Using Case-Cohort Studies to Estimate Excess Mortality in Large Cancer Cohorts

Hannah Twitchell (1)*, Clare Frobisher (2), Raoul C. Reulen (2), Michael M. Hawkins (2), Paul C. Lambert (1,3)

(1) Department of Health Sciences, University of Leicester, UK, (2) The Centre for Childhood Cancer Survivor Studies, University of Birmingham, UK, (3) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden.

Introduction

Case-cohort studies are an **efficient** way to analyse large cancer cohorts when additional detailed exposure information may be time-consuming or expensive to obtain for the entire cohort. They can be a **useful alternative to nested case-control studies**. We extend the case-cohort approach to model excess mortality, i.e. how much higher mortality is in a diseased group compared to that expected in the general population, using **flexible parametric models** [1].

A case-cohort study samples a random fraction of the entire cohort - the **subcohort**. The subcohort is determined at the start of time at risk. All **cases** (events) in the cohort are then used along with non-cases from the subcohort using appropriate **weights** in the analysis, so that non-cases are upweighted to represent multiple non-cases. This provides **unbiased** estimates of the quantities of interest, without a major loss in precision compared to analysing the full cohort.

Table 1: An example of a case-cohort sample drawn from the cohort, sampling fraction (SF) 0.2, Borgan II weights given, grey cells denote individuals contributing to the analysis. Total analysed 6006.

	In subcohort *(weight)*	Not in subcohort *(weight)*	Full cohort
Case	639 *(1)*	2410 *(1)*	3049
Non-case	2957 *(5.05)*	11975	14932
Total	3596	14385	17981

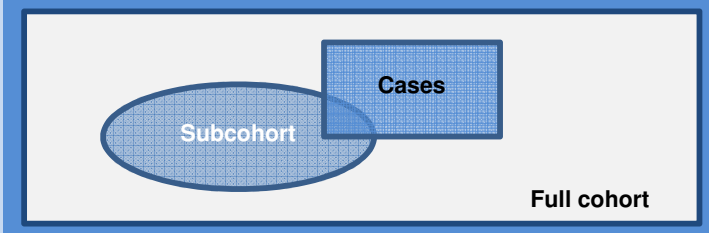

Figure 1: Visualising the case-cohort sampling procedure

Methods

Most previous analyses of case-cohort studies have used a weighted Cox model. Most case-cohort weights can be applied to flexible parametric models (FPM). Here Borgan II weights are used [2]. Such models are adopted as we are particularly interested in estimating **absolute rates** and **time-dependent effects**, i.e. non-proportional excess hazards. In many analyses of cancer data we want to be able to model time-dependent effects.

An application of case-cohort methods using FPMs is shown using 17,891 5-year survivors of cancer diagnosed before age 15 years from the British Childhood Cancer Survivor Study (BCCSS) [3]. We demonstrate the approach using a sampling fraction of 0.2.

A flexible parametric model is used with covariates sex, first cancer type and age (grouped), for illustration purposes. The baseline excess hazard and the time-dependent effect of first cancer group are modelled using restricted cubic splines.

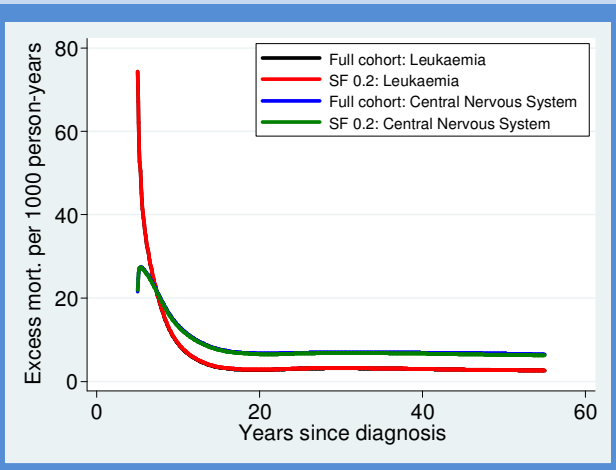

Figure 2: Excess mortality rates for 5-year childhood cancer survivors first diagnosed with CNS tumours, and those first diagnosed with leukaemia

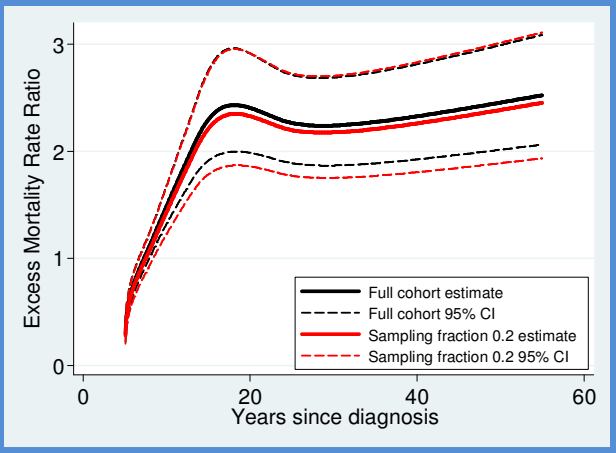

Figure 3: Excess mortality rate ratio for 5-year childhood cancer survivors first diagnosed with CNS tumours compared to those first diagnosed with leukaemia

Results

Using case-cohort sampling with appropriate sampling fractions, unbiased estimates can be obtained without much loss in precision. Previous work on this data indicates long-term excess mortality for these childhood cancer survivors. In such studies where detailed exposure information is needed, the numbers required could be estimated using far fewer individuals than would be the case if analysing the full cohort.

Conclusion

Case-cohort methods have not previously been used with excess mortality models. The flexible parametric modelling framework allows for their use. Historically the case-cohort design has been underused in cancer research, but is set to become more popular. It is particularly useful when detailed exposure information is required, such as genetic information using blood samples or, in the case of BCCSS, information on the doses of chemotherapy and/or radiotherapy received to each organ. Efficient absolute estimates can be made by using a small sample of a full cohort.

References

(1) Lambert and Royston, Further Development of Flexible Parametric Models for Survival Analysis, Stata Journal 2009, (2) Borgan et al., Exposure stratified case-cohort designs, Lifetime Data Anal. 2000, (3) Hawkins et al., The British Childhood Cancer Survivor Study, Pediatric Blood and Cancer, 2008

Contact: Hannah Twitchell, Email: hkt10@le.ac.uk