (ncohen@cvm.tamu.edu), Department of Large Animal Medicine and Surgery, College of Veterinary Medicine, Texas A& M University, College Station, TX 77843-4475, USA.

# 1 Introduction

In a classical logistic regression model based on a binary outcome D (case-control status), a covariate X and stratum level S is given by

$$\Pr(D = 1|X, S) = H\{X^T\beta + q(S)\} \tag{1}$$

where $H(\cdot)$ is a logistic distribution function, q(s) is a semiparametric function which includes the intercept and the unknown effects of the strata. The classical matched case-control study begins with model (1), but uses conditional analysis which makes all matched pairs independent since conditional analysis removes any stratum effect and q(s) (Hosmer and Lemeshow, 1989).

Analysis of a matched case-control study of this type was conducted by Cohen, *et al.* (1999) to determine whether dietary and other management covariates were associated with the development of colic in horses after matching on veterinarian and month of examination: equine colic is an important cause of disease and death in horses, frightening for the animal's human companions, and of considerable financial import. Results of this study indicated that a change in diet, regular parasitic control, and age were associated with the risk of colic. However, veterinarians who participated in the study contributed between one and twelve matched pairs of cases and controls. Therefore, a question arises whether characteristics of the veterinarian (a matching covariate) also affect the strength of the relationships. That is, whether there is effect modification for covariates of interest such as change in diet, regular parasitic control or horse age and risk of colic associated with a characteristic of the veterinarian, such as the number of pairs contributed by the veterinarian (cluster size).

It is generally accepted that covariates for which cases and controls are matched cannot exert a confounding effect on independent covariates included in analyzes (Breslow and Day, 1980) but we believe effect modification by matching covariates may occur. This question motivates our research. To our knowledge, methods for evaluating the magnitude and direction of effect modification by matching covariates in matched case-control studies are not well described. The aim of this paper is to describe methods for assessing and characterizing effect modification by a matching covariate in matched case-control studies: one of these methods is a new graphical method based on a semiparametric model. The graphs themselves are easily computed using any conditional logistic regression (CLR) program. The graphical method is developed using a varying coefficient model (Tibshirani and Hastie, 1987; Carroll, Ruppert and Welsh, 1998).

A simple parametric way to test for effect modification is to add one or more multiplicative interaction terms to the model, multiplying the potential effect modifier by the covariate in question, and then to test whether this derived variable has a statistically significant effect. This can be done either simultaneously, or in a series of CLR with each variable alone, but in turn having the multiplicative interaction.

## 2 Methods

Our method differs from the simple parametric way in two respects: (a) it is graphically based rather than based on statistical significance tests; and, (b) it does not assume that the effect modification is of a multiplicative form. The advantage of this graphical method based on varying coefficient model is that it can allow for nonparametric relationships between the effect modifier and regression parameters. The other proposed method is based on a generalized additive model (Hastie and Tibshirani, 1990) which requires specialized software. This alternative is not a varying coefficient model for effect modification except in the particular case where all covariates are binary.

## 3 Application

We performed several simulations to compare a parametric interaction method(PIM), a graphical method using a varying coefficient model(VCM) and an alternative method using a generalized additive model (GAM). Our simulations suggest that our approachs are more efficient than the simple parametric interaction method. Finally, we apply our approaches to the equine epidemiology example that motivated this work.

## REFERENCES

Breslow, N.E., and Day, N.E. (1980). *Statistical Methods in Cancer Research. Vol.1 -The analysis of case-control studies.* Lyon: International Agency on Cancer, 32, pp162-188; pp248-278.

Carroll, R.J., Ruppert, D. and Welsh, H. (1998). Local Estimating Equation, *Journal American Statistical Association*, 93, 214-227.

Cohen, N. D., Gibbs, P. G. and Woods, A. M. (1999). Dietary and Other Management Factors Associated with Colic in Horses. *Journal of American Veterinary Medical Association*, 215, 53-60.

Eilers, Paul H. and Marx, Brian D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89-121.

Hastie, T and Tibshirani, R. (1990). *Generalized Additive Models*, New York: Chapman & Hall.

Hosmer, D. W. and Lemeshow, S. (1989) *Applied Logistic Regression*, New York: Wiley.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation, *Journal American Statistical Association*, 82, 559-567.