Abstract 125

IMPROVING MODEL SELECTION IN OCCUPATIONAL EXPOSURE ASSESSMENT: AN EXAMPLE OF MULTI-MODEL INFERENCE AND MULTI-MODEL AVERAGING USING LINEAR MIXED-EFFECT MODELS

J Lavoué[1], PO Droz[2]
[1]Institute for Work and Health, [2]Institute for Occupational Health Sciences

Objectives: While still much present in the industrial hygiene literature, p-value-based model selection is currently being questioned by an increasing number of statisticians. A trend is emerging that emphasizes making statistical inferences from a set of plausible models rather than from a single model regarded as 'best'. The objective of this paper is to present a case study of modeling occupational exposure data using multi-model inference, as introduced in the monograph by Burnham and Anderson: Model selection and multi model inference 2nd edition (Springer, 2002), and applied in the context of linear mixed-effect models. **Methods:** Respirable dust exposure levels collected over the years by the Institute for occupational health sciences (Lausanne, Switzerland) were selected for this study. A set of plausible models was defined a priori by taking into account the sample size and previous knowledge of variables influent on exposure levels. A finite sample version of the Akaike information criterion (AICc) was calculated to evaluate the relative support of the data for each model. Multi-model inference was performed by averaging predictions and coefficients over the set of models, with weighing based on the values of AICc. Alternate weights were calculated using a bootstrap procedure. For comparison purpose, traditional forward stepwise selection was performed using p-values, the Aikaike information criterion, and the Bayesian information criterion. **Results:** The modeling datasets contained, respectively, 432 and 607 personal and area measurements taken between 1987 and 2006. The a priori set of models contained 96 models for both types of measurements. The relative weights of the different models showed that no single model had a clear support from the data at hand. As measured by cumulative relative weights, the five best models represented, respectively, 92 and 80% of evidence for the personal and area measurements. Model averaged predictions and accompanying confidence intervals, unconditional on a particular model, were calculated. **Conclusion:** Multi-model inference represents a promising procedure that incorporates the notion that several models can be supported by the data and permits to evaluate, to a certain extent, model selection uncertainty, which is seldom mentioned in current practice.

Key words: multi-model inference; linear mixed-effect models; model selection

Abstract 126

LONGITUDINAL ANALYSIS OF LUNG FUNCTION DATA AND IMPACT OF OCCUPATIONAL EXPOSURES - A METHODOLOGICAL CONSIDERATION OF DIFFERENT MODELS

J Cui[1], M Abramson[2], N de Klerk[3], M Dennekamp[2], A Del Monaco[2], G Benke[2], B Musk[3], M Sim[2]
[1]Monash University, [2]Department of Epidemiology and Preventive Medicine, Monash University, [3]School of Population Health, University of Western Australia

