

## 第 4 3 回 統計委員会 議事録

1 日 時 平成 23 年 2 月 24 日（木）16：40～17：30

2 場 所 中央合同庁舎第 4 号館 12 階 共用第 1208 特別会議室

3 出席者

### 【委員】

樋口委員長、深尾委員長代理、縣委員、安部委員、井伊委員、佐々木委員、  
首藤委員、椿委員、津谷委員、廣松委員、山本委員

### 【統計委員会運営規則第 3 条による出席者】

《国または地方公共団体の統計主管部課の長等》

内閣府経済社会総合研究所長、内閣府経済社会総合研究所総括政策研究官、  
総務省統計局統計調査部長、文部科学省生涯学習政策局調査企画課長、厚生  
労働省大臣官房統計情報部長、農林水産省大臣官房統計部長、経済産業省経  
済産業政策局調査統計部長、国土交通省総合政策局情報安全・調査課長、日  
本銀行調査統計局審議役、東京都総務局統計部社会統計課長

### 【事務局等】

西川内閣府総括審議官、乾内閣府大臣官房統計委員会担当室長、杉山内閣府  
大臣官房統計委員会担当室参事官、千野総務省政策統括官付統計企画管理官

4 議 事

- (1) 諮問第 35 号の答申「季節調整法の適用に当たっての統計基準」の設定について
- (2) 部会の審議状況について
- (3) その他

5 議事録

○樋口委員長 それでは、定刻を過ぎましたので、ただいまから第 43 回「統計委員会」を開催いたします。

本日は阿藤委員、宇賀委員が所用のためご欠席とのことでございます。

それでは、議事に入る前に本日用意されている資料について、事務局から説明をお願いします。

○乾内閣府統計委員会担当室長 お手元の資料 1 「諮問第 35 号の答申『季節調整法の適用に当たっての統計基準』の設定について（案）」。

資料 2 「人口・社会統計部会の審議状況について（報告）」。

資料 3 「匿名データ部会の審議状況について（報告）」があると思います。

あと、参考資料が 3 つございますので、ご確認ください。

○樋口委員長 それでは、議事に入ります。

まず「諮問第 35 号の答申『季節調整法の適用に当たっての統計基準』の設定について（案）」につきまして、山本部長からご説明をお願いいたします。

○山本委員 「季節調整法の適用に当たっての統計基準の設定」については、先月 26 日開催の統計委員会において総務大臣から諮問され、統計基準部会に審議が付託されたところです。

本件に関しては、去る 2 月 3 日に統計基準部会を開催して審議を行い、答申（案）をとりまとめましたので、部会の結果概要と併せてご報告申し上げます。

それでは、資料 1 答申（案）及び資料 1 の参考資料 2 の第 10 回統計基準部会の議事概要をご覧ください。なお、資料 1 の参考資料 1 は先日の統計委員会での諮問の際にお配りした資料です。

まず、資料 1 答申（案）の「1 設定の適否」では、「季節調整法の適用に当たっての統計基準」については、諮問案により、統計法第 28 条第 1 項に基づき統計基準として設定することは差し支えないとしています。

次にその具体的な理由として、まず「統計基準としての設定」については「2 理由」の「(1) 統計基準としての設定」に記載されているとおりであります。具体的には季節調整法の適用に当たり、適切性が評価されていない手法が使われたり、手法が頻繁に変更された場合、またオプション等の設定内容などの季節調整法の運用に関する情報が公表されず、季節調整値の客観性が確保されない場合には、季節調整値に基づく経済動向の判断や複数のデータを用いた総合的な判断が困難となります。

こうしたことから、平成 9 年に当時の統計審議会が了承した指針において、適切な手法を継続的に使用することや、季節調整法の運用に関する情報を公表することなどが実務上のガイドラインとして示され、これに沿って公的統計の分野において季節調整が行われていたところでもあります。

こうした状況に鑑みると、季節調整値について、その客観性の確保を図るとともに、統一性、総合性の確保を図るため、季節調整法を適用する場合の手法や運用に関する情報の公表事項等に係る基準を設けることが必要であると考えられます。

更に、平成 19 年に全部改正された統計法において、新たに、公的統計の作成に係る幅広い各種の技術的な基準として「統計基準」を設定することが可能となりました。

以上を勘案すると、季節調整法の適用に当たっての基準を統計法に規定されている統計基準として設定することは適当であると判断されるということでもあります。

「諮問案の内容」についても、これまでの季節調整法の適用に当たって実務上のガイドラインとして機能してきた平成 9 年の統計審議会指針をおおむね踏襲した内容となっていることに加え、利用者の利便の向上を図るとともに、情報通信環境の変化に対応する観点から、平成 9 年の指針に必要な変更を行ったものであります。したがって「2 理由」の「(2) 諮問案の内容」に記載のとおり、いずれの項目も適当であると判断いたしました。

