

疫学調査集団への地域がん登録データの活用

－標準DBSによる原爆被爆者集団と広島市・広島県地域がん登録のデータリンケージ－

杉山 裕美

(公財)放射線影響研究所疫学部 腫瘍組織登録室室長代理

放射線影響研究所(放影研)では、広島・長崎における原爆被爆者集団(約12万人)、胎内被爆者集団(約3,600人)、原爆被爆者の子供の集団(約76,000人)、約20万人について、1958年から原爆放射線ががん罹患に与える影響の研究を行っています。これらの主要集団におけるがん罹患データは、広島市、広島県および長崎県へがん登録データの資料利用申請を毎年行い、許可を得て取得しています。実際の作業は、広島、長崎とも地域がん登録中央登録室業務を放影研に委託されていますので、以前はどちらも放影研で開発したシステムを用いて個人照合およびがん罹患データ抽出を行っていました。広島県地域がん登録(2002年開始)では、2006年に標準データベースシステム(標準DBS)を導入したことにより、2010年に主要集団とがん登録データとの個人照合を標準DBS内で行う方式に切り替えました。そこで標準DBSを用いて主要集団と地域がん登録データとの照合を行った経験についてご紹介いたします。

(1) 主要集団データの確定

がん登録データと照合するための、放影研標本集団データ(広島で1999年1月1日時点で生存)を作成しました。データ項目は個人照合(同定)に必須の姓、名、性別、生年月日、および照合の参考にするために、住所、放影研内個人番号、死亡日としました。たとえば名前が2つ登録されているなど、1人でこれらのデータが複数ある場合があるので、複数の組み合わせを作成し、85,415件(82,189人)のデータセット(CSVファイル)を作成しました。

(2) 標準DBSの“外部データとの照合機能”を用いての照合

- ①システム設定:Z(照合のみ)を選択。
- ②同定リストパターンの選択:放影研では、同定リスト制限「0:すべてのデータをリストする」と、生年月日一致条件「1:生年月月までを一致とする」を選択しました。
- ③データ定義とインポート:ここで放影研個人ID番号が同定リストの病院コード欄に表示されるように設定しました。こうすることで、同定時に原票データに戻りたいときにすぐに戻ることが可能となります。そしてデータ件数が多かったため、20個のデータセット(CSVファイル)に分けて、標準DBSへインポートしました。
- ④個人照合:標準DBSでの通常の登録票の同定作業と同様です。

- ⑤確定済みデータのエクспорт、と進めました。詳細は標準DBS操作手順書をご参照ください。

実際の個人照合作業としては、約85,000件のデータを、当時標準DBSに登録されていたがん登録データ約16万人と照合し、結果として6,528人について、がん登録データからがん罹患情報を得ることができました。作業にかかった時間は、インポート開始からデータエクспорт作業終了まで5人で約2ヶ月を要しました。2回目の2012年は、主要集団で既に標準DBSに既登録のものは除外できるので、若干数が減って照合するデータは約82,000件(約80,000人)となり、がん登録データ約23万人と照合し、約2,100人の新規がん罹患情報が得られました。作業時間は、2010年同様5人で約2ヶ月を要しました。作業量は実際にリンケージできる数よりも、照合するデータ件数に依存しますので、大量データの個人照合はそれなりの覚悟が必要です。

このように放影研主要集団のような大きい標本集団であると、かかる手間も多いのですが、標準DBSを用いて他の集団とがん登録データを照合し、がん罹患データを得ることは可能です。例えば、検診受診者のその後のがん罹患情報を確認することや、医療機関で他の疾患でフォローしていた患者のがん罹患情報を確認することなども可能です。今後は様々な地域でがん登録データが臨床や疫学データへ活用されることを願います。

