Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computational Statistics & Data Analysis 50 (2006) 1807 – 1817

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Existence and uniqueness of relative incidence estimates in case-series analysis

Baibing Li[a,*], Xiangning Huang[b]

[a] Business School, Loughborough University, Ashby Road, Loughborough LE11 3TU, UK
[b] Pfizer Global Research and Development, Sandwich CT13 9NJ, UK

Received 4 February 2005; accepted 4 February 2005
Available online 14 March 2005

Abstract

Case-series analysis is used to estimate relative incidences of clinical events in defined time intervals after vaccination compared to a control period. It has advantages, in terms of both a reduction in data collection effort, because it uses only data on cases, and a reduction in the resultant variances of estimates, due to individuals being self-controlled. The existence and uniqueness of relative incidence estimates in case-series analysis are investigated. For the relative incidence of a clinical event, a simple condition for existence and uniqueness of the estimate of the parameter vector in a case-series model is established. An algorithm is developed to examine the established condition, which provides a clue for remedy when the condition for existence and uniqueness is not satisfied.

Keywords: Case-series analysis; Existence and uniqueness; Maximum likelihood estimate; Nonlinear programming; Relative incidence

1. Introduction

Evaluation of vaccine safety is an important aspect of vaccine programme surveillance. For example, several research projects have recently investigated the possible relationship between the MMR vaccine and autism (see e.g. Taylor et al., 1999; Farrington et al., 2001).

Most of the existing methods for evaluation are data-intensive, involving large cohorts or the careful selection and matching of controls (Begg and Miller, 1990; Farrington, 1995).

* Corresponding author. Tel.: +44 1509 228841; fax: +44 1509 223960.
E-mail address: b.li2@lboro.ac.uk (B. Li).

0167-9473/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2005.02.002

Farrington (1995) developed a case-series model that solely uses data on cases, thus substantially reducing the data collection effort. It combined aspects of the case control and cohort methods, using retrospectively ascertained vaccination histories in cases to estimate the relative incidence of an event on different time intervals after vaccination relative to a control period. Consequently, the resultant variances of estimates reduce due to individuals being self-controlled and individual effects and covariates being canceled out through conditioning. In contrast, in case control studies, individual or strata effects have to be taken into account during modeling and inference; see for instance, Kim and Cohen (2004) for a semiparametric/nonparametric modeling approach for effect modification in matched studies. The case-series model has, for example, been applied to assess vaccine safety in a study of intussusception among infants given an oral rotavirus vaccine (Farrington et al., 1995), and in a study of the association between diphtheria/tetanus/pertussis (DTP) vaccination and febrile convulsion (Murphy et al., 2001).

The unknown parameter vector of the case-series model developed by Farrington (1995) consists of a relative incidence part and an association part correcting age effects. After obtaining an estimate that maximizes the log-likelihood kernel of the case-series model, the estimate of log relative incidence can be used for the evaluation of vaccine safety.

The question arises as to whether a relative incidence estimate exists, and if so, whether it is unique. When it does not exist or is not unique, the solution obtained through an optimization algorithm may give misleading indications for vaccine safety evaluation. In general, one does not assume infinite parameter values in underlying populations, and thus infinite parameter estimates are not normally acceptable in practice (Albert and Anderson, 1984; Heinze and Schemper, 2002). The problems of infinite parameter values typically occur with small to medium-sized data sets. One possible solution is to reparameterize the relevant parameters during numerical calculations. However, interpretation of the results in terms of the original parameters, if they are of primary interest, may still be problematic. A well-known example is the separation problem in logistic regression analysis. The separation problem may result in at least one parameter estimate diverging to infinity, and lead to Wald confidence intervals of infinite width (Albert and Anderson, 1984; Heinze and Schemper, 2002).

The issue of existence and uniqueness has been given much attention in the statistical literature. Silvapulle and Burridge (1986) presented a general condition for the existence of maximum likelihood estimates (MLE) in regression models. Specifically, the MLE of a parameter vector, say $\boldsymbol{\theta}$, in a regression model exists if and only if there does not exist a non-zero vector $\mathbf{u}$ having the same dimension of $\boldsymbol{\theta}$ such that $\mathbf{Zu} \leqslant \mathbf{0}$, where $\mathbf{Z}$ is a matrix calculated from observation data. Mathematically, this condition is simple and can be verified by linear programming methods (Silvapulle and Burridge, 1986). Unfortunately, when an MLE does not exist, this condition does not provide any information for diagnosis and remedy. Recently, for the logistic regression analysis, Heinze and Schemper (2002) developed a procedure to remedy the separation problem in logistic regression analysis.

For a case-series model, the issue of diagnosis and remedy is particularly important and useful because the definition of age intervals in a case-series model can be adjusted to a considerable extent, and thus even if the MLE of a parameter vector does not exist under one partition of age intervals, it may exist under another. A diagnosis of the problem may thus provide information for re-defining age intervals. In this paper, we establish a simple