具体的には、(2)のアで、季節調整法を適用する場合は、手法の適切性について国際的に一般的な評価を受けている手法を継続することについては、国際比較可能性を確保しつつ、季節調整を適

切に行うために理論上また実務上適切な方法であることが国際的に広く認められている手法を使用する必要があること、手法を継続して使用することは季節調整における恣意性を排除し、季節調整値の客観性の確保に寄与するものであることから適当としています。

(2)のイの(ア)ですが、季節調整法の適用に当たっては、手法の名称、推計に使用するデータ期間、オプション等の設定内容及び設定理由等の運用に関する情報を、季節調整値と併せて公表することとしています。

これについては、同一の時系列データに同一の手法を用いたとしても、推計に使用するデータ期間などが異なれば算出される季節調整値は異なったものになるため、利用者が季節調整値に基づいて経済動向の判断を行う際や複数の季節調整値を用いた総合的な判断を行う際には、このような運用に関する情報が必要であることから、適当としています。

(2)のイの(イ)ですが、オプション等の設定内容について重大な変更があるときは、推計に使用するデータ期間などの運用に関する情報に加えて、オプション等の変更の影響も公表することとしています。

これについては、オプション等の設定内容の重大な変更は、季節調整値の動きに変化を生じさせ経済情勢の判断に影響を及ぼす可能性があるため、利用者が適切な判断を行うにはオプション等の変更の影響に関する情報が必要であることから適当としています。

(2)のウですが、適用している手法を変更するときは、あらかじめ、変更内容、変更理由及び変更の影響を公表することとしています。

これについては、手法の変更は、季節調整値の動きに大きな変化をもたらす経済情勢の判断に大きな影響を及ぼす可能性があるため、利用者が適切な判断を行うには、季節調整値の公表以前に、あらかじめ新たな手法の内容や変更の影響に関する情報を把握し、理解を深めておく必要があることから、適当としています。

(2)のエですが、季節調整法の運用に関する情報などの公表方法については、インターネットの利用その他の適切な方法によることとしています。

これについては、近年の経済動向の急速な変化に対応するため、利用者に対する情報の迅速な提供が必要であることから適当としています。

答申（案）についての説明は以上ですが、答申（案）の審議に関連して、季節調整法に関して意見や要望が出されたので、ここでご参考までに紹介したいと思います。資料1の参考資料2をご覧ください。

現在、季節調整の手法としてはX-12-ARIMAが世界的に最も広く用いられていることを踏まえて、基準案の第1項、季節調整法を適用する場合の手法においてX-12-ARIMA等と例示しています。しかしながら、X-12-ARIMAはある程度経験則に基づく手法であって、理論的な裏づけにやや弱いことから、将来的には理論に基づくモデルに基礎を置く手法へ移行することも検討すべきではないかと考えています。

例えばヨーロッパの一部の国で用いられているTRAMO-SEATSや、より理論的な状態空間法に基づくDecompという手法が挙げられます。政府は季節調整の手法の国際的な研究動向に関する情報取

集に努め、そうした手法を取り入れる余地を作ってほしいとの意見がありまして、私も同感であります。

これに対して、総務省からは、季節調整の手法の国際的な研究動向を今後とも注視していく旨の説明がありました。

また、統計基準を設定した際には、総務省から各府省に対して基準の運用に関する留意点などが文書で示されることになるので、私はこの中で X-12-ARIMA 以外の前述の手法についても何らかの形で言及することにより、このような意見にも配慮していただくようお願いしました。

私からの報告は以上です。

○樋口委員長 ありがとうございます。

ただいまの説明につきまして、ご意見、ご質問がございましたら、お願いいたします。よろしいですか。

廣松委員、どうぞ。

○廣松委員 この答申（案）の位置付けですが、今までは答申（案）の中にありますとおり、平成9年に統計審議会が了承した指針という位置付けだったわけですが、今回はそれを統計基準として設定するという趣旨ということでよろしいのですか。ということは、今後、季節調整法に関して、今、部会長から現在用いている以外の方法もあり得るという発言がありましたが、そういうものに移行する場合は、諮問、答申という形の手続を経るということによろしいのでしょうか。

○樋口委員長 いかがでしょうか。

○山本委員 それはここで挙げられている例のことでしょうか。

○廣松委員 はい。

○山本委員 これは例として挙げられているだけであって、実際に変わっていても構わないと思います。その際に実際に取り上げているものが変わったから、ここで審議しなければいけないということはないと思います。

