

厚生労働科学研究費補助金（循環器疾患・糖尿病等生活習慣病対策総合研究事業）
分担研究報告書

平成 17 年国民生活基礎調査－国民健康・栄養調査－歯科疾患実態調査のデータリンケージ状況
と性・年齢の不一致について

研究代表者：安藤 雄一（国立保健医療科学院・生涯健康研究部）
研究分担者 三浦 宏子（国立保健医療科学院）
研究協力者：佐藤 眞一（千葉県衛生研究所）
荒井 裕介（千葉県立保健医療大学健康科学部栄養学科）
齋藤 俊行（長崎大学医歯薬学総合研究科 口腔保健学）
北村 雅保（長崎大学医歯薬学総合研究科 口腔保健学）
川下由美子（長崎大学医歯薬学総合研究科 口腔保健学）
濱寄 朋子（九州女子大学家政学部 栄養学科）
加藤 佳子（神戸大学大学院人間発達環境学研究科 健康発達論）
若井 建志（名古屋大学大学院医学系研究科 予防医学）
深井 穂博（深井保健科学研究所）
大庭 志野（国立保健医療科学院・生涯健康研究部）

研究要旨

平成 17 年の国民生活基礎調査（基礎調）、国民健康・栄養調査（健栄調）、歯科疾患実態調査（歯実調）のデータリンケージを行ったところ、健栄調にリンケージできなかった歯実調レコードは 1.11%、基礎調にリンケージできなかった健栄調レコードは 3.47%であった。

リンケージデータについて、3 調査における性および年齢データの食い違いをみたところ、健栄調ベースのリンケージデータ（健栄調＋基礎調）では 799 件（8.66%）、歯実調ベース（歯実調＋健栄調＋基礎調）ベースのリンケージデータでは 372 件（8.36%）に何らかの食い違いが認められた。性と年齢を比較すると、年齢の食い違いが生じる頻度が高かった。

また、ID、性・年齢の食い違いが生じる頻度を地区別に検討したところ、食い違いの多くが一部の地域に偏って生じていた。

A. 目的

筆者らは、平成 23 年度の本厚生労働科学研究「口腔機能に応じた保健指導と肥満抑制やメタボリックシンドローム改善との関係についての研究」の一環として、食品・栄養摂取状況と口腔健康状態の関連を検討することを主たる目的に、平成 17 年の国民生活基礎調査（以下、基礎調）、国民健康・栄養調査（以下、健栄調）、歯科疾患実態調査（以下、歯実調）の目的外使用を申請し、厚生労働省の各調査の担当窓口より利用許可を得て、この 3 調査についてデータリンケージを行い、各種分析を行った。

この 3 調査の関係は、健栄調の調査対象地区が、基礎調の世帯票における 1,056 の調査対象地区¹⁾より設定された単位区から層化無作為抽出された 300 単位区であり²⁾、歯実調の対象者は健栄調と同一であり³⁾、ID（都道府県コード・地区コード・世帯コード・世帯員コード）の一致によりデータリンケージを行うことができる。

政府統計をデータリンケージして個票データを分析する意義や価値の高さ^{4,5)}については論を俟たない※、それに関する実証的な報告例はあまり多くないようである。

また、ID によりリンケージできたとしても、別の人に同じ ID が付与されている可能性が残る。そのような場合、性と年齢の情報を用いれば、ある程度のチェックが可能と考えられる。しかしながら、そのような検討は今までほとんど行われていないようである。そこで、筆者らは、本目的外使用で主たる目的とする分析に先立ち、平成 17 年の基礎調、健栄調、歯実調の 3 データのリンケージデータにおける各調査の性および年齢情報の一致度について分析を行い、興味深い結果を得たので報告する。

B. 方法

1. 用いたデータ

表 1 にデータリンケージを行った 3 調査の名称・調査票名、提供時のレコード件数を示す。国民生活基礎調査（世帯票）では、提供データのレコードが世帯を単位としているもの（ex. 3 人世帯では 1 レコードに 3 人分の情報が入っている）であったが、個人単位のファイルを作成して分析を行った。

