

変数名	属性	内容	定義の相違	凡例など	備考
v1605	数値	子宮がん検診受診場所(保健所・保健センター)	一部相違	1: Yes (欠損=)	女性のみ定義。大阪東成は「保健所の検診車」。宮城と愛知は欠損。
v1606	数値	子宮がん検診受診場所(職場)	一部相違	1: Yes (欠損=)	v1600=1かつ当該場所の受診有無・受診年いづれか回答あり女性のみ定義。v1600=1かつ当該場所の受診有無・受診年いづれか回答ありなら1。
v1607	数値	子宮がん検診受診場所(病院など)	一部相違	1: Yes (欠損=)	女性のみ定義。v1600=1かつ当該場所の受診有無・受診年いづれか回答ありなら1。
v1608	数値	子宮がん検診受診場所(東成母子会自己採取)	一部相違	1: Yes (欠損=)	女性のみ定義。大阪東成のみ。
v1609	数値	子宮がん検診受診場所(その他)	一部相違	1: Yes (欠損=)	v1600=1かつ当該場所の受診有無・受診年いづれか回答あり女性のみ定義。v1600=1かつ当該場所の受診有無・受診年いづれか回答ありなら1。
v1610	数値	初経年齢	共通	(欠損=)	女性のみ定義。大阪は欠損。
v1611	数値	現在月経有り	共通	1: Yes, 0: No (欠損=)	女性のみ定義。大阪は欠損。
v1612	数値	自然閉経年齢	共通	(欠損=)	女性のみ定義。大阪は欠損。V1611に関わらず定義。宮城はv1613との重複回答あり。
v1613	数値	手術閉経年齢	共通	(欠損=)	女性のみ定義。大阪は欠損。V1611に関わらず定義。宮城はv1613との重複回答あり。
v1614	数値	出産経験あるか	共通	1: Yes, 0: No (欠損=)	女性のみ定義。
v1615	数値	出産人数	共通	(欠損=)	愛知都市地区は欠損。女性かつ「出産経験なし以外」は定義(出産経験不明例の出産人数ママ、出産あり例の0人ママ)。女性かつ「出産経験なし」は0人。
v1616	数値	初産年齢	共通	(欠損=)	愛知都市地区は欠損。女性かつ「出産経験なし以外」は定義(出産経験不明例の出産人数ママ、出産あり例の0人ママ)。女性かつ「出産経験なし」は0人。
v1700	数値	健康保険カテゴリ	一部相違	1: 国民健康保険, 2: 健康保険(政府管掌), 3: 健康保険(組合等), 4: 共済, 5: その他の保険, 6: 健康保険(政府管掌・組合), 9: 不明 (欠損=)	選択肢8は大阪府対照地区のみ(2とは欠損)。
v1701	数値	健康保険本人・家族別	共通	1: 本人・世帯主, 2: 家族 (欠損=)	
v1800	数値	ご飯1日何杯	共通	(欠損=)	宮城と愛知は2杯。大阪は1杯。
v1801	数値	パン摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1802	数値	肉摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1803	数値	魚類摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1804	数値	卵摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1805	数値	牛乳摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1806	数値	緑黄色野菜摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1807	数値	その他の野菜摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1808	数値	果物摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1809	数値	みそ汁摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1810	数値	漬物摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1811	数値	インスタント食品摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	宮城は欠損。
v1900	数値	緑茶摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上 (欠損=)	大阪能勢は「1.ほとんど飲まない」
v1901	数値	紅茶摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上 (欠損=)	大阪能勢は「1.ほとんど飲まない」
v1902	数値	コーヒー(非インスタント)摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上 (欠損=)	大阪能勢は「1.ほとんど飲まない」 宮城は欠損。
v1903	数値	インスタントコーヒー摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上 (欠損=)	大阪能勢は「1.ほとんど飲まない」 宮城は欠損。
v1904	数値	コーヒー(豆またはインスタント)摂取頻度カテゴリ	一部相違	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上 (欠損=)	宮城のみオリジナル変数。 他はv1902とv1903のOR結合(摂取多い選択肢優先で、選択肢1はAND結合)で定義。 大阪能勢は選択肢4はなし
v2000	数値	飲酒習慣カテゴリ	一部相違	1: ほとんど毎日, 2: ときどき, 3: 集まりのときだけ, 4: 以前はよく飲んだが最近飲まない, 5: 以前から飲まない (欠損=)	大阪能勢は選択肢4はなし
v2100	数値	喫煙習慣カテゴリ	一部相違	1: 吸っている, 2: やめた, 3: 吸わない, 4: 時々吸っている (欠損=)	選択肢4は大阪能勢のみ。
v2101	数値	喫煙開始年齢 連続変数	共通	(欠損=)	宮城、大阪東成、大阪能勢は生涯喫煙以外の回答ママ。 年齢との比較など論理チェックせず(10未満あり)。 99はママ(愛知)。
v2102	数値	喫煙本数/日 連続変数	共通	(欠損=)	99はママ(愛知、大阪)。
v2103	数値	禁煙年齢 連続変数	共通	(欠損=)	宮城、大阪東成、大阪能勢は生涯喫煙以外の回答ママ。 年齢との比較など論理チェックせず。 宮城、大阪東成、大阪能勢は過去喫煙以外の回答ママ。 99はママ(愛知)。
v2104	数値	吸い込むか	共通	1: Yes, 0: No (欠損=)	宮城、大阪東成、大阪能勢、大阪能勢は生涯喫煙以外の回答ママ。
v2200	数値	父親の既往 脳卒中	共通	1: Yes (欠損=)	
v2201	数値	父親の既往 心臓病	共通	1: Yes (欠損=)	
v2202	数値	父親の既往 がん	共通	1: Yes (欠損=)	
v2203	数値	父親の既往 上記3つなし	共通	1: Yes (欠損=)	宮城、大阪能勢はv2200, v2201, v2202との矛盾ママ。
v2204	数値	父親の既往 知らない	共通	1: Yes (欠損=)	大阪能勢は欠損。宮城、大阪東成、大阪能勢はv2200, v2201, v2202との矛盾ママ。
v2205	数値	父親の既往 がん 食道	共通	1: Yes (欠損=)	宮城、大阪東成、大阪能勢はv2202との矛盾ママ。
v2206	数値	父親の既往 がん 胃	共通	1: Yes (欠損=)	宮城、大阪東成、大阪能勢はv2202との矛盾ママ。
v2207	数値	父親の既往 がん 大腸	共通	1: Yes (欠損=)	大阪能勢は欠損。宮城、大阪東成、大阪能勢はv2202との矛盾ママ。
v2208	数値	父親の既往 がん 肝臓	共通	1: Yes (欠損=)	大阪東成、宮城、大阪能勢はv2202との矛盾ママ。
v2209	数値	父親の既往 がん 肺	共通	1: Yes (欠損=)	大阪東成、宮城、大阪能勢はv2202との矛盾ママ。
v2210	数値	父親の既往 がん 膀胱	共通	1: Yes (欠損=)	大阪東成、宮城、大阪能勢はv2202との矛盾ママ。
v2211	数値	父親の既往 がん 白血球	共通	1: Yes (欠損=)	大阪東成、宮城、大阪能勢はv2202との矛盾ママ。
v2212	数値	父親の既往 がん その他	共通	1: Yes (欠損=)	大阪東成、宮城、大阪能勢はv2202との矛盾ママ。
v2300	数値	母親の既往 脳卒中	共通	1: Yes (欠損=)	
v2301	数値	母親の既往 心臓病	共通	1: Yes (欠損=)	
v2302	数値	母親の既往 がん	共通	1: Yes (欠損=)	
v2303	数値	母親の既往 上記3つなし	共通	1: Yes (欠損=)	宮城、大阪能勢はv2300, v2301, v2302との矛盾ママ。
v2304	数値	母親の既往 知らない	共通	1: Yes (欠損=)	大阪能勢は欠損。宮城、大阪東成、大阪能勢はv2300, v2301, v2302との矛盾ママ。
v2305	数値	母親の既往 がん 食道	共通	1: Yes (欠損=)	宮城、大阪東成、大阪能勢はv2302との矛盾ママ。
v2306	数値	母親の既往 がん 胃	共通	1: Yes (欠損=)	宮城、大阪東成、大阪能勢はv2302との矛盾ママ。
v2307	数値	母親の既往 がん 大腸	共通	1: Yes (欠損=)	大阪能勢は欠損。宮城、大阪東成、大阪能勢はv2302との矛盾ママ。
v2308	数値	母親の既往 がん 肝臓	共通	1: Yes (欠損=)	宮城、大阪東成、大阪能勢はv2302との矛盾ママ。
v2309	数値	母親の既往 がん 肺	共通	1: Yes (欠損=)	宮城、大阪東成、大阪能勢はv2302との矛盾ママ。
v2310	数値	母親の既往 がん 膀胱	共通	1: Yes (欠損=)	大阪能勢は欠損。宮城、大阪東成、大阪能勢はv2302との矛盾ママ。
v2311	数値	母親の既往 がん 白血球	共通	1: Yes (欠損=)	宮城、大阪東成、大阪能勢はv2302との矛盾ママ。
v2312	数値	母親の既往 がん 子宮	共通	1: Yes (欠損=)	宮城、大阪東成、大阪能勢はv2302との矛盾ママ。
v2313	数値	母親の既往 がん 乳房	共通	1: Yes (欠損=)	宮城、大阪東成、大阪能勢はv2302との矛盾ママ。
v2314	数値	母親の既往 がん その他	共通	1: Yes (欠損=)	宮城、大阪東成、大阪能勢はv2302との矛盾ママ。
v2400	数値	小学生の頃の父親喫煙有無	一部相違	1: 吸っている, 2: やめた, 3: 吸わない, 4: 知らない (欠損=)	大阪能勢は時期を特定せず「お父さん(実父)はかつておられましたか」。 選択肢2は能勢のみ。愛知は受領シート選択肢1(吸う)2(吸わない)0だったのでそれぞれ1.3.(欠損)に分配。
v2401	数値	小学生の頃の母親喫煙有無	一部相違	1: 吸っている, 2: やめた, 3: 吸わない, 4: 知らない (欠損=)	大阪能勢は時期を特定せず「お母さん(実母)はかつておられましたか」。 選択肢2は能勢のみ。愛知は受領シート選択肢1(吸う)2(吸わない)0だったのでそれぞれ1.3.(欠損)に分配。
v2500	数値	現在同居家族喫煙あり	共通	1: Yes, 0: No (欠損=)	
v2501	数値	現在夫喫煙あり	共通	1: Yes (欠損=)	女性のみ定義。宮城、大阪東成、大阪能勢はv2500=1以外でも回答あり。