○廣松委員 そうですか。当然のことながら、変える場合にはこの答申（案）にあるように公表するということが義務付けられるということですね。

○山本委員 そういうことです。ここでは案の手法の1つとして例示しただけと考えています。

○樋口委員長 よろしいですか。

○廣松委員 わかりました。

○樋口委員長 ほかにどうでしょうか。よろしいでしょうか。

資料1の参考資料1にありますような諮問に対して、ただいまの資料1のように答申をしたいと考えておりますが、それでよろしいでしょうか。

（「異議なし」と声あり）

○樋口委員長 ありがとうございます。それでは、資料1によって総務大臣に対し答申します。

山本部会長始め、統計基準部会に所属されている先生方、どうもご苦勞様でした。

それでは、次に人口・社会統計部会の部会報告につきまして、本日、阿藤部会長がご欠席ですので、津谷部会長代理からご報告をお願いいたします。

○津谷委員 それではご報告させていただきます。お手元の資料2をご覧くださいと思います。今回は医療施設調査及び患者調査の変更についてでございます。

まず、部会の開催状況について簡単にご報告いたします。平成23年に実施される医療施設調査及び患者調査に関しまして、諮問が行われました平成22年12月17日金曜日の統計委員会以降、12月21日火曜日に第1回目の部会、1月24日月曜日に第2回目の部会を開催いたしました。

第2回目の部会は1月24日に開催されたものですけれども、患者調査の変更に関する最初の審議を行いました後、第1回目の部会で医療施設調査に関して示された要検討事項を審議いたしました。それでは、部会における主な議論等についてご報告いたします。

「(1) 患者調査の変更について」でございますが、実施主体の厚生労働省が計画している変更内容についてご説明いたします。

「(ア) 標本設計」についてでございますが、資料2の1ページの中段、5の(1)の(ア)でございますが、そこがございますとおり、病院の層化の基準の変更でございますが、これについては適当とされました。

「(イ) 調査事項」についてでございますが、お手元の資料の1ページの下段から2ページの上段にありますとおり、厚生労働省が計画している調査事項の変更については、おおむね適当とされました。ただし、診療費等支払方法欄における医療保険の選択肢、中でも被用者保険分の統合に関しましては、分析に支障が生じるおそれがあるということから、厚生労働省で再検討の上、次回の部会で改めて審議することとされました。

「イ 基本計画及び前回答申における指摘への対応について」ご報告申し上げます。

「(ア) 行政記録情報等の活用」についてでございます。資料2の2ページの中段にありますとおり、平成23年調査において行政記録情報等を活用しないことについてはやむを得ないとされました。ただし、平成26年以降の調査で、先ほどからお話が出ておりますレセプトやDPC調査のデータが活用されるよう、厚生労働省において工程表を作成し、それに基づいて着実に検討を進める必要があるとされました。

「(イ) セカンドオピニオンが利用されている疾患の把握」についてでございます。資料2の2ページの下段にありますとおり、セカンドオピニオンについては既に一般統計調査で把握が開始されておりまして、厚生労働省の対応は適当とされました。

「(ウ) 退院票の拡充」についてです。資料2の2ページの下段から3ページ上段にありますとおり、平成23年調査において退院票の標本規模や調査対象期間を拡充しないということについては、やむを得ないとされました。ただし、現行の標本規模は都道府県別、二次医療圏別分析を中心とした地域別の比較的発生が稀な疾患等の分析をするには十分ではないと考えられるため、平成26年以降の調査で標本規模の拡大が図られるよう、厚生労働省において検討を進める必要があるとされました。なお、標本規模の拡大は行政記録情報等活用による報告者負担の軽減と併せて検討されるべきであるとされました。

「(エ) オンライン調査の導入」についてです。資料2の3ページの中段から下段にありますとおり、平成23年の患者調査において、政府統計共同利用システムを用いたオンライン調査の導入を

見送るということについては、やむを得ないとされました。ただし、電子調査票による報告は、報告者や審査を行う地方公共団体の負担軽減に資することから、今後積極的に推進するべきであるとされました。

次に「(2) 医療施設調査の変更について」お話いたします。資料2の3ページの下段から4ページにありますとおり、第1回目の部会で医療施設調査に関して示された要検討事項について、厚生労働省から回答が行われ、おおむね適当とされました。ただし、「看護師数及び准看護師数の削除」を取りやめることに関しましては、報告者負担や利活用の観点で疑義があるため、厚生労働省において再整理の上、次回の部会で改めて審議することになりました。

