令和７年度厚生労働科学研究費補助金（地域医療基盤開発推進研究事業）
分担研究報告書

【分担研究７】匿名加工処理および仮名加工処理に伴う情報欠損を評価する研究

藤井　進・中村　直毅・宮本　太郎・川本　章太・横田　崇・園部　真也

【要旨】
匿名加工医療情報が、研究目的に十分利用可能な信頼性と妥当性を有しているか、ソースデータおよびオプトアウトデータと比較した。当該情報は、母集団代表性、統計的整合性のいずれも満たしており、研究利用における信頼性と妥当性を有することが実証された。

【目的】
東北大学病院において作成された匿名加工医療情報が、研究目的に十分利用可能な信頼性と妥当性を有しているか、ソースデータおよびオプトアウトデータと比較し、統計的に検証する。

【方法】
以下の定義を用いた。ソースデータ：電子カルテに記録されたすべての患者情報。オプトアウトデータ：次世代法に基づき通知を受けた患者の情報。匿名加工データ：そのうち提供拒否を示した患者を除外し、非識別化および統計的丸め処理を施した情報。2022 年 4 月～2023 年 3 月に収集された DPC 関連データを用い、①ソースデータ、②オプトアウトデータ、③匿名加工データの 3 群において、主傷病名（ICD10 分類）、処方、年齢階層、会計情報の 4 カテゴリで上位 10 項目を抽出し、構成比の差をχ^2検定（有意水準 $p<0.05$）で評価した。各群の対象件数は 10 万件を超えており、十分な検出力を有していた。
観察研究であり、オプトアウト方式で実施した。

【結果】
主傷病名では、COVID-19（U071）が最多で、ソース 288 名、オプトアウト 111 名、匿名 111 名と完全に一致し、前立腺癌（C61）は 199→162→162 名、大腸ポリープ（K635）は 175→159→159 名と、上位 10 疾患全体でのχ^2=17.8、p=0.274 と統計的有意差はなかった。
処方では点滴注射（8,896→6,302→6,302 件）、経口抗菌薬（3,104→2,205→2,205 件）、胃酸分泌抑制薬（2,895→2,116→2,116 件）など、すべてオプトアウトと匿名加工で一致し、χ^2=26.26、p=0.094 と差は認められなかった。
会計情報では外来再診（70,149→52,942→52,942 件）、検査料（24,176→18,333→18,333 件）、初診料（10,204→7,735→7,735 件）等においても一致し、χ^2=0.12、p=0.73 と統計的な差異はなかった。
年齢階層においても、65～74 歳（25.1%→24.8%→24.9%）、85 歳以上（8.4%→8.2%→8.3%）など、全項目で 0.2%以内の差に収まっており、匿名加工処理による構成比の変動は認められなかった。

【考察】
匿名加工処理後のデータはオプトアウト除去後の構造を忠実に保持しており、ソースデータとの統計的な乖離もきわめて小さく、研究目的での利活用に支障はないと考えられる。主要カテゴリにおける構成の安定性と一致性は高く、バイアス導入の兆候はみられなかった。また、加工処理手順は定型化可能であり、多施設への展開も現実的である。

【結論】
東北大学病院で作成された匿名加工医療情報は、母集団代表性、統計的整合性のいずれも満たしており、研究利用における信頼性と妥当性を有することが示された。本研究は、制度の信頼性を高め、次世代医療基盤法に基づくデータの利活用促進に資する重要なエビデンスとなる。

【論文発表】
該当なし

【学会発表】
藤井進，中村直毅，川本章太，野中小百合．横田崇，園部真也．次世代医療基盤法に基づく匿名加工データの母集団代表性と研究利用の妥当性に関する検証．第 45 回医療情報学連合大会．

【知的財産権の出願・登録状況】
該当なし