変数名	属性	内容	定義の相違	凡例など	備考
v2502	数値	現在喫煙あり	共通	1: Yes (欠損=)	男性のみ定義。宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2503	数値	現在父喫煙あり	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2504	数値	現在母喫煙あり	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2505	数値	現在子喫煙あり	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2506	数値	現在その他喫煙あり	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2600	数値	暖房 なし	共通	1: Yes (欠損=)	
v2601	数値	暖房 電気(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2602	数値	暖房 エアコン(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2603	数値	暖房 クリーンヒーター(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2604	数値	暖房 煙突付きストーブ(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2605	数値	暖房 煙突なしストーブ(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2606	数値	暖房 炭(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2700	数値	住居が/スの通る道路に面しているか	共通	1: Yes, 0: No (欠損=)	
v2800	数値	生まれてからいまのところに住んでいるか	共通	1: 生まれてから住んでいる, 0: 他の市区町村から引っ越し(欠損=)	
v2801	数値	転入何年前か	一部相違	(欠損=)	変知は「在住年数(計算)」として受領データセットにあり。年齢との比較など論理チェックせず0以上ならママ。大阪は欠損。
v2802	数値	転入年の元号	共通	1: 明治, 2: 大正, 3: 昭和(欠損=)	宮城は欠損。v2800,v2803との整合性チェックせずママ。
v2803	数値	転入年(和暦)	共通	(欠損=)	宮城は欠損。v2800,v2802との整合性チェックせず年数1以上ならママ。
v2804	数値	転入年(西暦)	共通	(欠損=)	宮城は欠損。v2802,v2803から単純計算(実在元年年かつ調査年以前のケースのみ)。
v2901	数値	職業 病院・診療所	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の下位選択肢(場所→病院・診療所・歯科診療所・歯科技工所)。
v2902	数値	職業 美容院・理髪店	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の下位選択肢(場所→美容院・理髪店)。
v2903	数値	職業 紡績・織布・タオル工場	一部相違	1: Yes (欠損=)	大阪熊取は「木綿・亜麻・大麻の工場」。宮城は2段階質問形式の下位選択肢(物じん→織物)。
v2904	数値	職業 出版・印刷工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場→印刷工場)。
v2905	数値	職業 メッキ作業所	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の下位選択肢(現場→メッキ作業所)。
v2906	数値	職業 運輸業務	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の第1問。
v2907	数値	職業 鉄鋼工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場→鉄鋼工場)。
v2908	数値	職業 金属製品工場	一部相違	1: Yes (欠損=)	大阪熊取は欠損。
v2909	数値	職業 鋳物工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(物じん→金属製品)。
v2910	数値	職業 陶磁器・ガラス工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(物じん→陶磁器・製造→ガラス製品)。
v2911	数値	職業 プラスチック加工	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(製造→プラスチック製)。
v2912	数値	職業 コム製品製造	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(製造→コム製品)。
v2913	数値	職業 耐熱・保温作業	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場→耐熱保温作業)。
v2914	数値	職業 造船所	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場→造船所)。
v2915	数値	職業 部品解体・再生工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場→部品解体・再生現場)。
v2916	数値	職業 建設・解体現場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場→建設・解体現場)。
v2917	数値	職業 石油精製・石油製品の製造	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場→石油精製所・製造→石油製品)。
v2918	数値	職業 ピッチ・タール・アスファルト製造	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(製造→ピッチ・タール・アスファルト)。
v2919	数値	職業 バッテリー・電池製造	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(製造→バッテリー・電池)。
v2920	数値	職業 セメント・生コン製品製造	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(扱う仕事→セメント・生コン)。
v2921	数値	職業 石綿工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(扱う仕事→石綿または石棉含有物質)。
v2922	数値	職業 ランドリー・洗濯工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(場所→ランドリークリーニング)。
v2923	数値	職業 土石採取・石材加工	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(物じん→石材・採土)。
v2924	数値	職業 自動車修理・整備	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の下位選択肢(現場→自動車修理作業所)。
v2925	数値	職業 写真現像所	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の下位選択肢(場所→写真現像所)。
v2926	数値	職業 炭灰・その他の鉱山、鉱石の精錬	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の第1問(鉱業)。
v2927	数値	職業 サービス業(公務員も含む)	共通	1: Yes (欠損=)	宮城、大阪熊取は欠損。
v2928	数値	職業 林業	一部相違	1: Yes (欠損=)	大阪東成、大阪熊取は欠損。宮城は2段階質問形式の第1問。
v2929	数値	職業 農業(畜産・果樹園芸も含む)	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の第1問。
v2930	数値	職業 皮革製品製造	一部相違	1: Yes (欠損=)	宮城、大阪熊取のみ。宮城は2段階質問形式の下位選択肢(製造→皮革製品)。
v2931	数値	職業 殺虫剤・枯草剤製造	一部相違	1: Yes (欠損=)	宮城、大阪熊取のみ。宮城は2段階質問形式の下位選択肢(製造→殺虫剤・枯草剤・殺菌剤)。
v2932	数値	職業 家具製造	一部相違	1: Yes (欠損=)	宮城、大阪熊取のみ。宮城は2段階質問形式の下位選択肢(製造→家具)。
v2933	数値	職業 卸売業・小売業	共通	1: Yes (欠損=)	大阪東成のみ。
v2934	数値	職業 有機溶剤	共通	1: Yes (欠損=)	宮城のみ。
v2935	数値	職業 放射線・放射性物質	共通	1: Yes (欠損=)	宮城のみ。
v2936	数値	職業 ベリリウム	共通	1: Yes (欠損=)	宮城のみ。
v2937	数値	職業 タルク(滑石)	共通	1: Yes (欠損=)	宮城のみ。
v2938	数値	職業 染料	共通	1: Yes (欠損=)	宮城のみ。
v2939	数値	職業 なし	共通	1: Yes (欠損=)	宮城欠損。v2901-v2932と矛盾ありママ(大阪熊取、変知)。
v2940	数値	最も長く就いた仕事	一部相違	(欠損=)	3桁(変知のみ2桁)コードそのまま添込(0, 99, 999ママ)。コードの出典は不明。
v2941	数値	最も長く就いた仕事	一部相違	(欠損=)	大阪熊取のみ。4桁コードそのまま添込。大阪東成はすべて0のため欠損に。コードの出典は不明。
v2950	数値	従軍年数	共通	(欠損=)	0以上はそのまま添込(0, 99ママ)。