最後に今後の予定についてご説明いたします。第3回目の部会ですが、これは一昨日2月22日火曜日に開催したところでございます。第2回目の部会で示された要検討事項及び答申で今後の課題とする事項についてここでは審議をいたしました。

第4回目の部会は3月15日火曜日に開催を予定しておりまして、そこでは答申案のとりまとめを行う予定でございます。

なお、第3回目の部会は一昨日開催済みではございますけれども、開催後間もないため、審議状況の報告は3月の統計委員会で行わせていただきたいと思っております。

以上です。

**○樋口委員長** ただいまの説明、ご報告につきまして、ご質問がございましたら、お願いいたします。

まだ、いろいろと課題が残っているようで、これについて説明者からございますか。

**○厚生労働省** ございません。

**○樋口委員長** よろしいですね。工程表を作るということもありましたので、よろしく願いいたします。

ほかにもございますでしょうか。

それでは、人口・社会統計部会の委員におかれましては、今後も引き続きご審議のほどよろしくお願いいたします。

それでは、次の議題です。匿名データ部会の報告につきまして、椿部会長からお願いいたします。

**○椿委員** それでは、お手元の資料3をご覧くださいければと思います。第5回匿名データ部会のご報告をさせていただければと思います。

日時はそこにありますように、平成23年2月7日、時間はかなり細かく書かれているところがございます。

第5回の匿名データ部会におきましては、前回の部会で提出されました委員の意見を踏まえて、論点を修正させていただきまして、それに基づいて審議を行った次第です。

なお、前回の統計委員会でご承認いただいたことですが、今回から2年前の第1回から第3回の部会で、匿名データに関する技術的な検討をしていただいた2人の委員に入っていたところですので。

初めに厚労省から前回の論点に沿った説明を受けた後、論点の項目ごとに審議が行われたところ

ですが、1 ページの下の「ア リサンプリング単位と地域区分の関係について」ということで、前回、そもそもリサンプリングの方法と地域情報の話を分離して議論することがなかなかできないということで、リサンプリングと地域情報などの識別情報の削除の両方は、いずれも基本的な匿名化技法であって併せて議論する。

この話と更に2 ページにある「イ 地域区分の提供について」という話も非常に密接に関連している話ですが、論点の整理自体を行った形になっております。

これに関して、今回の諮問では世帯員単位でリサンプリングされた匿名データを作っていなかったわけですが、部会の中ではこのような匿名データが、公衆衛生や疫学分野での有用性がある、必要であるということと、そもそも世帯員単位でリサンプリングをすれば地域情報を付加できる可能性があり、メリットもあるのではないかとということで検討が必要であるという意見がありました。一般的な関係を分析するときには、世帯員同士の影響を受けないように世帯員レベルでサンプリングされた匿名データの方が、そもそもバイアスも少なく望ましいという意見もあったところです。

ただ、今回の諮問の対象になっておりますデータに関して、二段抽出の方法自体は妥当であり、これによって結果として2 割のリサンプリングになったことはやむを得ない。これは今回の諮問に対する考え方になっております。今後、今のような議論があったことを受けて、公衆衛生・疫学分野におけるニーズ、世帯員単位の提供ならば地域情報を付加できる可能性があるというメリットの大きさを勘案して、この検討は続けていく必要であるとまとめさせていただいた次第です。これは私の方のまとめにあったところでございます。

それと、今、申し上げた密接関連ですけれども、今回、地域区分の提供を行っていないということでもあります。実は今回、厚生労働省にかなりご尽力いただいて、試算していただいて、「15 万人以上の市及びそれ未満の市及び郡部」という地域区分を付与した場合、リサンプリングによる最終的なサブサンプルがどれくらい減るかということの評価していただいた次第です。ただ、現実には「15 万人以上の市及びそれ未満の市及び郡部」という区分では、この種の情報を提供していただいたとしても、それほど有用性は見込まれないのではないかと意見が大半を占めました。

そこで、今回は、むしろ地域区分を付けないという形にして、今後、「三大都市圏」といったような、もう少し有用性が見込まれる地域区分の提供について、開示リスクということは非常に大きくなる可能性があるので、今後の課題として検討していただければという形でまとめさせていただいた次第です。

「ウ 世帯員の年齢について」ということを論点にしておりますが、これはトップコーディングに関わる基準になっております。前回の総務省の4 調査は0.5%基準だったのですけれども、今回は開示リスクを考慮して全体の上位1%に関してトップコーディングを行うということになっております。ただ、これに関しましても、現在、我が国の人口構成において一番急激に増加している85 歳から89 歳というところが上位1%に入ってしまう。その階級の情報については利用者のニーズが高いというご指摘があって、有用性の観点からいかなものかということがありました。今回の1%基準というのは余裕をもって、今後余りトップコーディングの閾値を変えたくないというご

