ANALYSIS OF INFECTIOUS DISEASE DATA FROM A SAMPLE OF HOUSEHOLDS

Niels G. Becker
La Trobe University, Australia.

Summary

Observations are made on the individuals from a sample of households at two points in time. At the first time point it is determined who is susceptible to a certain infectious disease. At the second time point it is determined which of the susceptibles have been infected. A method for the analysis of such data is derived with the aid of martingale theory and suitable epidemic model assumptions. In contrast to a previously proposed analysis the present approach requires only a modest amount of manual calculation and is based on less restrictive assumptions. An application to influenza data is given.

1. Introduction Consider the spread of an infectious disease through a community of households. Suppose the disease is transmitted primarily by sufficiently close contacts between susceptible and infectious individuals, and that such contacts may take place between any pair of individuals.

1.1. Data on Sizes of Outbreaks in Households. Each of the individuals from a random sample of households is tested at time 0 and again at time τ. At time 0 it is determined who is susceptible to the infectious disease and at time τ it is determined which of the susceptibles has been infected since time 0. The time interval $(0, \tau)$ might be the time interval during which the epidemic passes through this community, or the 'epidemic season' of a disease which is endemic in the community. It is important to consider this type of infectious disease data, because it is relatively easy to obtain and is often reliable. When based on tests of sera one tends to obtain more objective data and can hope to cope with the presence of subclinical infections.

It has been the practice to analyse such infectious disease data by assuming that outbreaks within households essentially evolve independently of each other. However, it seems safer to base the analysis on a model which permits both between and within household infection. Such an analysis is described by Becker & Hopper (1983) for a community in which every household is kept under observation. Complete observation is impractical for large communities. Haber et al. (1988) propose an analysis which can be applied to data from a random sample of households from a larger community. Here we demonstrate that the ap-