表2 各変数の分類数とユニークセル数、カテゴリとグループについて

グループ	カテゴリ名	変数名	内容	分類数	ユニークセル数	
個人特性	性別	v0200	性別	2		
	年齢	v0300	追跡開始年齢(40-99)	60		
	居住地	v0402	地区通し番号	6		
追跡	日付	v0500	追跡開始日	7		
		v0501	10年コホート追跡終了日(地域ごとの定義)	7		
		v0502	10年観察終了日(個人ごとの定義)	12651	304	
	転帰	v0503	10年転帰	3		
	死因	v0600	死因ICD-9コード4桁	836	311	
自覚症状	自覚症状	v1000	自覚症状あるか 咳	2		
		v1001	自覚症状あるか 痰	2		
		v1002	自覚症状あるか 血痰	2		
		v1003	自覚症状あるか 動悸	2		
		v1004	自覚症状あるか 息切れ	2		
		v1005	自覚症状あるか 食欲不振	2		
		v1006	自覚症状あるか 便秘	2		
		v1007	自覚症状あるか 不眠	2		
		v1008	自覚症状あるか やせてきた	2		
		v1009	自覚症状あるか 疲れやすい	2		
		v1010	自覚症状あるか 喘息の発作	2		
	v1011	自覚症状あるか なし	4			
	既往	既往	v1100	既往 高血圧	4	
			v1101	既往 心臓病	4	
			v1102	既往 脳卒中	4	
			v1103	既往 糖尿病	4	
			v1104	既往 肝臓病	4	
			v1105	既往 腎臓病	4	
			v1106	既往 胃十二指腸潰瘍	4	
			v1107	既往 胆石	4	
			v1108	既往 虫垂炎	2	
			v1109	既往 結核ろく膜炎	4	
			v1110	既往 肺炎	4	
			v1111	既往 喘息	4	
			v1112	既往 慢性気管支炎	4	
			v1113	既往 肺気腫	4	
			v1114	既往 じん肺	4	
			v1115	既往 貧血	2	
			v1116	既往 アレルギー性鼻炎	4	
			v1117	既往 蕁麻疹	4	
			v1118	既往 湿疹・皮膚炎	4	
	v1119	既往 痛風	2			
	v1120	既往 なし	2			
	身長・体重	身長・体重	v1200	身長(cm)	489	97
			v1201	体重(kg)	433	126
	まるめの身長・体重	まるめの身長・体重	v1200	身長(cm)※小数点切り捨て	89	6
			v1201	体重(kg)※小数点以下切り捨て	100	19
	胸部レントゲン	胸部レントゲン	v1300	胸部レントゲン検査受診有無	3	
			v1301	胸部レントゲン検査受診有無(1年以内)	2	
			v1302	胸部レントゲン検査受診有無(1年より前~2年以内)	2	
			v1303	胸部レントゲン検査受診有無(2年より前)	2	
			v1304	胸部レントゲン検査受診場所(住民健診)	2	
			v1305	胸部レントゲン検査受診場所(保健所・保健センター)	2	
			v1306	胸部レントゲン検査受診場所(職場)	2	
			v1307	胸部レントゲン検査受診場所(病院など)	2	
			v1308	胸部レントゲン検査受診場所(農協)	2	
			v1309	胸部レントゲン検査受診場所(その他)	2	
	胃レントゲン	胃レントゲン	v1400	胃レントゲン検査受診有無	3	
			v1401	胃レントゲン検査受診有無(1年以内)	2	
			v1402	胃レントゲン検査受診有無(1年より前~2年以内)	2	
			v1403	胃レントゲン検査受診有無(2年より前)	2	
			v1404	胃レントゲン検査受診場所(住民健診)	2	
			v1405	胃レントゲン検査受診場所(保健所・保健センター)	2	
			v1406	胃レントゲン検査受診場所(職場)	2	
			v1407	胃レントゲン検査受診場所(病院など)	2	
			v1408	胃レントゲン検査受診場所(農協)	2	
			v1409	胃レントゲン検査受診場所(その他)	2	
	v1410	胃レントゲン検査受診回数(5年間)	16			
	健康診査または血圧検査受診有無	健康診査または血圧検査受診有無	v1500	健康診査または血圧検査受診有無	3	
			v1501	健康診査または血圧検査受診有無(1年以内)	2	
			v1502	健康診査または血圧検査受診有無(1年より前~2年以内)	2	
			v1503	健康診査または血圧検査受診有無(2年より前)	2	

グループ	カテゴリ名	変数名	内容	分類数	ユニークセル数
アンケート	健康診査	v1504	健康診査または血圧検査受診場所(住民健診)	2	
		v1505	健康診査または血圧検査受診場所(保健所・保健センター)	2	
		v1506	健康診査または血圧検査受診場所(職場)	2	
		v1507	健康診査または血圧検査受診場所(病院など)	2	
		v1508	健康診査または血圧検査受診場所(農協)	2	
		v1509	健康診査または血圧検査受診場所(その他)	2	
	子宮がん	v1600	子宮がん検診受診有無	3	
		v1601	子宮がん検診受診有無(昨年)	2	
		v1602	子宮がん検診受診有無(一昨年)	2	
		v1603	子宮がん検診受診有無(一昨年より前)	2	
		v1604	子宮がん検診受診場所(検診車・住民検診)	2	
		v1605	子宮がん検診受診場所(保健所・保健センター)	2	
		v1606	子宮がん検診受診場所(職場)	2	
		v1607	子宮がん検診受診場所(病院など)	2	
		v1608	子宮がん検診受診場所(東成母子会自己採取)	2	
		v1609	子宮がん検診受診場所(その他)	2	
	出産・月経	v1610	初経年齢	17	2
		v1611	現在月経有り	3	
		v1612	自然閉経年齢	51	4
		v1613	手術閉経年齢	44	6
		v1614	出産経験あるか	3	
		v1615	出産人数	19	3
		v1616	初産年齢	41	4
	健康保険	v1700	健康保険カテゴリ	8	
		v1701	健康保険本人・家族別	3	
	食習慣	v1800	ご飯1日何杯	16	
		v1801	パン摂取頻度カテゴリ	6	
		v1802	肉摂取頻度カテゴリ	6	
		v1803	魚類摂取頻度カテゴリ	6	
		v1804	卵摂取頻度カテゴリ	6	
		v1805	牛乳摂取頻度カテゴリ	6	
		v1806	緑黄色野菜摂取頻度カテゴリ	6	
		v1807	その他の野菜摂取頻度カテゴリ	6	
		v1808	果物摂取頻度カテゴリ	6	
		v1809	みそ汁摂取頻度カテゴリ	6	
		v1810	漬物摂取頻度カテゴリ	6	
		v1811	インスタント食品摂取頻度カテゴリ	6	
	お茶・コーヒー	v1900	緑茶摂取頻度カテゴリ	6	
		v1901	紅茶摂取頻度カテゴリ	6	
		v1902	コーヒー(非インスタント)摂取頻度カテゴリ	6	
		v1903	インスタントコーヒー摂取頻度カテゴリ	6	
		v1904	コーヒー(豆またはインスタント)摂取頻度カテゴリ	6	
	飲酒	v2000	飲酒習慣カテゴリ	6	
	喫煙	v2100	喫煙習慣カテゴリ	5	
		v2101	喫煙開始年齢	83	11
		v2102	喫煙本数/日	54	5
		v2103	禁煙年齢	76	3
		v2104	吸い込むか	3	
	父親の既往	v2200	父親の既往 脳卒中	2	
		v2201	父親の既往 心臓病	2	
v2202		父親の既往 がん	2		
v2203		父親の既往 上記3つなし	2		
v2204		父親の既往 知らない	2		
v2205		父親の既往 がん 食道	2		
v2206		父親の既往 がん 胃	2		
v2207		父親の既往 がん 大腸	2		
v2208		父親の既往 がん 肝臓	2		
v2209		父親の既往 がん 肺	2		
v2210		父親の既往 がん 膀胱	2		
v2211		父親の既往 がん 白血病	2		
v2212	父親の既往 がん その他	2			
母親の既往	v2300	母親の既往 脳卒中	2		
	v2301	母親の既往 心臓病	2		
	v2302	母親の既往 がん	2		
	v2303	母親の既往 上記3つなし	2		
	v2304	母親の既往 知らない	2		
	v2305	母親の既往 がん 食道	2		
	v2306	母親の既往 がん 胃	2		
	v2307	母親の既往 がん 大腸	2		
	v2308	母親の既往 がん 肝臓	2		
	v2309	母親の既往 がん 肺	2		
	v2310	母親の既往 がん 膀胱	2		