意向があったということは重々承知しているところですが、これにつきまして、やはり将来的に人口構成が変化した場合には、適宜その閾値について見直すことが必要という形でまとめさせていただいた次第です。

同様に「エ トップコーディングの閾値について」、これは若干テクニカルな話になっているところですが、これにつきましてご意見をいただいたところです。今回、同一の調査票からAとBという2つの匿名データを作成することになっておりまして、その両者には分布に大きな差がなかったので問題なしという判断をいたしましたけれども、今回のように同一調査票から複数の匿名データを提供する場合には、基本的に元となる対象サンプルの分布が異なっていた場合には、トップコーディングの閾値を変えることが有用性、秘匿性の両方の観点から必要ではないかという意見があった次第です。これに対して厚生労働省の方からは、そもそもトップコーディングの閾値というのは母集団情報から決まるものであるから、そういうことは滅多にないのではないかというご意見があったところですが、トップコーディングの閾値が母集団自体から決められないときには、実際に作られた対象サンプルの分布に違いがあるかないかということは確認する必要があるのではないかとこのことをまとめさせていただいた次第です。この話は若干テクニカルな部分になりますので、どういう形でまとめるかということに関しては、今後最後の部会で検討したいと思っております。

3ページにいただきまして「オ 外観識別可能性の低い項目の秘匿措置の緩和について」ということです。特に健康票に関する項目では外観識別が不可能と思われる項目に秘匿措置がかけられているのではないかという意見がありました。これはそういう項目があるのかないのかということに関して、包括的にもう一度厚生労働省に確認していただくという作業を次回の部会までをお願いしたところです。

カでございますけれども、これは前回、樋口委員長からご指摘を受けたところですが、「カ トップコーディングした階級の平均値などのメタデータの提供について」ということで、これに関して平均値以外にもさまざまな要望が上がる可能性があるという指摘がなくなるといった意見もあったところです。また、この次の話の方が本質的だったのかもしれませんが、まだ本調査の利用実績がないから、こういう要望がどれぐらいあるか見極めた上で、必要なデータは匿名データの提供に併せて提供すればよいのではないかと。どちらかというところ、運用上で何とかできないのかというご意見でした。この点は匿名データの提供を進める中でニーズを見ながら判断すべきではないかという形で、まとめさせていただいております。

それから、有用性に関わることとして3ページの最後ですが「キ 所得票及び貯蓄票に関する事項について」ということで、ここではいろいろな意見があったところです。

まず、所得の内訳は分析者として非常に重要な情報であるということは皆さん方も認識を共有していたところです。ただ、今回、厚生労働省のデータに関しては、大変大きな情報量を持っているので、秘匿が開示されるというリスクの問題があり、安全性を優先するというところで、必要な措置として理解するけれども、むしろ今後、総務省なども含めて提供の方法を検討してほしいというご意見であります。



それから、匿名データを作成する場合、匿名化技法に関してさまざまな方法論があるわけです。ラウンディングのようなものとかカテゴリー化、その他もろもろ幾つかの方法があって、今やっているのはトップコーディングやリコーディングだけをやっているわけですが、さまざまな匿名化技法を使えば、秘匿が難しいデータについても提供を検討できるのではないかと。したがって、そういう方法を勘案して有用な情報を提供した方がいいのではないかとという意見がありました。

一方、そういうラウンディング等を行ってしまうよりは、できる限り現データに近い形で提供してほしいということで、むしろどのようにデータを加工するかということは研究者や研究目的によっても異なるので、加工は最小限にしてほしいという意見がありました。この意見は実は部会の中で両論出てきたもので、特にこれがまとまるという形は考えられませんでした。

したがって、とりまとめとしては両論併記の形になってしまっておりますが、開示リスクを考えると現段階での所得等の内訳を提供することは困難である。一方、ノイズの付加などこれまで匿名化措置として採用してこなかったものを用いてでも内訳項目を提供するか、それともこれまでどおりのトップコーディングなどのみで対応し、できるだけ個票に近い形で提供していくのかは、今後の研究課題として整理させていただいた次第です。

「ク 外観識別可能な項目について」です。世帯人員が8人以上の世帯、3つ子以上の世帯、年齢差の大きい夫婦について削除することは適当であるということを確認いたしました。この際、専門委員から併せて年齢差の大きい親子についてもそういう必要があるのではないかとということで、これも厚生労働省の方に確認をお願いした次第です。

「ケ 統計調査以外の情報との関係について、本人が識別できる場合について」ということでございます。匿名データと外部情報との照合の可能性に関しては、前回の部会以来、一般人基準と特定人基準、どのような情報を外部から使えるかということによって、どの基準を使うべきかということがかなり論点となっていたところです。