表 1. データリンケージを行った調査の一覧

調査年	調査名	調査票名	レコード件数
2005(平成17)年	国民生活基礎調査	世帯票	44,999 (世帯数) ^{注1}
2005(平成17)年	国民健康・栄養調査	栄養摂取量票	9,561 (人数) ^{注2}
		食品群別摂取量票	
		身体状況・生活習慣票	
2005(平成17)年	歯科疾患実態調査	—	4,606 (人数)

注1: 人数=12,636人

注2: 国民生活基礎調査における各調査票ごとの調査人数は実際のところ異なっている。

2. 分析方法

1) データリンケージ

データリンケージは、3 調査に共通する以下の ID 項目を用いて、すべてが一致する場合にデータをリンケージ（結合）させた。

都道府県
地区番号
単位区番号
世帯番号
世帯員番号

データリンケージは、①歯実調を健栄調にリンケージ、②そのリンケージデータを基礎調にリンケージ、という手順で行い、それぞれの段階で ID 不一致によりリンケージできなかったレコード件数（以下、ID 不一致件数）を把握した。

2) 性・年齢の一致度

データリンケージ完了後、各調査における性および年齢の情報が一致しているか否かを検討した。その際、性については男か女かが一致するか否かを単純に判定した。年齢については、基礎調の実施時期が 2005 年 6 月で¹⁾、健栄調と歯実調が 11 月である^{2,3)}ことから、基礎調の年齢と健栄調・歯実調の年齢との差が 0 または -1 である場合を年齢一致とし、それ以外は不一致とした。健栄調と歯実調については、歯実調では調査日における年齢の特定が可能であるが、健栄調では 3 つの調査票（身体状況／栄養摂取状況／生活習慣）があり、調査実施日に関する情報がない。そのため、両者の差で一致度を評価する際の許容範囲が -1 または 0 または + 考えられることから、差の値が -1 ～ +1 の場合を一致、それ以外の値を不一致とした。

3) 地区単位でみた分析

ID 不一致および性・年齢の不一致が認められたレコードについて、調査を行った地域単位別に集計し、地域的な偏りが生じているか否かを検討した。この際、地域単位は上述した ID のうち、都道府県・地区番号・単位区番号により識別した。

以上の分析に用いたソフトは Stata 12⁸⁾で、この merge コマンドを用いてリンケージを行った。地区単位の分析は Stata 12 で出力された結果を Microsoft Excel 2010⁹⁾を用いて分析処理した。

C. 結果

1. ID によるデータリンケージ

図 1 に ID を用いてデータリンケージを行った結果を示す。

歯実調では、提供されたデータ（4,606 人）から、うち ID の不明と重複を除いたデータ（4,602 人分）を健栄調にリンケージを試みたところ、ID 不一致は 51 件で、提供データの 1.11%であった。

基礎調では、提供データが世帯単位（44,499 世帯分）だったので、これを人単位に変換し

(120,636 人分)、健栄調と歯実調の調査地区への絞り込んだところ、14,645 件であった。このデータに歯実調とリンケージされた健栄調データのリンケージを試みたところ、ID 不一致は 332 件で、健栄調の全データ (9,561 件) の 3.47% に相当した。

最終的に、国民健康・栄養調査のリンケージデータとして利用できたものは 9,299 件 (9,299 人分) であった。また、さらに歯科疾患実態調査とのリンケージデータとして利用できたものは 4,450 件 (4,450 人分) であった。

