

Statistical Methods in Medical Research § 掲載論文**「生存時間解析での不完全な喫煙歴データの取り扱い」**

古川恭治、Dale Preston、三角宗近、Harry Cullings

“Handling Incomplete Smoking History Data in Survival Analysis”*Stat Methods in Med Res* 2014 (October 26) (冊子版出版前の電子版)

(doi: 10.1177/0962280214556794)

今回の調査で明らかになったこと

原爆被爆者の寿命調査集団 (LSS) におけるがん罹患率調査などで問題となる不完全な喫煙歴データを伴う解析に対し、新しい多重代入法^{*}による解析法を提案した。提案手法を用いた LSS 肺がん罹患率解析では、中程度の喫煙者の方が、重度喫煙者よりも、放射線に関連した肺がんリスクが高くなる傾向を確認した。また、今回の解析では、従来の欠測指標 (欠測データを扱う上で比較的単純な方法) を用いた単純な解析で見られたような、非常に強い喫煙と放射線の交互作用はやや弱まり、より妥当と思われる結果が得られた。提案手法は、既知の情報を用いることにより推定効率を維持する一方、従来の解析方法では避けられない推定バイアスを減少させ、より精度の高い放射線リスク推定を行うことが期待できる。

(注) 多重代入法とは、欠測値を含むデータ解析手法のひとつで、欠測値に値を複数回代入し、結果を平均化する手法である。

解 説

原爆被爆者のがん罹患率調査などでは、放射線影響を調べる上で、喫煙など他の重要なリスク因子の影響も考慮して解析することは重要である。そのようなリスク因子の情報は、データが欠測したり、不完全に得られることが多い。一般に欠測データをどのように扱うかはリスク推定結果に大きな影響を及ぼす。特に、欠測している対象者を外すなどの単純な解析は、推定結果に大きなバイアスをもたらすことが知られている。本研究は、LSS 肺がん罹患率調査において、欠測したり不完全に得られる喫煙歴データを伴う解析方法として、多重代入法を用いた新しい方法を提案し、従来の欠測値に対する単純な解析から得られた結果と比較した。

1. 調査の目的

欠測したり不完全に得られる時間依存の喫煙データに対して、多重代入法を用いた新しい解析方法を提案し、それを用いて LSS 肺がん罹患率解析を行う。

2. 調査の方法

1958年から1999年の期間のLSS肺がん罹患率において、対象者105,401人の約40%

に対し喫煙情報が欠測しており、また、その他の対象者の多くに対しても、不完全な喫煙歴が得られた。不完全に観測される喫煙歴に対する新しい多重代入法として、喫煙開始年齢、さらに、喫煙を開始した場合、喫煙頻度と喫煙中止年齢の欠測値を予測するモデルを推定し、それらのモデルに基づいて、欠測している個人喫煙歴を予測した。予測データを含む完全な喫煙データを、多重代入法によって解析した。

3. 調査の結果

LSS 肺がん罹患率データの解析において、放射線被曝と喫煙の両方が起こった場合には、中程度の喫煙者の方が、重度喫煙者よりも、放射線関連の肺がん過剰相対リスクが高くなる傾向が確認できた。従来の欠測指標を用いた単純な解析で見られた、非常に強い喫煙と放射線の交互作用はやや弱まり、同時影響について、より妥当な解釈ができる結果となった。提案する方法は、欠測する喫煙歴のような時間依存データを含む解析において、既知の情報をできるだけ多く用いることにより推定効率を維持する一方、単純な解析では避けられない推定バイアスを減少させることが期待でき、LSS のような大規模コホート調査で特に有益である。

放射線影響研究所は、広島・長崎の原爆被爆者および被爆二世を 60 年以上にわたり調査してきた。その研究成果は、国連原子放射線影響科学委員会（UNSCEAR）の放射線リスク評価や国際放射線防護委員会（ICRP）の放射線防護基準に関する勧告の主要な科学的根拠とされている。被爆者および被爆二世の調査協力に深甚なる謝意を表明する。

§ *Statistical Methods in Medical Research* 誌は、医療情報科学分野の主要な学術誌のひとつであり、疫学、臨床研究における研究デザイン・統計解析方法に関連する幅広い課題の原著および総説を査読を経て掲載している。（2013 年のインパクト・ファクター：2.957）