グループ	カテゴリ名	変数名	内容	分類数	ユニークセル数	
		v2311	母親の既往 がん 白血病	2		
		v2312	母親の既往 がん 子宮	2		
		v2313	母親の既往 がん 乳房	2		
		v2314	母親の既往 がん その他	2		
	周囲の喫煙	v2400	小学生の頃の父親喫煙有無	5		
		v2401	小学生の頃の母親喫煙有無	5		
		v2500	現在同居家族喫煙あり	3		
		v2501	現在夫喫煙あり	2		
		v2502	現在妻喫煙あり	2		
		v2503	現在父喫煙あり	2		
		v2504	現在母喫煙あり	2		
		v2505	現在子喫煙あり	2		
		v2506	現在その他喫煙あり	2		
		暖房	v2600	暖房 なし	2	
			v2601	暖房 電気(複数回答)	2	
			v2602	暖房 エアコン(複数回答)	2	
	v2603		暖房 クリーンヒーター(複数回答)	2		
	v2604		暖房 煙突付きストーブ(複数回答)	2		
	v2605		暖房 煙突なしストーブ(複数回答)	2		
	v2606		暖房 炭(複数回答)	2		
	住居	v2700	住居がバスの通る道路に面しているか	3		
	転居	v2800	生まれてからいまのところに住んでいるか	3		
		v2801	転入何年前か	101	5	
	職業	v2901	職業 病院・診療所	2		
		v2902	職業 美容院・理髪店	2		
		v2903	職業 紡績・綿布・タオル工場	2		
		v2904	職業 出版・印刷工場	2		
		v2905	職業 メッキ作業所	2		
		v2906	職業 運輸業務	2		
		v2907	職業 鉄鋼工場	2		
		v2908	職業 金属製品工場	2		
		v2909	職業 鋳物工場	2		
		v2910	職業 陶磁器・ガラス工場	2		
		v2911	職業 プラスチック加工	2		
		v2912	職業 ゴム製品製造	2		
		v2913	職業 耐熱・保温作業	2		
		v2914	職業 造船所	2		
		v2915	職業 部品解体・再生工場	2		
		v2916	職業 建設・解体現場	2		
		v2917	職業 石油精製・石油製品の製造	2		
		v2918	職業 ビッチ・タール・アスファルト製造	2		
		v2919	職業 バッテリー・電池製造	2		
		v2920	職業 セメント・生コン製品製造	2		
		v2921	職業 石綿工場	2		
		v2922	職業 ランドリー・洗濯工場	2		
		v2923	職業 土石採取・石材加工	2		
		v2924	職業 自動車修理・整備	2		
		v2925	職業 写真現像所	2		
		v2926	職業 炭坑・その他の鉱山、鉱石の精錬	2		
		v2927	職業 サービス業(公務員も含む)	2		
v2928		職業 林業	2			
v2929		職業 農業(畜産・果樹園芸も含む)	2			
v2930		職業 皮革製品製造	2			
v2931		職業 殺虫剤・枯草剤製造	2			
v2932	職業 家具製造	2				
v2933	職業 卸売業・小売業	2				
v2934	職業 有機溶剤	2				
v2935	職業 放射線・放射性物質	2				
v2936	職業 ベリリウム	2				
v2937	職業 タルク(滑石)	2				
v2938	職業 染料	2				
v2939	職業 なし	2				
v2940	最も長く就いた仕事	583	62			
v2950	従事年数	81	7			

表3 ベースライングループ(【個人特性】【追跡】)に対するまるめの処理について

個人特性	処理内容	分類数	ユニークセル数
1	処理なし(性別×年齢×居住地)	673	19
2	年齢5歳階級かつ85歳以上まるめ	120	0
追跡	処理内容	分類数	ユニークセル数
1	処理なし(日付:日まで×転帰×死因ICD-9コード4桁)	20176	16631
2	日付:月まで×転帰×死因ICD-9コード3桁	11398	7936
3	追跡期間(月)×転帰×死因ICD-9コード3桁	6225	3654
4	追跡期間(月)×転帰×(死因削除)	243	0

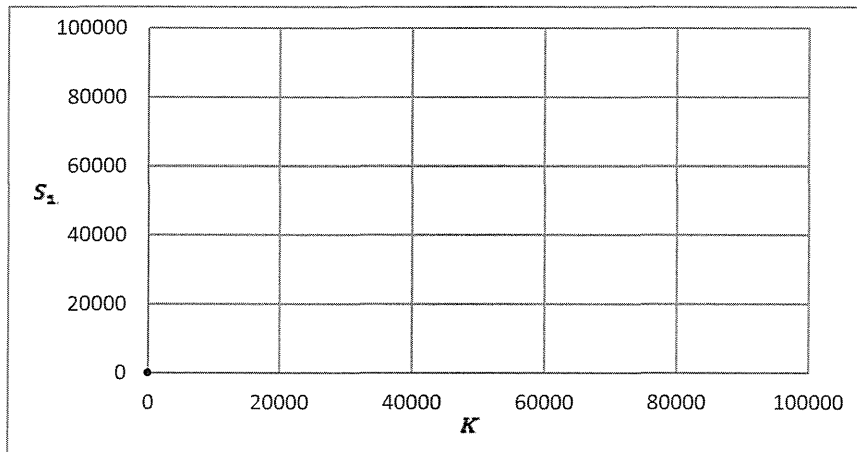


図1 分類数(K)とユニークセル数(S_1)の関係

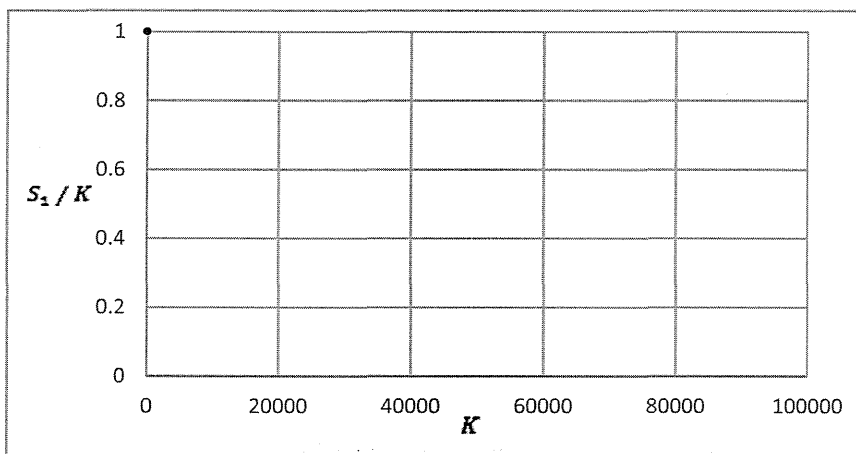


図2 分類数(K)と分類数に占めるユニークセル数の割合(S_1/K)の関係

図3 分類数(K)とユニークセル数(S_z)の関係(標本数別)

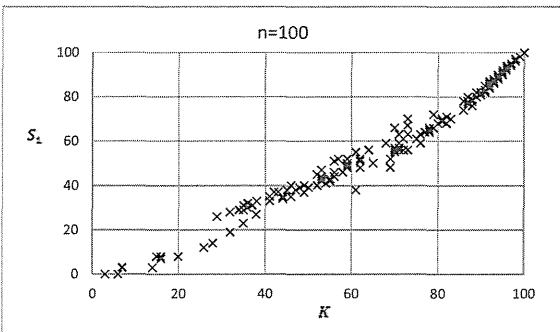
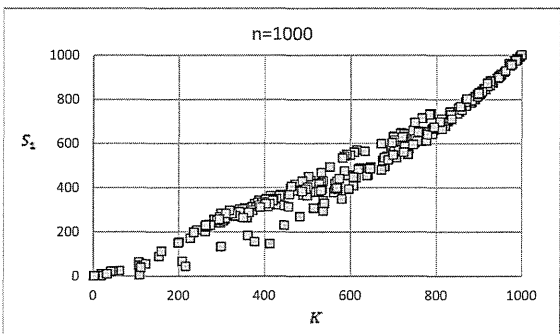
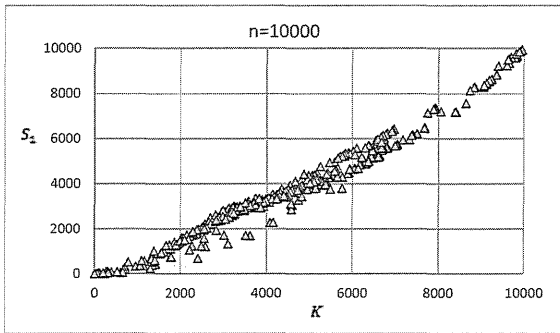
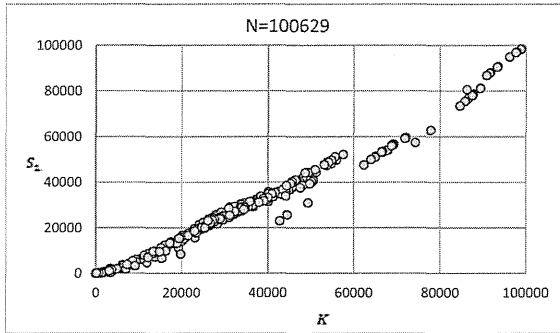
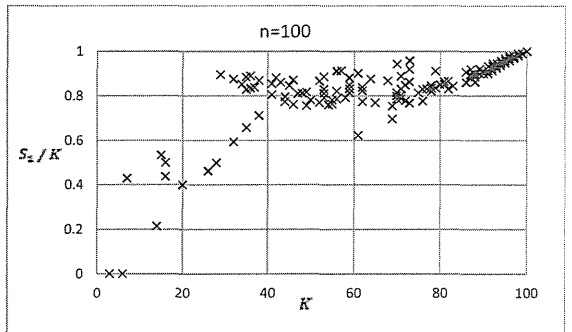
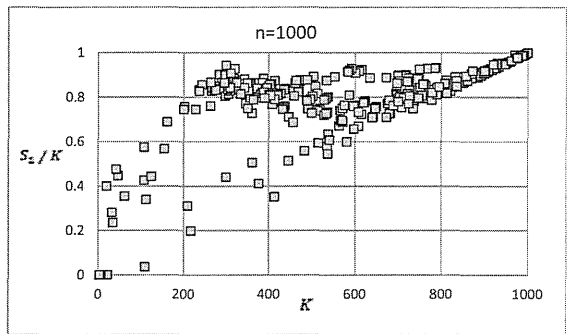
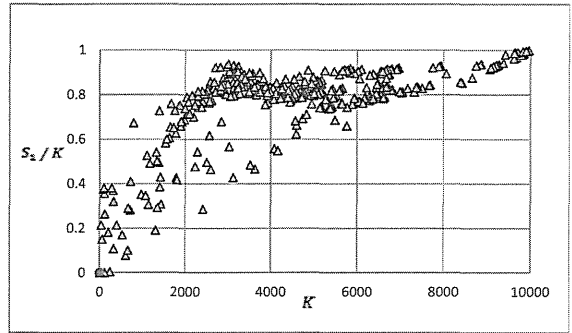
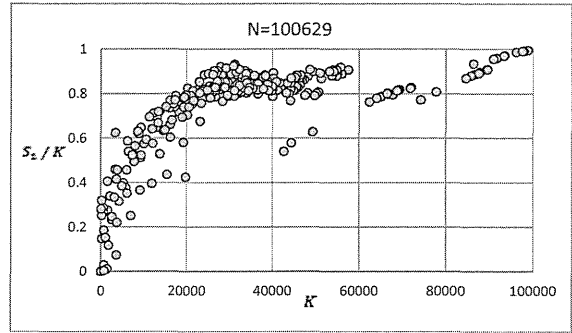