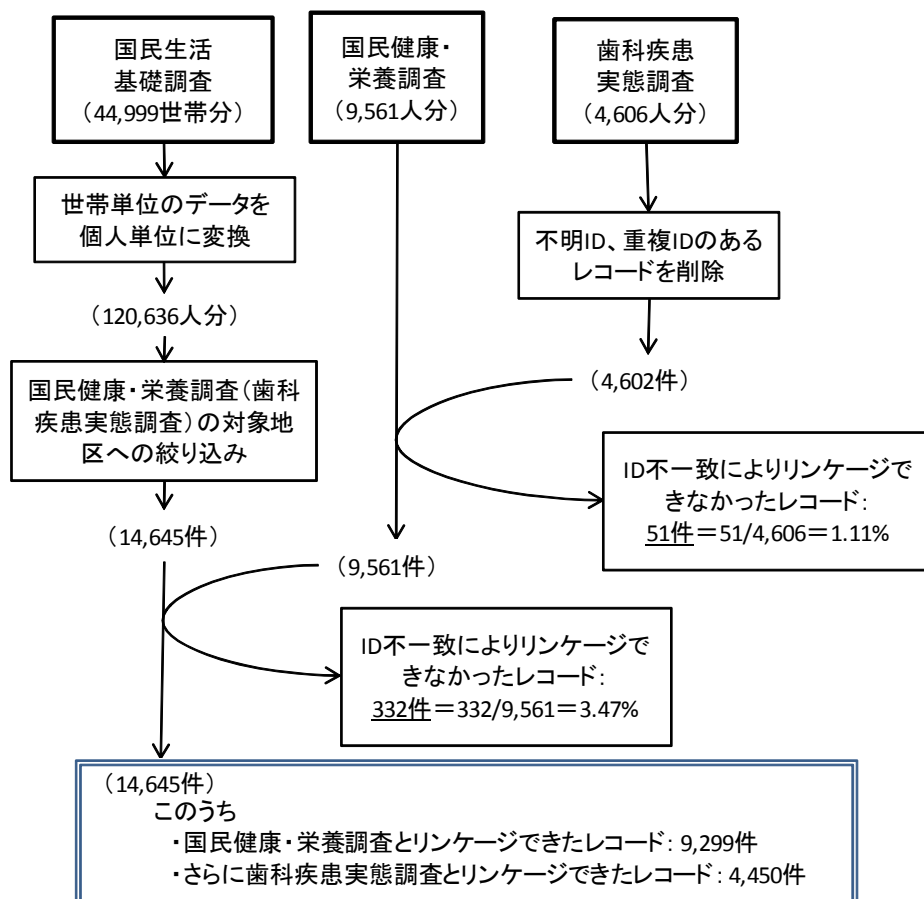


図1. IDによるデータリンケージ結果

2. 性および年齢の一致度

表 2～表 4 に各調査間における性の一致状況を示す。不一致の割合は、基礎調 vs 健栄調が 4.3% (表 2)、基礎調 vs 歯実調が 3.7% (表 4) と、健栄調 vs 歯実調の 0.4% (表 3) に比べて高い割合を示した。

表 5～表 7 に各調査間における年齢の一致状況を示す。不一致の割合は、全般的に性の不一致よりも高く、基礎調 vs 健栄調が 8.7% (表 5)、基礎調 vs 歯実調が 7.4% (表 7) と、健栄調 vs 歯実調の 0.9% (表 6) に比べて高い割合を示した。

性の不一致と年齢の不一致のどちらかが生じた場合を「性・年齢不一致あり」として、その件数(割合)をみたところ、健栄調ベースのリンケージデータ(健栄調+基礎調)では 799 件 (8.66%)、歯実調ベース(歯実調+健栄調+基礎調)ベースのリンケージデータでは 372 件 (8.36%) であった。

表2. 性の一致度(基礎調 vs 健栄調)

		健栄調					
		人数			割合		
		男	女	計	男	女	計
基礎調	男	4,166	207	4,373	45.1%	2.2%	47.4%
	女	189	4,667	4,856	2.0%	50.6%	52.6%
計		4,355	4,874	9,229	47.2%	52.8%	100.0%
不一致		人数 396		割合 4.3%			

表3. 性の一致度(健栄調 vs 歯実調)

		歯実調					
		人数			割合		
		男	女	計	男	女	計
健栄調	男	1,864	11	1,875	41.9%	0.2%	42.1%
	女	8	2,567	2,575	0.2%	57.7%	57.9%
計		1,872	2,578	4,450	42.1%	57.9%	100.0%
不一致		人数 19		割合 0.4%			

表4. 性の一致度(基礎調 vs 歯実調)

		歯実調					
		人数			割合		
		男	女	計	男	女	計
基礎調	男	1,807	100	1,907	40.6%	2.2%	42.9%
	女	65	2,478	2,543	1.5%	55.7%	57.1%
計		1,872	2,578	4,450	42.1%	57.9%	100.0%
不一致		人数 165		割合 3.7%			