まず、法制度を所管しております総務省政策統括官室から、匿名データは利用者や利用目的が限定した運用をされており、また開示請求に対しても不開示。これらを踏まえれば、現時点では外部情報との照合の可能性については一般人基準で運用することが適当と考える。今後、利用者や利用目的が拡大することになった場合には、特定人基準で考える必要があるかもしれないという、この問題に関してかなり重要な説明があった次第です。

一方で、他者からは秘匿されていても、本人が自分のことだと識別できる場合にはどうか。調査の影響を考慮すべきではないかという点についても、本人には識別できるからといって制限してしまうと、匿名データという制度が使い物にならなくなるという意見が大勢を占めました。現時点では研究者がそれなりの誓約を行った上で使う状況においては、特定人基準として厳格に運用する必要はない。また、本人自身が識別できる場合についても、現時点では二次利用は利用者や利用目的の範囲が限定されていること、そもそもリサンプリングという秘匿措置を行っており、原理的には完全には特定できないことから、そこまで基準を厳しくする必要はないということでまとめさせていただきました。ただし、今後、二次利用の利用者の範囲が拡大した場合には、この基準を厳しくする可能性もあるという形でまとめさせていただいた次第です。

「コ 提供時期について」ですけれども、本調査は3年周期なので、5年経過後の提供とすると直近の結果が出ないという不都合が発生するかもしれないということで、提供時期について議論すべきではないかという意見もありました。これについては、本調査は先行する総務省の4調査よりも開示リスクが大きいということで、現時点では5年経過後で妥当といたしました。有用性の観点から提供期間を短くすることに関しては、今後の課題として整理するつもりであります。

最後に「サ その他の意見について」です。介護票の匿名データ化です。これも有用性の観点から提供ニーズが非常に大きいことは理解できる場所ですけれども、介護票自体のデータ数が大変少ないこともあって、匿名化技法の検討状況を見る必要があるのではないかとというまとめ方しております。

また、後続調査との関係については、今回、地域情報はすべて削除することになっているなど、開示リスクは十分に低くなっているということで、問題がないとまとめさせていただきました。

なお、匿名データ部会は次回が最後の審議となります。3月8日火曜日に開催いたしまして、宿題としております課題の審議を行った後に答申（案）を審議することになっております。来月の統計委員会には答申（案）をご報告させていただく予定となっております。

匿名データ部会の報告は以上でございます。

○樋口委員長 ありがとうございます。

それでは、ただいまのご報告につきまして、ご質問等がございましたら、お願いいたします。

安部委員、どうぞ。

○安部委員 この調査はこれまでに試行的提供がありませんでした。以前に試行的提供がありました、総務省の4調査だったか、3調査だったか覚えていませんけれども、これらについては、試行的提供と比較して、一部分は利用者から見れば使い勝手がよくなり、一部分不便になったものが現在の匿名データになっているというのが私の理解です。

例えば、私自身の経験でいいますと、試行的提供でも利用して、良くなった点も悪くなった点もあるけれども、また利用したいという気持ちになったのです。この調査は試行的提供の実績がないので、しかも、私自身の目から見てどうかと思うのは、リサンプリング率が非常に低くなってしまったので、そもそも使いたいと思う人がどのぐらいいるかということが予測しにくい面があるような気がするのです。それで、利用の方による評価を調査するような機会を設けることは考えておられるのでしょうか。

○樋口委員長 お願いします。

○椿委員 二次利用全般に言えることですが、匿名データに限らず、いわゆるオーダーメイド集計についてもどれぐらい利用があったかということは、この統計委員会に毎年報告される形になっているのではないかと認識しているところです。今回の匿名データは大変大きな情報量を持っていますけれども、今回初めて提供されていくものですので、安部委員がおっしゃられたように私どもも正直予測はつきません。ただ、今おっしゃられたような意味での利用の実績に関しては、注視しておくべきことであることは当然だと思っています。

○樋口委員長 よろしいですか。

どうぞ。

**○厚生労働省統計情報部** 私はこの個票を使った分析のことはよく分からないものですから、お伺いしたいのですが、こういう世帯調査の場合、もともとの母集団があつて、そこから一定の率でサンプリングをされていて、そこからリサンプリングしているわけです。本質的に問題になるのは、私はリサンプリング率ではなくて、母集団に対してリサンプリングされたものがどれぐらいの割合であるか。あるいは再度抽出された匿名標本データの標本サイズが重要なのではないかと思うのですが、リサンプリング率というのはそこまでこだわる必要があるのでしょうか。