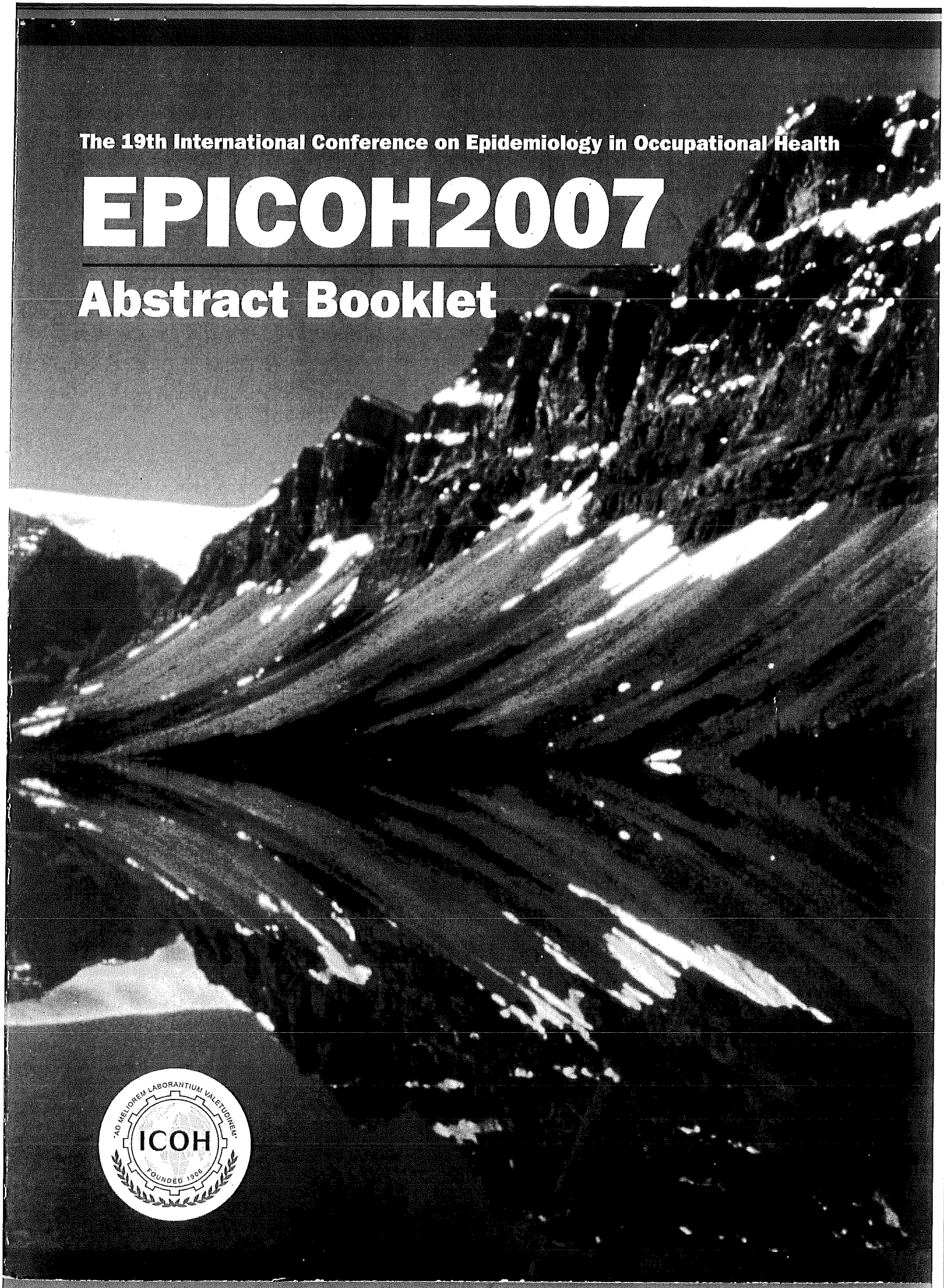
Objectives: Occupational health longitudinal studies often need to analyze the dose-response relationship between relevant exposure variables and health outcomes. However, exposure data are usually not normally distributed and not every individual is exposed to the contaminant of interest. The aim of this study was to use several different statistical methods to model longitudinal occupational exposure data and select the optimal approach to analyze the risk of exposure on a health outcome. **Methods:** The main outcome variable was the FEV1 (Forced Expiratory Volume in one second) measured on 446 individuals at their annual interviews in an Inception Cohort study in aluminum smelters. The main exposure variable was the cumulative exposure of fluoride (mg/m3-year) since start of employment. A linear Mixed Effects Model and GEE methods were used in the analysis of the longitudinal data. Different methods were used to analyze the exposure data. The Quasi-likelihood under Independence Model Criterion (QIC) was used to select the best model. **Results:** The model using tertiles of fluoride exposure had the lowest QIC and thus gave the best fit to the data. The next best model was to use the trend of the tertiles. The model using the original scale gave the worst fit to the data and the model using log-transformation was slightly worse than the model using trend over tertiles. **Conclusion:** We recommend using the tertile method in similar analyses of longitudinal data. The model using the original exposure data gave the worst fit and should be avoided. The model using the trend over tertiles can also be used in evaluation of a dose-response relationship.

Key words: longitudinal; lung function; statistical models

The 19th International Conference on Epidemiology in Occupational Health

EPICOH2007

Abstract Booklet



ABSTRACTS

EPICOH2007

19th International Conference on
Epidemiology in Occupational Health

Banff, Alberta, Canada
October 9-12, 2007

Frontiers of Occupational Epidemiology