表5. 年齢の一致度(基礎調 vs 健栄調)
差=基礎調-健栄調

差の値	人数	%		
~-80	3	0.03%	91.34%	8.66%
-79~-70	6	0.07%		
-69~-60	17	0.18%		
-59~-50	20	0.22%		
-49~-40	9	0.10%		
-39~-30	55	0.60%		
-29~-20	86	0.93%		
-19~-10	36	0.39%		
-9~-6	23	0.25%		
-5	18	0.20%		
-4	14	0.15%		
-3	33	0.36%		
-2	57	0.62%		
-1	3,380	36.62%		
0	5,050	54.72%		
1	64	0.69%		
2	51	0.55%		
3	37	0.40%		
4	16	0.17%		
5	13	0.14%		
6~9	22	0.24%		
10~19	34	0.37%		
20~29	71	0.77%		
30~39	31	0.34%		
40~49	7	0.08%		
50~59	34	0.37%		
60~69	18	0.20%		
70~	8	0.09%		
欠値	16	0.17%		
計	9,229	100.00%		

表6. 年齢の一致度(健栄調 vs 歯実調)
差=健栄調-歯実調

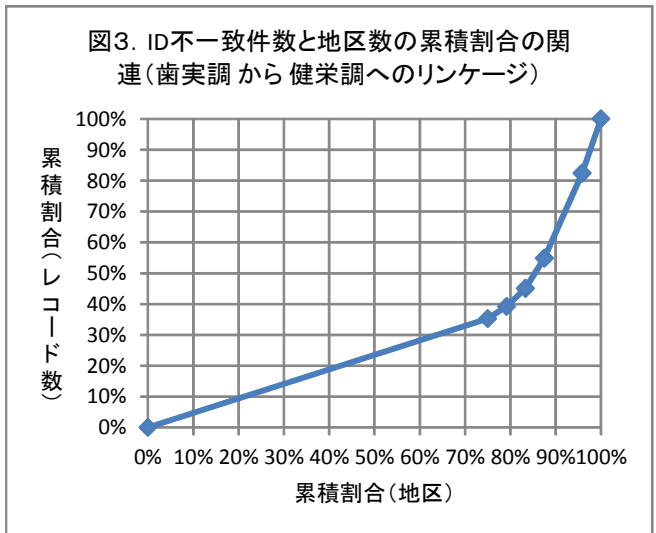
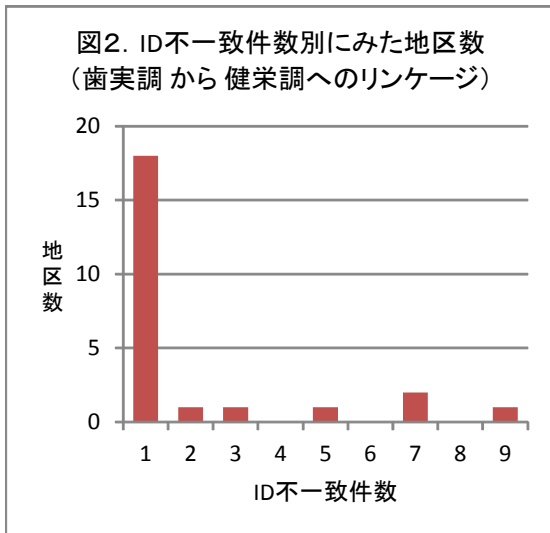
差の値	人数	%		
~-30	6	0.13%	99.06%	0.94%
-29~-20	2	0.04%		
-19~-10	2	0.04%		
-9~-6	3	0.07%		
-5	2	0.04%		
-4	4	0.09%		
-3	2	0.04%		
-2	3	0.07%		
-1	354	7.96%		
0	4,048	90.97%		
1	6	0.13%		
2	6	0.13%		
3	0	0.00%		
4	1	0.02%		
5	1	0.02%		
6~9	1	0.02%		
10~19	3	0.07%		
20~29	3	0.07%		
30~39	0	0.00%		
40~49	0	0.00%		
50~59	1	0.02%		
60~	2	0.04%		
欠値	0	0.00%		
計	4,450	100.00%		