図4 分類数(K)と分類数に占めるユニークセル数の割合(S_z/K)の関係(標本数別)



疫学追跡終了後コホートデータの共通利用（アーカイブ化）の際の 死因データ利用に関する検討

研究分担者 大橋 靖雄 中央大学理工学部人間総合理工学学科生物統計学
研究協力者 原田亜紀子 東京大学大学院医学系研究科生物統計学

研究要旨

死因情報を付与した形でのデータアーカイブ化が難しい現状を鑑みて、現制度下での運用案を提案した。一つ目はアーカイブセンターなどに死因情報以外のデータを集約管理し、必要に応じて従来通りの死因照合作業を実施し、死因を付加したデータセットを作成し解析を行う方法である。二つ目は、データを保持者のもとに置いたまま、必要と判断した情報だけを選択的に共有させる分散型ネットワークによる方法である。提案1の運用案を複数のコホート研究から構成される国内循環器疫学研究(JALS)に適用したところ、死因照合率は99.7%であり、照合作業の技術的側面、作業手順化の面で問題はなく、運用案の一つになりうると考えられた。提案2については、医療(臨床)、疫学研究で先行している分散型ネットワークの事例を収集し、本検討事例である疫学追跡終了後コホートデータの共通利用の場面に応用可能な事例があるか検討を行った。疫学共同研究における分散型ネットワークを用いたデータ連携の例などから、リモート環境で様々な統計解析を柔軟に行えること、ローカルサイト(大学、研究所、医療機関を想定)へアクセスを行う上でのセキュリティ、さらには、人口動態統計の二次利用をローカルサイト単位で各々申請し、死因を付与したデータを統合して用いることや、施設外部のPC上のRAMへの一時的な書き出し(保存はされない、電子データの「一時的蓄積」)の可否など、法令解釈の問題にも留意する必要があると考えられた。疫学研究追跡終了後のコホートデータの共通利用に関して、提案した二つの方法で運用していくことは可能と考えられるが、実際に運用する場合には、前者においてはデータ利用の規約、データ利用の物理的環境、サポート体制、後者は、人口動態二次利用申請による死因データ利用の法令解釈などの課題があると考えられた。

A. 目的

米国では、National Heart, Lung, and Blood Institute (NHLBI) などの公的研究資金で実施した臨床試験や疫学観察研究のデータは、Biologic Specimen and Data Repositories Information Coordinating Center (BioLINCC) に対して追跡終了後一定の期間、条件を充足した後に研究データを登録し公開するよう進められている

(<https://biolincc.nhlbi.nih.gov>)。BioLINCCに収集されたデータは、使用申請を行うことにより、研究利用でき、研究データの二次利用により新たな研究や若手研究者の教育・研究機会が創出されている。翻って我が国の現状をみると、厚生労働科学研究費補助金を受けた研究では、National Bioscience Database Center (NBDC) 等へのデータ提供が求められ、バイオサイエンス基礎研究でのデータアーカイブ化には進展はみら

れるが、個人の健康情報を含む疫学観察（コホート）研究領域では、未だデータの二次利用環境は整っていない。このように疫学観察研究において、データの共有化、オープンデータ化が進まない背景には、コホート研究が様々な研究費で維持されており、米国 NHLBI などに代表される大型の研究費を受けた先へのデータ提供という単純な構図になりえないこと、多くのコホート研究では主要なアウトカムとなる死因情報を人口動態統計調査の二次利用申請によって得ており、死因を付与した状態でのデータ公開が難しい点、研究企画時点から、データアーカイブ化を見据えた倫理的諸問題の対応、調査票やデータベース構造の標準化の準備されていないなどが原因として考えられる。そこで本研究では、死因情報を付与した形でのデータアーカイブ化が難しい本邦の現状を鑑みて、現制度下で疫学追跡終了後コホートデータの共通利用を行うための運用案を提案する。

B. 研究方法

1. 死因情報を外したアーカイブ環境を想定し、必要時に死因を人口動態二次利用申請し、アーカイブセンター（データセンター）にて照合・集計・解析を行う運用

国内大規模循環器疫学研究（JALS）を例に、研究を統括する中央研究事務局が人口動態二次利用申請し、提案する運用例にならって死因を照合することで生じる問題点を考察し、提案例をアーカイブセンターで運用する際に想定される課題を検討した。

1) 死因データを必要時に連結させる方法の提案

JALS では、ローカルコホートから追跡調査データ（生存・死亡（死因はなし）、発症）の提供があり、死因の同定については中央事務局（データセンター）で人口動態の二次利用申請を行い一括で照合作業を行っている。この JALS で採用しているシステムを例にして、研究コンソーシアムに属する

研究者が、このデータを利用した研究を計画し、人口動態の二次利用申請で死因情報を得ることによって、センターのスタッフ（必ずしも専門知識を必要としない）により簡便に照合作業が行えるような環境を検討した。

2) 人口動態統計の二次利用申請

Japan Arteriosclerosis Longitudinal Study (JALS) は、国内 58 市町村、8 職域の計 11 万人を追跡している。この JALS コホートのうち、職域コホートを除き、研究開始（2002 年 ただし一部コホートは 1999 年）から 2012 年 12 月までの異動状況が、住民基本台帳（住民票）情報により確認されている追跡対象者で期間中に死亡した 7,137 件を死因照合の対象とした。総務省による「統計法第 33 条の運用に関するガイドライン」を参照した上で、厚生労働省大臣官房統計情報部企画課審査解析室に人口動態調査二次利用の申請を行った。申請作業と並行し研究事務局において 2012 年 12 月までの死亡例について、性別、生年月日、死亡日、死亡時の居住地（市町村コード）のリストを作成した。

3) 提案例に従った照合作業

人口動態調査の使用許可がおりた後、性別、生年月日、死亡年月日、死亡時の居住地（市町村コード）を照合変数とし、人口動態調査データを用い原死因を確定した。

2. 分散型ネットワーク国内外の医療（臨床）データ連携、疫学共同研究利用などの先行事例を収集し、疫学追跡終了後コホートデータの共通利用に活用できるかどうか検討を行う

