

A research paper published in the *Statistical Methods in Medical Research*<sup>§</sup>

**“Handling incomplete smoking history data in survival analysis”**

Kyoji Furukawa, Dale L. Preston, Munechika Misumi, and Harry M. Cullings

*Stat Methods in Med Res* 2014 (October 26) (electronic version prior to print publication)

(doi: 10.1177/0962280214556794)

**Study Findings**

This study proposes a new method based on multiple imputation\* for analysis involving incomplete smoking history data, a situation that could be problematic when analyzing cancer incidence data in the Life Span Study of atomic bomb survivors. Applying the proposed approach to the LSS lung cancer incidence analysis, we observed the radiation-associated lung cancer risk to be larger among moderate smokers than heavy smokers. The smoking-radiation interaction, which had been estimated to be unusually strong in the previous analysis using a naive approach (relatively simple methodology to analyze missing data) with missing data indicators, was slightly weaker in the current analysis, which seems to be more interpretable. The proposed approach is expected to reduce estimation bias that might be unavoidable in naive analyses while preserving efficiency by using all available information, which should lead to radiation risk estimation with improved precision.

\*A method to analyze data involving missing observations by averaging the estimation results across multiple imputed data sets.

**Explanation**

In studying the radiation effects on, for example, cancer incidence among atomic bomb survivors, it is important to account for effects of other important risk factors, such as smoking habit. However, data for such risk factors are frequently missing or only available incompletely. Generally, the choice of a method to analyze data involving missing data may have large impacts on the results of risk estimation. In particular, it is well known that a naive approach, such as the simple deletion of subjects with missing data, can lead to substantial bias in estimation. This study proposed a new approach using multiple imputation for analysis involving missing or incompletely observed smoking history data, applied it to analysis of the LSS lung cancer incidence, and compared the estimation results with those based on the simpler missing-data approaches.

(1) 1. Study purpose

To develop a new multiple imputation method to analyze data involving missing or incompletely observed time-dependent smoking history data and apply it to the LSS lung cancer incidence analysis.

2. Study methods

In analyzing the LSS lung cancer incidence during the period 1958–1999, about 40% of the 105,401 eligible subjects had no smoking data, and many of the others had smoking history known for only part of the time period. As a new approach of multiple imputation for incompletely observed smoking history data, prediction models for the age of smoking initiation and, given initiation, smoking intensity and cessation age, were estimated; based on which missing smoking data were predicted and imputed. Complete data sets including the imputed data were analyzed by multiple imputation.

3. Study results

In an analysis of the LSS lung cancer incidence, we confirmed that when one had both radiation exposure and smoking habit, the radiation-associated excess relative risk of lung cancer would be higher if he/she were a moderate smoker than a heavy smoker. The smoking-radiation interaction, which had been found to be unusually strong by a naive approach using indicators for missing data, was slightly weaker with the new approach, which seems to be more interpretable for the joint effect of radiation exposure and smoking habit. The proposed approach is particularly appealing in a large-scale cohort study such as the LSS, since it can reduce estimation bias that might be unavoidable in naive analyses while maintaining efficiency by retaining known information, which is expected to lead to radiation risk estimation with improved precision.

**The Radiation Effects Research Foundation** has studied A-bomb survivors and their offspring in Hiroshima and Nagasaki for more than 60 years. RERF's research achievements are considered the principal scientific basis for radiation risk assessment by the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) and for recommendations regarding radiation protection standards by the International Commission on Radiological Protection (ICRP). RERF expresses its profound gratitude to the A-bomb survivors and survivors' offspring for their cooperation in our studies.

<sup>§</sup>*Statistical Methods in Medical Research*, a highly ranked academic journal in the field of medical informatics, carries original papers and commentaries on a wide range of subjects related to study designs and statistical methods in epidemiology and clinical studies. (Impact factor in 2013: 2.957)