表7. 年齢の一致度(基礎調 vs 歯実調)
差=基礎調-歯実調

差の値	人数	%		
~-70	2	0.04%	92.56%	7.44%
-69~-60	7	0.16%		
-59~-50	8	0.18%		
-49~-40	7	0.16%		
-39~-30	32	0.72%		
-29~-20	40	0.90%		
-19~-10	19	0.43%		
-9~-6	12	0.27%		
-5	3	0.07%		
-4	4	0.09%		
-3	12	0.27%		
-2	23	0.52%		
-1	1,934	43.46%		
0	2,185	49.10%		
1	17	0.38%		
2	24	0.54%		
3	10	0.22%		
4	9	0.20%		
5	7	0.16%		
6~9	9	0.20%		
10~19	13	0.29%		
20~29	26	0.58%		
30~39	12	0.27%		
40~49	3	0.07%		
50~59	10	0.22%		
60~69	10	0.22%		
70~79	3	0.07%		
80~	2	0.04%		
欠値	7	0.16%		
計	4,450	100.00%		

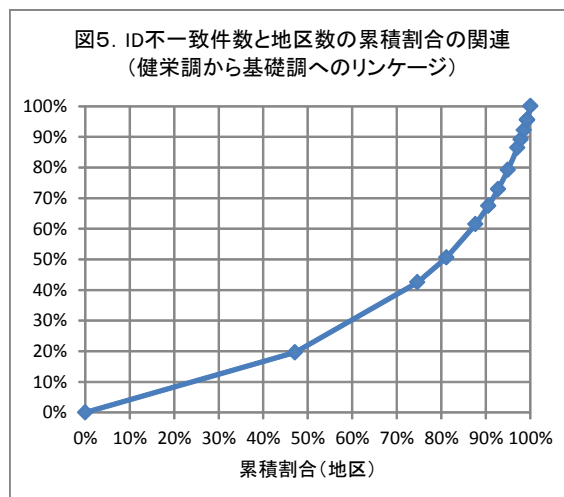
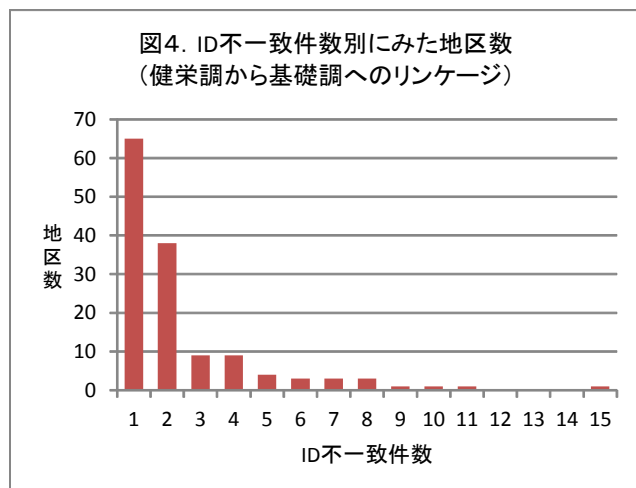
3. 地区単位でみた分析

① リンケージ



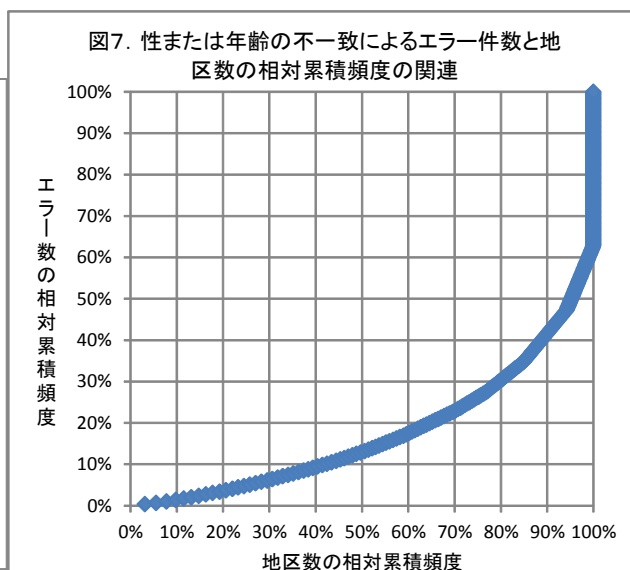
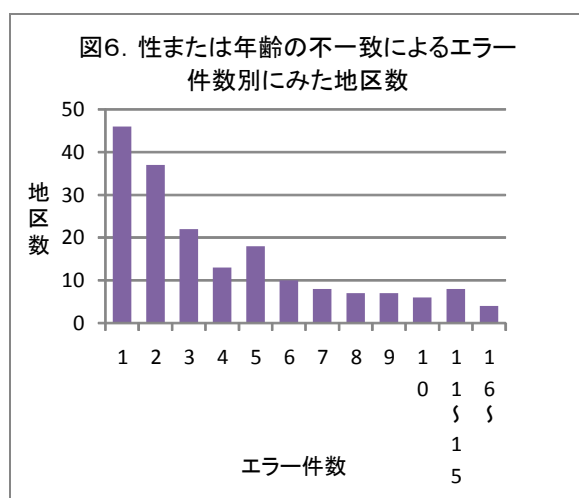
健栄調にID不一致によりリンケージできなかった歯実調の51レコード(図1)は24地区で生じていた。図2は、この24地区の分布をID不一致件数別に示したのである。24地区のうち4分の3が、ID不一致件数が1件のみであった。ID不一致件数は一部の地区に偏在しており、図3に示すように、ID不一致件数の約7割は約3割の地区で生じていた。