C. 結果

1. 死因情報を外したアーカイブデータを利用する運用案

1) データ提供からアーカイブ化まで

提案するデータ利用基盤の概略を[図 1]に示し

た。研究コンソーシアムに参加する各研究が、基本データ（生活習慣、検査データなど）と死因を除いた追跡データをアーカイブセンターに提供する（①データ提供）。アーカイブセンターでは、基本データベースと追跡データベースを分けて構築しておき、基本データベースは原則登録時から修正なしの状態、追跡データベースは、今後の追跡継続に応じて更新できる構造とする（②アーカイブ化）。追跡データは、今後の死因照合作業で必要となる「死亡地（市町村）」、「死亡日」、「生年月日」、「性別」を含むよう設計した。

2) 死因情報の申請

コンソーシアム内の研究者（あるいは一定の条件を設け、研究グループ外の研究者も可能とするか）が、このデータベースを使用する研究を計画し、死因情報を得るために厚生労働省に対して人口動態調査二次利用申請を行うとした。（③死因情報申請）。

3) 照合作業から研究データセット作成

承認後に提供を受けた死因情報をアーカイブセンター内で、「死亡地（市町村）」、「死亡日」、「生年月日」、「性別」をキー変数として、保有する追跡情

報と照合し（④データ照合）、死因を付与した一時的な解析データセットを作成し、研究計画に基づいた解析に使用する（⑤解析用データセットの作成）。研究終了後は死因情報を削除（抹消）し、厚生労働省に利用後報告を行うという流れである（⑥利用報告、死因情報の削除）。死因を付与したデータをアーカイブせずに、研究を計画するたびに、この①から⑥の作業をアーカイブセンターで手順化し対応するという運用方法の提案である。

4) JALS を例にした上記作業の検証

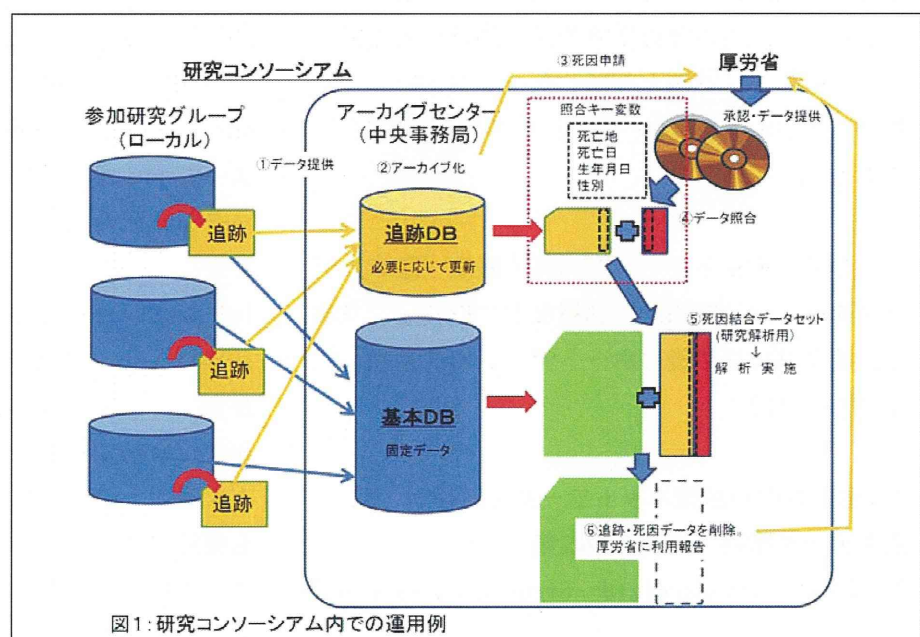
2015年1月29日付で利用許可がおり、JALS対象地域の市町村で1999年1月1日から2012年12月31日までに発生した死亡の調査票情報を受領した。

① 提供を受けた人口動態データの読み込み

提供を受けるデータの構造（提供変数）に応じて変更は必要となるが、データの構造と読み込み用のSASプログラムを作成した。

②研究全体の市町村（コード）をリストアップした一覧表を読み込み（SASプログラム）

③追跡調査データベースの対象データから死因照合を行いたい死亡者を抽出する作業（SAS プログ



ラム)

④照合作業 (SAS プログラム)

上記①と③で作成されたデータセットについて、「死亡地(市町村コード)」、「死亡日」、「生年月日」、「性別」でマッチマージしていく。JALS の対象者で、職域コホートと死亡調査データが確定していないコホートを除き、死亡が特定できていたのは 7,137 件であった。性別、生年月日、死亡年月日、死亡時の居住市町村名を照合変数とし、人口動態調査データと一致がみられたのは 7,099 件 (99.5%) であった。

⑤照合例、未照合例のリスト化 (SAS プログラム→Excel ファイル)

死因一致例の中に複数の候補例(一意に決まらない例)が存在する対象がいないか、また未照合例の情報確認を行うために該当者のリストアップを行う必要がある。これらを SAS から Excel ファイルに出力するプログラムを作成した。複数の候補例が見つかったのは 5 件あり、いずれも東日本大震災の被災地域での死亡者で、死因が同じであったためその死因で照合を行った。死因が照合できなかった例は 33 件であり、このうち 14 件については、以前に JALS が行った死因照合作業において既に未照合が判明しており、各コホートに対して死亡時情報を確認したがいずれも情報に誤りのなかった例で、人口動態統計作成の過程で入力間違い等が発生した事例と判断した。このため、今回照合出来なかった例は、実質として 19 件 (0.27%) であった。

2. 分散型ネットワーク国内外の医療(臨床)データ連携、疫学共同研究利用などの先行事例収集と、疫学追跡終了後コホートデータの共通利用についての検討

1) 国内外の分散型ネットワーク先行事例(医療・臨床データ連携(交換)の収集

①国内: Standardized Structured Medical

record Information eXchange (SS-MIX)

厚生労働省電子的診療情報交換推進事業(Standardized Structured Medical record Information eXchange)で策定された『電子的診療情報を他システムとの交換や地域医療連携で利用するために、診療情報を標準的な形式で蓄積・管理するデータとして保存できる領域』の仕様のことである。SS-MIX では、これらの医療情報を「標準化ストレージ」というツールに医療情報を標準化した形式で格納・蓄積することにより、複数ベンダー間・複数システム間の相互運用性を高めることを目的としている。標準化ストレージの構造は、コンピュータの一般的なファイル格納形式と同様に階層化されたフォルダのディレクトリー構造を用いている。電子カルテシステムなどではほとんどの業務で個々の患者を軸として診療情報が格納されているため、標準化ストレージにおいても患者にひも付く各種の情報をフォルダの階層構造にルールを決めて格納している。それ以外の項目について研究間での統合をはかるために、拡張ストレージを設け統合することで、標準化ストレージを核にし、施設間連携を構築し、診療情報の研究利用、地域医療連携、災害時連携などを目指している)。

②米国: Sentinel Initiative データ交換モデル

米国では 2008 年より、Food and Drug Administration Amendments Act (FDA 改革法)によって FDA 承認医薬品および医療製品の安全性モニタリングのため、既存のデータベースを用いたアクティブ・サーベイランスである Sentinel Initiative が実施されることとなった。Sentinel Initiative ではデータインフラと手法の開発を目的として、Harvard Pilgrim Health Care Institute が中心となって医療データベースを所有する数十の医療機関が協力して Mini-Sentinel も実施され、各データストレージは分散したままで、コモンデータモデルに従って共通プログラム

で解析するシステムが構築されている。

② 米国：Health Information Exchange (HIE)

CDX (Crossflo Data Exchange®) ソフトウェアは、Crossflo 社が開発したデータ交換ソフトウェアで、HL7、GJXDM (Global Justice XML Data Model)、NIEM (National Information Exchange Model)、EDXL、CAP、NCPDP など主要医療データシステムおよび国家的なデータ標準、業界標準に準拠したデータ共有方式を実装している。

本システムの医学領域の応用事例としては、モンタナ州および連邦政府の要請を受けて、国立医療情報科学センターの協力のもと、モンタナ州の4つの病院のED(救急科)と州保健福祉局(Montana Department of Public Health and Human Services)とをデータ接続するモンタナ州医療情報交換システムプロジェクト(Montana Health Information Exchange Pilot Project)があげられる。CDXによるシステムで、ほぼリアルタイムでの医療情報の交換を実現し、異種の医療データソースを迅速および効率的に接続する疾病サーベイランスシステムが構築されている。

2) 国際共同疫学研究等での分散型ネットワークによるデータ連携例

ViPAR:Virtual Pooling and Analysis of Research Data (Karter KM et al. *Int. J. Epidemiol.* 2015)

ViPARは、International Collaboration for Autism Registryにおいて、国際共同研究を実施する6つのサイトデータを統合的に解析することを目的に開発された(Karter KM et al. *Int. J. Epidemiol.* 2015)。多くの医学研究では、倫理的、法的な理由から保有するデータの管理と保持を各サイトにおいて行う必要があるが、本研究では、解析の目的で一時的に「Virtual pooling」サイトに移動することは許可されたことから、一か所にデータを持続的に収集することなしに、物