**○樫委員** 技術的にはその種の検討がされているということをもっと申し上げたいと思います。これは厚生労働省の立場を擁護するような形になりますけれども、これは20%のリサンプリングを行ったのではなくて、結果としてそうなった。この調査ではクラスター抽出という非常に特殊な抽出を行っている、まず地域を抽出した後、その地域に対して悉皆調査をするので、匿名データの中にどの地域が入っているかということが分かってしまうことを厚生労働省が極めて恐れた。そのために地域情報もすべて残さない形でリサンプリングするという措置を、部会では妥当と認めた。

一方で、非常に人口の大きい地域、小さい地域という形になってくると、先ほどの全国全体で乗率を一本化する、ウェイトを一本化する。都市部と非都市部との乗率を全て一致させることによって、ある地域の実際のリサンプリング率が極めて低下してしまったことによって、結果としてこうなっているということで、リサンプリング率を2割にするということを我々は認めているわけではなく、一連のプロセスを審議する中で妥当としているということです。

この内容を詳細に統計委員会の中で報告するとなると、秘匿措置自体が分かってしまう。これは秘匿の開示リスクが増えてしまうということで、若干テクニカルな部分の説明を公開議事録の中には入れてなくて、詳細なものは部会の中で報告されているという形になっています。大変申し訳ないのですが、ご容赦いただければと思います。

**○樋口委員長** 何かございますか。

**○厚生労働省統計情報部** 方法論についてはよく分かるのです。研究者の皆さんの立場として、私はリサンプリング率というよりも、元の標本の抽出率、例えば国民の半分ぐらいは抽出してあるというものから10%をリサンプリングしようが、もともと10%の標本から半分をリサンプリングしようが、リサンプリング率としては10%と50%で大きく差があるのですけれども、それはリサンプリング率だけを評価するのではなくて、元のサンプリング率との評価が必要ではないのだろうか。

更に標本理論からいえば、リサンプリングされた後の標本サイズそのものの方がむしろリサンプリング率より重要なのではないかと考えているのですけれども、そうではないのかということをお伺いしたかったわけです。

**○樋口委員長** どうぞ。

**○樫委員** 申し訳ありません。その意味では、今回作られている匿名データに関して、サブサンプルのサイズは一般的な計量経済的な分析には耐えるものになっていると考えます。

それから、もう一点、当然、この審議の中でチェックしていることは、もともと厚生労働省が持っているデータが示しているさまざまな分布と、こうして作った匿名データから推計される分布と

が比較的一致している。あるレベル以上乖離していないということをチェックしているところです。勿論、これは母集団情報ではなくて、もともとの標本情報が一致していることにすぎませんけれども、そこでの乖離が非常に少ないということ、それから、十分なサブサンプル数は確保しているという形で有用なデータではないかと判断しているところです。

○樋口委員長 二段階無作為抽出になっていないところが最初のいろいろな問題となっているわけで、サンプル数をみても、それは個々のサンプル数は多いにしても、抽出された地域の数も限られている可能性がある。

○椿委員 特定の地域が選ばれている可能性があるので、逆にいえば、地域情報に関しては非常にナーバスになっているということです。

○樋口委員長 もともと、この統計調査は行政記録情報であったという、過去何十年間の経緯があって、このような設計になってきているのだらうと思いますが、ある意味では二段階抽出ではない方法を用いることによる限界が今回の議論の中でいろいろと出てきているのだと理解しますが、それでよろしいのでしょうか。

○椿委員 基本的には今のクラスター抽出、集落抽出というのが極めて特徴的な設計であると判断しています。

○樋口委員長 よろしいですか。安部委員、どうぞ。

○安部委員 これは利用者がどういう目的を持っているかによるので、私の意見が利用者の中で一般的かどうかはわからないということをお断りした上で、やはり属性別の集計などをしようとしたしますと、サンプル数というのは重要なことで、必ずしも抽出率ということではなくて、提供される匿名データのサンプル数によって、利用したいかどうかの意向が変わるのではないかと思います。

ただ、それだけではなくて、ほかにも幾つも論点が出ていることも含めて、やはり利用者にとって利用したいと思うデータかどうかというのが予測のつかない部分があります。勿論、利用実績を把握していくわけですけれども、匿名化の手法が一定の方法だったことによって、利用が少なくなってしまうということもあるかもしれないので、そのようなことに関して、例えば利用者から意見を聞くとか、そういうことも含めてという意味ですが、検討する可能性は考えていらっしゃるのかということです。