基礎調にID不一致によりリンケージできなかった健栄調+歯実調の332レコード(図1)は138地区で生じていた。図4は、この138地区の分布をID不一致件数別に示したものである。138地区のうち4分の3が、ID不一致件数が2件以下であった。ID不一致件数は一部の地区に偏っており、図5に示すように、ID不一致件数の約3分の2が約3分の1の地区で生じていた。



② 性または年齢の不一致

健栄調ベースのリンケージデータ(健栄調+基礎調)では、性または年齢の不一致がみられた799件は186地区で生じていた。図6は、エラー件数別にみた地区数を示したものである。186地区のうち、6割弱がエラー件数が3件以内であった。エラー件数は少数地区に偏っており、エラーの4分の3が4分の1の地区で、また、エラーの約半分が5%程度の地区で生じていた。



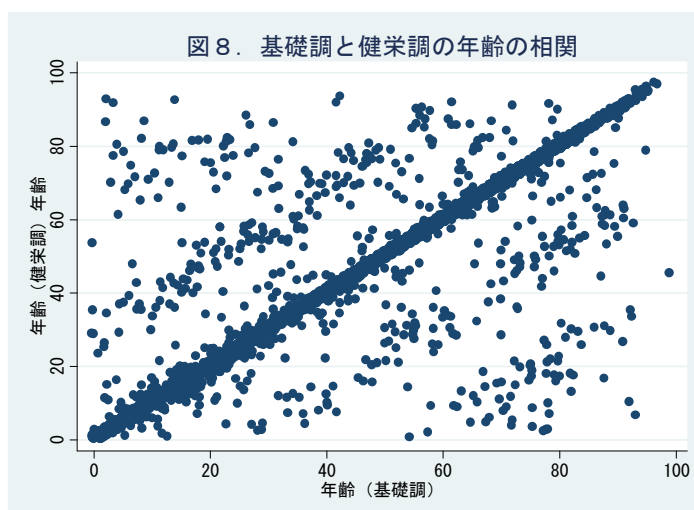
D. 考察

今回レコードリンケージを行った3調査のうち、歯実調は健栄調の身体状況調査と同一会場で行われるなど、健栄調との関連性は基礎調に比べると強いといえる。歯実調と健栄調との間のID不一致(図1)と性・年齢の不一致(表2～表7)が生じた割合が低値を示したのは、そのためと考えられる。一方、基礎調は他の2つの調査と調査時期が異なり(基礎調は2005年6月、健栄調と歯実調は同年11月)、調査対象も広範囲であり、IDや性・年齢情報に食い違いが生じやすい面があるものと思われる。

これらのエラーを抑制する必要があることは論を俟たないが、ゼロに抑えることは困難が予想され、一定程度のエラー発生は許容せざるを得ないと思われる。しかしながら、これらのエラー発生状況を調査地区別にみると、エラー発生が一部の地区に偏って発生していることが明らかであった(図3、図5、図7)。これらの地区では、データの管理面で何らかの問題がある可能性が高く、改善の余地が残されていると考えられる。