理的に離れたリモートサイトのデータを仮想的(一時的)に集約し分析を可能とするシステムが開発された。ViPARのリモートサイトとマスターサーバの関連の関連を図4に示したが、ViPARのマスターサーバは、リモートサイトのPCへ接続されており、研究データはそれぞれリモートサイトで管理されている。利用者はWeb上のanalytical portalからアクセスし、データをリモートPCからマスターサーバのRAM上に抽出し、仮想的にデータをプールして解析を行い、終了後は保存することなく消去する。このため外部にデータを収集することなく多施設データを統合した解析が可能となっている。

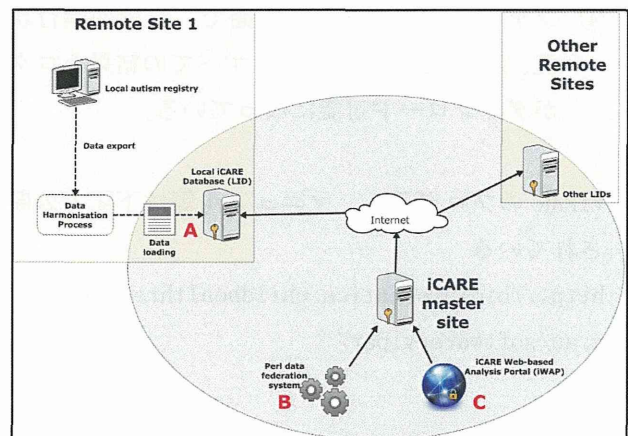


図2 ViPARのリモートサイトとマスターサーバの関連

(Karter KM et al. *Int. J. Epidemiol.* 2015)

ViPARの環境のうち、リモートサイトのデータベースは、データstorageとしてMySQLサーバを、マスターサイトとの通信のためSSHサーバが導入され構成されている(いずれもオープンソース)。実際に解析を行うマスターサーバは、ViPAR daemonとweb-basedのポータルで構成されている。ViPAR daemonは、ログ出力、統計パッケージのアクセスとコントロール、ローカルサーバから抽出したデータの統合作業を担う。web-basedのポータルは、解析、データマネージメントを行うインターフェイスであり、新規の解析スタート画面、解析アウトプット画面、コードのマネージメント

画面の三つからなる。ViPAR を用いた具体的な解析の流れは、以下の通りである。

- ① 解析用インターフェイスを開き変数やサイトを選択
- ② 統計パッケージの種類 (R, SAS, Stata) を選択し、テキスト入力部分に解析のための syntax を入力しサブミットする
- ③ リモートデータベースにデータのリクエストが送られ、マスターサーバの RAM に読み込まれ、バーチャルプールされる (これらのデータは保存されることはない)。バーチャルプールデータセットは、選択された統計パッケージに読み込まれ、解析が行われる
- ④ ファイルマネージャー画面で、解析の進行状況、完了状況が確認でき、すべての結果とログがダウンロード可能になっている。

ViPAR のプログラム、マニュアル等は下記で公開されている

<http://bioinformatics.childhealthresearch.org.au/software/vipar/>

D. 考察

1. 死因情報を外したアーカイブ環境を想定し、必要時に死因を人口動態二次利用申請し、アーカイブセンター (データセンター) にて照合・集計・解析を行う運用

本検討では、死因情報を外したアーカイブ環境を想定し、必要時に死因を人口動態二次利用申請し、アーカイブセンターにて照合・集計・解析を行う運用例を提案し、国内大規模循環器疫学研究 (JALS) で、実際に運用し死因照合を行った。JALS は各コホート研究 (ローカルサイト) データを中央事務局で統合して実施している研究であることから、この研究で行う死因照合作業の実際が、本検討で想定している研究データアーカイブの環境づくりに対して参考になる点が多いように思われた。死因データの提供を受けてから実施する照合

作業については、JALS の研究進捗にあわせて、統計プログラム (SAS) を使い、データの読み込み、Excel などへの帳票出力に至るまで一連の作業の手順化をすすめてきた。多くの作業をルチーン化してきており、大部分の課題は解決してきているが、本提案例をアーカイブセンターで運用する際に予想される課題について検討する。

1) 厚生労働省から受領する統計フォーム (データ形式) に大幅な変更がないことを前提とする

本検討で提案した仕組みは、人口動態二次利用申請により提供される死因データの構造に大幅な変更がないことを前提としている。構造の変更を伴わない小規模の変数追加等の変更であれば、受領後の確認作業を手順化すること (変数表の確認作業やプログラム修正箇所の明示) で、アーカイブセンタースタッフ、疫学研究に従事する研究者等で十分対応できる作業と考えられる。

2) 死因照合率はコホートから提出される照合キー変数の精度に依存している

死因照合の一致率については、各コホート研究から提出される照合キー変数 (死亡地、性別、死亡年月日、生年月日) の情報精度に依存している。JALS では、自治体の住民基本台帳などの確認後の情報に基づき照合作業を実施しているが、未照合となる例は、提供情報に何らかの誤りがあり、情報の再確認後に照合可能となる例が大部分である。一方で、提供情報に誤りがないにもかかわらず照合例を検索できない例が少なからず存在しており、これらは性別、生年月日、死亡日について、人口動態統計作成の過程で発生した何らかの入力ミス等が原因として考えられる。しかし、本検討で想定しているのは追跡終了後コホート研究のデータアーカイブ化であるので、これまでの研究実施の過程 (各研究での死因照合作業などの過程) でこれらの変数は十分確認されてきており、情報の精度については、進行中の研究に比べて十分高く、こうした問題の発生は少ないと考えられる。

3) 死因付与データが使用できる環境が限定され

る

提案する方法では、死因付与後のデータセットの利用場所がアーカイブセンター（あるいは申請書に記載した研究者の所属する機関）に限られることになる。したがって、研究の解析も同様にこの利用場所に限られることになり、データ解析を行える環境（物理的な環境、統計家の配置等）についても検討する必要がある。また、死因データの申請者の所属機関で実施する場合には、アーカイブデータの外部利用の規約等の整備も必要があるといえる。

2. 分散型ネットワーク環境を利用し、疫学追跡終了後コホートデータの共通利用が可能かどうか

分散型ネットワークが開発され、利用されている背景には、倫理的規定などからデータ使用の場が制限され物理的に施設外に出せないが、何らかの方法でデータ利用を促進したいというニーズが存在する。疫学研究でのデータ連携では、医療（臨床）データ連携が必要とされるような即時的なデータ連携、多種多様なデータをそのまま連携したいというニーズは高くはない。一方で、以下のような医療データ連携にはない、疫学研究で重視すべき項目が考えられる。

① 様々な統計解析を柔軟に行えること

医療（臨床）データ連携のように、定型化された集計作業（例：Mini sentinel）が中心ではないので、ViPARのような各種統計パッケージ等と連動して運用可能であるかどうかは重視すべき点であると考えられる。ただし、ViPARについては、RAM上で解析を行い、外部にデータ保存が行われないというだけであり、施設外にデータが出るといふことの定義や解釈が利用の際の要件となってくる。

② ローカルサイト（大学、研究所、医療機関を想定）へアクセスを行う上でのセキュリティ

昨今、施設によっては外部からのアクセスに対して制限を設ける場合も多くなっている点を考

慮すべきである。

③ 人口動態統計の二次利用に関連した法令解釈

- ・ローカルサイト単位で各々申請し、死因を付与したデータを統合して用いることの可否
- ・ViPARでの運用のように、施設外部のPC上のRAMへの一時的な書き出し（保存はされない、電子データの「一時的蓄積」）の可否

などが問題として考えられる。

文部科学省文化審議会著作権分科会等でも機器利用時・通信過程における一時的なデータ蓄積については議論されており、「一時的固定（複製）」については、次のように整理されている（コンピュータに該当する部分抜粋）。

ア. 瞬間的・過渡的な蓄積であり「複製」ではないもの

- ・処理装置(CPU)の読み込み
- ・ビデオRAMへの書込み

イ. 一時的固定（複製）のうち、「複製」と判断すべきものではないもの

- ・主記憶(RAM)への蓄積
- ・補助記憶のドライブキャッシュ注釈
- ・CPUにおける1次キャッシュ2次キャッシュ

ウ. 一時的固定（複製）のうち「複製」と判断すべきもの

- ・主記憶(RAM)への蓄積（常時蓄積）

このような解釈は、技術動向を見極めて判断されるものでもあり難しい課題であるといえる。

一施設にデータを集約しない分散型ネットワークによるデータ連携（交換）は、本検討の既存研究データの連携に限ったことではなく、今後新たに疫学研究を行う際にも検討していく必要がある。このような方法を活用していくことで、共同研究の促進や研究のマネジメントの効率化・生産性の向上が期待できるほか、信頼性の高いデータ収集へつなげていくものと考えられる。

