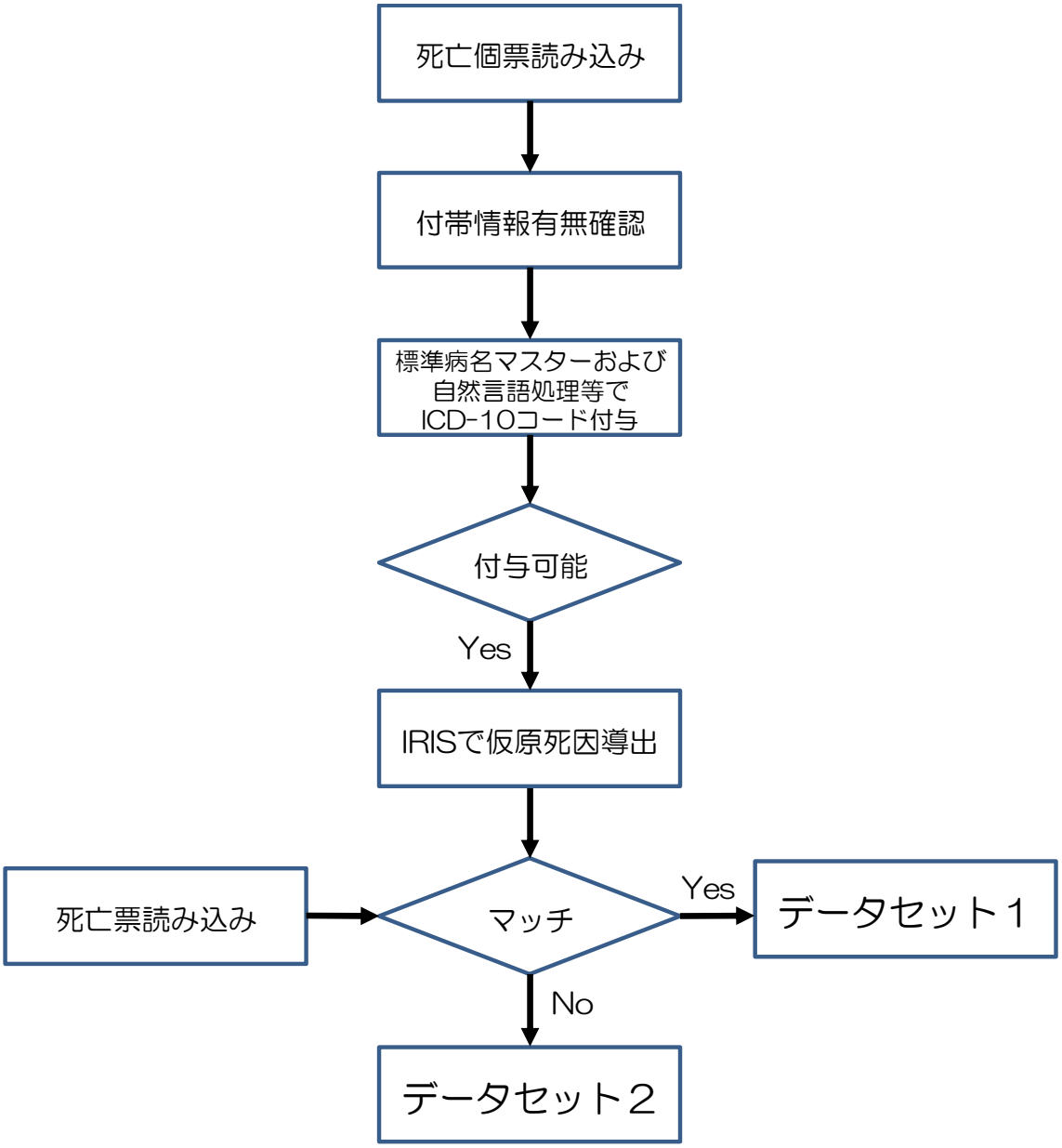
別添資料４　死亡票・死亡個票の調査票情報利用申請における集計方法

本研究で行う統計的研究の流れ図は以下の通りである



対象とする死亡個票は、「死亡の原因」に含まれる I 欄・II 欄病名の全てについて標準病名マスターおよびこれを用いた自然言語処理により ICD-10 コードを付与できたものとする。

そして平成 27 年度～30 年度の死亡票・死亡個票に対し、上記の ICD-10 コードを用い、欧州で利用されている原死因確定ツール IRIS を用いて導出した原死因と、死亡票に含まれ

る確定原死因コードを比較し、合致したものをデータセット１、合致しなかったものをデータセット２とする。また、これらのデータセットを機械学習モデル開発のための訓練セットとテスト用セットに分割する。

IRIS が導出した原死因コードと合致するか否かには、「死亡の原因」に含まれるⅠ欄・II 欄病名以外の情報の影響が考えられる。これは従来人手チェックにて行われており、これを支援するための機械学習モデルを開発する。どのような情報が原死因の変更の有無に寄与するかを学習するため、データセット１およびデータセット２に含まれる以下の情報を利用する：　死亡届出地の都道府県と市町村、死亡したところの都道府県と市町村、死亡したところの種別、「死亡の原因」に含まれるⅠ欄・II 欄病名ならびにその原死因、外因符号と期間、性別、生年月日時分、死亡した年月日時分、手術の部位及び所見、手術年月日、解剖の部位及び所見、死因の種類、障害が発生した年月日時分、傷害が発生したところの種別、傷害が発生したところ、その他の記述、傷害発生場所、手段及び状況、単胎・多胎の別、多胎の子数、多胎の出産順位、妊娠週数、妊娠・分娩時における母体の病態又は異状の有無と詳細、母側病態、母の生年月日、前回までの妊娠結果、その他付言すべき事柄、備考欄。

死亡票と死亡個票は、以下の項目を用いて結合する。死亡個票：処理年月届出地（都道府県・保健所・支所符号・市区町村（種類）・市区町村（順位））・事件簿番号、死亡票：調査年・提出年月・届出地・事件簿番号。

その後、開発したアルゴリズムをテストセットデータに適用し、原死因を導出する。その結果と死亡票の確定原死因・外因符号の結果を照らし合わせ、精度検証と課題確認、またアルゴリズムの改善を繰り返し行う。