○樋口委員長 お願いします。

○椿委員 なかなか難しい問題ですけれども、実際にこの匿名化データを使っていた利用者がどう考えているかということは非常に重要な問題だと思います。現時点では先ほどの総務省の4調査と違って、我々の部会の態度は保守的にならざるを得ないということになっておりますので、安部先生がおっしゃられたように、これが有用性という観点から非常に不満が出てくることはあり得ると思います。

でも、一方で匿名データ部会としては、開示リスクという問題の正当性を議論することがミッションと考えておりますので、むしろ議事概要に幾つかありましたけれども、匿名化技法に関するテクニカルな委員というか、私とよく似た統計方法論側の委員からすると、特定の方法論自体の慣用

性がないか。一方で、使う方からすれば、それは有用性の減少であるという形で、今後ユーザーの声とそれに対してどういう技術を提供できるかということに関してのトレードオフというのは、比較的大きな問題として今後も継続するのではないかと思うのが私の私見です。

○樋口委員長 私も一利用者というか研究者から、いろいろトップコーディングについてご審議いただきまして、ありがとうございます。

トップコーディングあるいはボトムコーディングにつきまして、各項目の上位1%という基準で、秘匿するということは匿名データの安全性を確保するということから必要なこと、あるいはいたしかたないことだと理解します。

研究者としてこれを使って学術論文を書こうとなったときに、必ず求められるのが標本の「基本統計量」を記述しろということです。これはおおよそのアカデミックペーパーであれば、あるいはジャーナルであれば要求されるだろうと思います。ところが、今のやり方ですと、必ずしもサブサンプルについての平均値を提供すること、例えば、年齢であるとか、所得であるとか、これらトップコーディングが検討されている項目には、トップコーディングの閾値を入れることになっていると思いますが、この間も申し上げて繰り返して恐縮ですが、トップコーディングについて1%以上のところの平均値がどうなっているかということが出てこない、「基本統計量」を示すことができないということが起こって、結果的に学術論文に匿名データというのが利用できないのではないかと懸念されるのですが、その点について可能であればもう一度ご審議いただきたいと思います。

いかがでしょうか。

○椿委員 基本統計量、特に平均値を出すためにはトップコーディングした閾値より先の部分の平均値の公表が必要になります。標準偏差でしたら2次のモーメント情報が必要だということを知っていますけれども、少なくとも基本統計量が算出できるという観点に絞って、技術的にはできる話ですから、開示のリスク等の問題があるかどうかということについては、次回の部会でもう一度検討させていただこうと思います。

○樋口委員長 よろしく願いいたします。

ほかにいかがでしょうか。

よろしければ、まだ議論が続くかとは思いますが、椿部会長を始め匿名データ部会の委員の皆様、今後ともご審議のほどよろしくお願いいたします。

それでは、本日の議題は以上ですが、総務省政策統括官から連絡があると聞いておりますので、よろしくお願い致します。

○総務省政策統括官付統計企画管理官 本日お配りしております資料の一番下に、赤白の1枚紙があると思いますので、それをご覧いただきたいと思います。

これは政府統計の統一ロゴデザインの募集のお知らせでございます。政府統計の統一ロゴですが、「公的統計の整備に関する基本的な計画」の中にございます統計に対する国民の理解の促進の施策の一環として作成するものです。このロゴによりまして、政府が実施するすべての統計調査について、それが政府の統計調査であること、調査に安心してご回答いただけることを調査の対象となった方々にお示ししたいと考えています。

統一ロゴのデザイン決定後は、政府が実施するすべての統計調査の調査票や封筒などにこのロゴが印刷されることとなります。これは平成24年4月からの使用を予定しております。

このロゴの策定に際しまして、2月1日から3月15日まで広く一般の方からデザインを募集しているものです。応募対象に制限はございませんので、委員の先生方でも、また各府省の皆様方も広く応募が可能となっております。より多くの方々から募集を募りたいと考えておりますので、お知り合いの方々などに対してご周知いただければと思います。

どうぞよろしくお願いいたします。

**○樋口委員長** どうぞよろしくお願いいたします。このロゴが付いていない統計は、たとえ政府が実施していても、あるいは各部局が実施していても、政府統計とは認めないという趣旨もあるそうですので、すばらしいロゴをご用意いただけたらと思います。品質証明というような役割もこれで果たしたいということですので、よろしくお願いいたします。

それでは、最後に次回の日程につきまして、事務局から連絡をお願いいたします。

**○乾内閣府統計委員会担当室長** 次回の統計委員会につきましては、3月17日木曜日15時から、本日と同様にこの会議室において開催いたします。詳細につきましては、正式な開催通知をもってお知らせいたしますので、よろしくお願いいたします。

**○樋口委員長** それでは、本日の「統計委員会」は以上で終了します。どうもありがとうございました。