E. 結論

死因情報を付与した形でのデータアーカイブ化が難しい本邦の現状を鑑みて、現制度下での運用案を提案した。提案1の運用案を複数コホート研究から構成される国内の循環器疫学研究に適用したところ、死因照合率は99.7%であり、照合作業の技術的側面、作業手順化の面では問題はなく、運用案の一つになりうると考えられた。想定するアーカイブセンター等で実際に運用を行う際には、データ利用の規約、データ利用の物理的環境、サポート体制などについては検討すべき必要があると考えられた。提案2については、疫学共同研究における分散型ネットワークを用いたデータ連携の先行事例などから、様々な統計解析を柔軟に行えること、ローカルサイト(大学、研究所、医療機関を想定)へアクセスを行う上でのセキュリティ、さらには、人口動態統計の二次利用をローカルサイト単位で各々申請し、死因を付与したデータを統合して用いることや、施設外部のPC上のRAMへ

の一時的な書き出し(保存はされない、電子データの「一時的蓄積」)の可否など、法令解釈の問題にも留意する必要があると考えられた。

F. 健康機器情報

特になし

G. 研究発表

1. 論文発表 なし
2. 学会発表 なし

H. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得
2. 実用新案登録
3. その他

いずれもなし

アーカイブデータ提供のためのガイダンスに含める必要がある項目の整理

研究代表者	玉腰暁子	北海道大学大学院医学研究科
研究分担者	祖父江友孝	大阪大学大学院医学系研究科
研究協力者	辻 一郎	東北大学大学院医学系研究科

研究要旨

特に追跡の終わったコホート研究の利活用を進めることを念頭に、データ提供の際に従うべきガイダンスに含める必要がある項目を5点に整理した。その5点とは、①対象者の個人情報保護、②インフォームド・コンセントと倫理審査、③知的財産権の帰属、④寄託する項目/しない項目の判断、⑤データ提供先の制限、である。

A. 目的

疾病の発生関連要因の検討を目的としたコホート研究では、生活習慣の詳細や既往歴等の情報をベースに、その後の疾病発生や死亡状況等を追跡し、情報を蓄積する。したがって、追跡を終了したコホート研究データのアーカイブ化を進めるにあたっては、利用を促進すると同時に倫理面や研究参加者の個人情報保護に配慮した適切なデータ提供が求められる。しかし、研究開始時点のデータ取り扱いや対象者への説明同意のあり方は、現在従うことが求められている人を対象とした医学系研究に関する倫理指針で定めるものとは異なっている。そこで、今後利活用を進めることを念頭に、データ提供の際に従うガイダンスに含める必要があると考えられる項目の整理を試みた。

B. 研究方法

昨年度までの研究成果、ならびに、国内のヒトデータベース(ライフサイエンス分野)の収集・公開を担うバイオサイエンスデータベースセンター

(NBDC)のガイドライン、社会科学分野のデータアーカイブの運営を行っている東京大学社会科学研究所附属社会調査・データアーカイブ研究センターSSJDAのデータ寄託手引きを参考に、特に研究開始時期のデータ取り扱いが現在のものと異なり、かつ多数を取り扱う疫学研究データのアーカイブ化の場合にガイダンスに含めておくべき項目を検討した。

(倫理面への配慮)

特に必要なし。

C. 結果

含める必要があると考えられたのは、以下の項目であった。

- ① 対象者の個人情報保護: SSJDAでは回答者のプライバシー保護に最大限の注意が払われており、調査地域を粗く束ねるなどの手段を用いて、個々の回答者の識別を不可能にしている。我々の実データを用いた過去2年間の検討でも調査項目が多岐に渡り、かつ対象者数が多

いコホート研究では、データの一意性はあるものと考えた方がよいとしており、この点に関してデータ寄託者、アーカイブ運営者ともに十分に認識する必要がある。

② インフォームド・コンセントと倫理審査：NBDCでは対象者からデータ寄託に関する同意が得られていることを原則とした上で、寄託者の所属機関で倫理審査を受けることを義務付けている。しかし、既に追跡が終わったコホート研究の場合、その開始は今から20年以上前であると考えられ、研究の説明同意プロセスは現在と異なる。また、コホートによっては、地域との関係が現在まで維持されていないことも考えられる。そのため、対象者の同意あるいは再同意の取得を前提とすることは困難であり、所属機関での倫理審査に加え何らかの形で二次利用に関する広報・情報公開を行うことが適切と考えられた。

③ 知的財産権の帰属：生体試料を伴わない情報のみのアーカイブでは、知的財産権の発生は稀と思われるが、生体試料を収集・保管するコホート研究では、研究終了時にも試料が残ることが想定される。試料・情報提供者と直接提供を受ける研究者の間であれば、知的財産権は研究者（研究機関）に帰属することとなり、最近では説明事項に含まれることも多い。一方、二次利用者とデータ寄託者の間の分配方法については、一定のルールが定まっていないことから、利用開始時に取り決めを交わしておくことが望ましい。

④ 寄託する項目/しない項目の判断：①で述べたような個人同定の可能性が特に高い項目は、丸め処理をするか、提供しないことが望まれる。また、調査方法や設定が特殊である等何らかの理由により、研究内容を十分に理解しない二次利用者の解析・解釈によって誤った結論を引き出すおそれがある項目についても、提供しないことが推奨される。

⑤ データ提供先の制限：二次利用者の資格、利用上の制約等については、各研究関係者の意向を反映するような区分けの周知とそれに沿った対応が求められると考えられた。

D. 考察

米国に比べ日本では、未だ疫学研究データの二次料体制が整っていないとはいえない。しかし、日本においても社会科学系では既に多くの調査データがアーカイブ化され、利用されていることから、その促進のための教育や環境整備が望まれる。一方で、疫学研究では生活習慣や生活環境のみならず、個人の心身面の健康状態情報を収集することから、その取り扱いには慎重であることが望まれる。今回、特に追跡の終わったコホート研究の利活用を進めることを念頭に、データ提供の際に従うべきガイダンスに含める必要がある項目を5点に整理した。今後そのような研究の寄託を進める際の指標になるものと考えられる。

E. 結論

特に追跡の終わったコホート研究の利活用を進めることを念頭に、データ提供の際に従うべきガイダンスに含める必要がある項目を5点に整理した。その5点とは、①対象者の個人情報保護、②インフォームド・コンセントと倫理審査、③知的財産権の帰属、④寄託する項目/しない項目の判断、⑤データ提供先の制限、である。

F. 健康機器情報

特になし

G. 研究発表

1. 論文発表 なし
2. 学会発表 なし

H. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得 なし
2. 実用新案登録 なし

3. その他

コホート研究で人口動態統計資料から得た死因情報を データアーカイブに付加して提供するための統計法解釈に関する検討

研究代表者	玉腰暁子	北海道大学大学院医学研究科
研究分担者	磯 博康	大阪大学大学院医学系研究科
研究分担者	大橋靖雄	中央大学理工学部人間総合理工学科
研究分担者	祖父江友孝	大阪大学大学院医学系研究科
研究協力者	辻 一郎	東北大学大学院医学系研究科
研究協力者	友岡史仁	日本大学法学部

研究要旨

現在までのところ、多くのコホート研究では統計法の規定に則り人口動態統計資料から死因情報を得ているため、死因をつけた形でのデータアーカイブの公開利用はできないと考えられる。エンドポイントとしての死因情報は非常に価値が高いものであることから、追跡が終了したコホート研究の二次利用を進めるために、死因情報のソースとなる人口動態統計調査の有効活用の方策を模索することが望まれ、その際の論点を2つ提示した。

A. 目的

通常のコホート研究では、追跡情報として重要な死因情報を人口動態統計資料（死亡個票）から、定められた手続きを経て入手する。この人口動態統計資料は、統計法に基づき作成されることから、同法40条では、同法に特別の定めがある場合を除き行政機関の長等が行った統計調査の目的以外の目的のために、当該統計調査に係る調査票情報を自ら利用したり提供してはならないとされていることから、研究目的での提供は同法第三章に規定される一定の条件の下で行われる。したがって、我々研究者が死因情報の提供を受ける多くの場合は、同法33条2号の規定が適用されることから、入手にあたっては、公的研究費を受け続けて

いることなど、以下の様な一定の要件が課せられることになる。

平成23年6月厚生労働省「調査票情報の提供に関する利用申出手引」における

イ 公的機関から委託を受け、又は公的機関と共同して行う調査研究に係る統計の作成等、

ロ その実施に要する費用の全部又は一部を公的機関が公募の方法により補助して行う調査研究に係る統計の作成等、

ハ 行政機関の長又は地方公共団体の長その他の執行機関が、その政策の企画、立案等に有用であると認める統計の作成等。

しかし、このような要件は、追跡が終わったコ