今回の分析では、ID、性、年齢の3情報の各調査による食い違いを比較したわけであるが、これらのうち、食い違いが生じた割合が最も高かったのは年齢について基礎調と他の2調査を比較した場合で、全レコードの7.44～8.66%にエラー(食い違い)が生じていた(表5、表7)。

図8は、基礎調と健栄調の年齢の関連を示した散布図であり、両者の差(基礎調の年齢－健栄調の年齢)の許容範囲である-1～0の範囲、すなわち $Y=X$ または $Y=X+1$ の直線上に多くの点が分布している。その一方で、この直線に平行する点の集まりが幾つか認められるが、これは元号の記入ミスである可能性が高いと考えられる。また、国民生活基礎調査(世帯票)では、年齢に関する情報入力欄生年月のみで、生年は西暦ではなく元号であり、年齢そのものの入力欄はない。この際に元号御記入が生じた割合が比較的高いことが想像される。また、もし、この際、年齢そのものの情報の記入が求められれば、一種のチェックが働きエラー抑制につながる可能性が期待できると思われる。調査票の記入項目を最小限にとどめることは負担軽減の意味で重要であるが、これが10%近いエラー発生の原因になっているのだとすれば、元号を西暦に改める、或いは年齢の記入を追加するといった改善策を検討する必要があると思われる。



E. 結論

平成17年の基礎調、健栄調、歯実調のデータリンケージを行ったところ、健栄調にリンケージできなかった歯実調レコードは1.11%、基礎調にリンケージできなかった健栄調レコードは3.47%であった。

リンケーデータについて、3 調査における性および年齢データの食い違いをみたところ、健栄調ベースのリンケーデータ（健栄調＋基礎調）では 799 件（8.66%）、歯実調ベース（歯実調＋健栄調＋基礎調）ベースのリンケーデータでは 372 件（8.36%）に何らかの食い違いが認められた。性と年齢を比較すると、年齢の食い違いが生じる頻度が高かった。

また、ID、性・年齢の食い違いが生じる頻度を地区別に検討したところ、食い違いの多くが一部の地域に偏って生じていた。

F. 研究発表

なし

G. 知的財産権の出願・登録状況

なし

H. 引用文献

- 1) 平成 17 年 国民生活基礎調査の概況－調査の概要
<http://www.mhlw.go.jp/toukei/saikin/hw/k-tyosa/k-tyosa05/gaiyo.html>（厚生労働省ウェブサイト、2012 年 5 月 12 日アクセス）
- 2) 平成 17 年国民健康・栄養調査報告－目次・調査の概要
<http://www.mhlw.go.jp/bunya/kenkou/eiyoubu07/dl/01-mokuji.pdf>（厚生労働省ウェブサイト、2012 年 5 月 12 日アクセス）
- 3) 平成 17 年歯科疾患実態調査結果について－調査の概要
<http://www.mhlw.go.jp/topics/2007/01/tp0129-1a.html>（厚生労働省ウェブサイト、2012 年 5 月 12 日アクセス）
- 4) 井伊雅子. 社会的共通資本としての統計情報（特集：社会的共通資本としてのヘルスケアシステム）. 保健医療科学 2010 ; 59(1) : 17-20.
- 5) 橋本英樹. 今後の国民生活基礎調査の在り方についての一考察 健康票を中心に. 厚生 の 指 標 2009 ; 56(1) : 1-8.
- 6) 吉池信雄. 国民栄養調査における兵法抽出方法等による検討. 平成 10 年厚生科学研究費補助金健康科学総合研究事業「国民栄養調査の再構築に関する研究」分担研究報告書. 1999.
- 7) 西信雄、中出麻紀子、猿倉薫子、野末みほ、坪田恵、三好美紀、卓興鋼、由田克士、吉池信男. 国民健康・栄養調査の協力率とその関連要因. 厚生 の 指 標 2012.4 ; 59(3) : 10-15.
- 8) Stata <http://www.stata.com/>（Stata 社ウェブサイト、2012 年 5 月 12 日アクセス）
- 9) Excel 2010 <http://office.microsoft.com/ja-jp/excel/>（Microsoft 社ウェブサイト、2012 年 5 月 12 日アクセス）