

調査の新しい潮流

—ESRA で得た知見から考察する—

New Trend in Surveys: Considerations from Acquired Findings at ESRA Conference

江口 達也
Tatsuya Eguchi

1. はじめに
2. Mixed-Mode 調査
3. Push2Web
4. 確率パネル
5. オンライン調査でのモバイルデバイスへの対応
6. 非確率オンライン調査のデータ品質
7. 混合デバイス調査
8. モバイルデバイスでの新しい測定機能
9. パッシブモバイルデータの収集
10. ビッグデータと調査
11. おわりに

〈要旨〉

2017年7月に行われた European Survey Research Association (ESRA) のカンファレンスでは、インターネットを利用した調査における工夫や課題について多くの報告が行われた。Dillman が提唱する Push-to-Web を用いることにより、インターネットを利用した Mixed-Mode 調査でも回収率を高めることができる。カナダの国勢調査ではネットによる回収が約7割を占めた。オンライン調査におけるモバイル端末への対応に関する工夫も報告された。画面サイズが小さいモバイル端末に不向きなグリッド形式の質問の欠点を解消するため、de Leeuw は Horizontal Scrolling Matrix (HSM) 形式を、Barlas らはアコーディオングリッド形式を提案している。また、オンライン調査では、スマートフォンによる回答が増えており、質問文や画面レイアウトなどは「モバイルファースト」で作成すべきだという意見もあった。本稿は、日本におけるインターネットを活用した調査方法論の発展に資するため、これらの知見を整理して論じるものである。

At the European Survey Research Association (ESRA) conference held in July 2017, there were many reports on contrivances and issues with regards to Internet surveys. By using the Push-to-Web for which Dillman advocated, it is possible to increase response rates using the Internet, even with Mixed-Mode surveys. In the Canadian census, Internet responses accounted for approximately 70% of responses. Contrivances that used mobile devices in online surveys were also reported. De Leeuw proposed the Horizontal Scrolling Matrix (HSM) format and Barlas et al. proposed the accordion grid format to solve the weak points in grid format questions unsuitable for mobile devices with small screen sizes. Furthermore, smartphone responses increased in online surveys. Therefore, there were also suggestions that questionnaires and screen layouts should be created with "Mobile First." We organize these findings to discuss and contribute to the development of Internet-based survey methodologies in Japan.

1. はじめに

2017年7月17～21日に、ポルトガルのリスボンで European Survey Research Association (ESRA)の第7回カンファレンスが開催された。今回のカンファレンスでは、欧州を中心に850人を超える研究者が参加し、延べ187のセッションが開かれている。ESRAは2008年に設立された組織であり、2年に1度カンファレンスを開いている。発行している学術誌「Survey Research Methods」には、それらの研究成果も掲載されている。

本稿は、筆者がこのカンファレンスで得たインターネットやモバイルを利用した調査、Mixed-mode 調査(複数の調査手法を組み合わせてより良い調査結果を目指す)の工夫など最先端の事例や知見を整理し、日本の調査法開発に資するよう論じる。なお、本稿では敬称を省略する。

2. Mixed-Mode 調査

カンファレンス初日の17日午後、ユトレヒト大学のEdith Desiree de Leeuwによる基調講演が行われた。講演のタイトルは「Never A Dull Moment : Mixed-Mode Surveys In Past, Present & Future」。Mixed-Mode 調査についての基礎から最新事情までを論じた。各要点をde Leeuwの語り口に沿って整理する。

2-1. Mixed-Mode 調査とは

Mixed-Mode 調査と一口に言っても、その形式にはいくつか種類がある。例えば、

- 〈1〉 対象者への接触を異なるモードで行う
- 〈2〉 特定の質問について、全ての対象者に異なる回答モードを与える
- 〈3〉 異なる回答モードを、異なる回答者のグループに与える

の3つに大別できる。〈2〉は「センシティブな質問については自記式で回答してもらおう」というケースが当てはまる。また、国際調査を実施する場合、国によって調査上の制約が異なるために〈3〉の形になる場合があるだろう。後述するPush2Web (Push to Webのtoを2に置き換えて1語で表記されることが多い)も〈3〉に該当する。

では、なぜMixed-Modeなのか。Mixed-Modeがもたらす効用としてde Leeuwは以下の4点を挙げた。

- ①カバレッジの改善
- ②回答を増やす
- ③コストを減らす
- ④より良い測定の可能性

2-1-1. カバレッジの改善

「良いカバレッジ」とは、サンプリングフレームに関与する集団の全てのユニットが含まれている必要性を意味している。例えばWeb調査を考えると、インターネット普及率が高まってきているとはいえ、国によって違いがある(図表1)。北米では88.1%、ヨーロッパで77.4%などとなっているが、ヨーロッパの中でも国によって100%から43%とかなり幅広い違いがある。

図表1 世界のインターネット普及率
(引用: Internet World Stats 2017年Q1時点)

北アメリカ	88.1%
ヨーロッパ	77.4%
オーストラリア/オセアニア	68.1%
ラテンアメリカ/カリブ海	59.6%
中東	56.7%
世界平均	49.7%
アジア	45.2%
アフリカ	28.3%

また、カバレッジが十分高くても、デジタル・ディバイドは存在していて、年齢や学歴、性別で違いがある。時間の経過とともに減少するだろうが、まだ偏りが存在しているのが現状である。

こうした状況に対する解決策として、一部の対象者に対して異なる回答モードを与える

GESIS パネル

ドイツの研究機関 GESIS が運営する確率ベースのMixed-Mode パネル。約4700人のパネルリストが登録されている(2017年11月時点)。登録者は、ドイツに住む18歳から70歳(採用時)のドイツ語を話す人々。

パネルリストは個別面接方式で募集され、続いて自記式でプロフィール調査が行われた。回答モード(オンラインまたは郵送)は参加者自身が選んでいる。

GESIS パネルを使って集められたデータは、学術研究者は無料で二次研究の実施などに使用することができる。

Mixed-Mode 調査が考えられる。例えば、ドイツの GESIS パネルでは、オンラインと郵送の Mixed-Mode 調査が採用されている。

2-1-2. 回答者を増やす

ノンレスポンスは国を超えて時とともに増加している。回収サンプル数が少なくなり、有効1票あたりのコストが高くなっている。また、回答者と非回答者との間で、主要変数に違いがある可能性がある。これが「ノンレスポンスバイアス」である。

こうした状況に対する解決策として、シーケンシャル Mixed-Mode アプローチがある。この手法では、非回答者に対して、コスト的に手軽なものから順に異なる回答モードを提示していく。例えば American Community Survey では、郵送(紙またはオンライン)、電話(CATI)、面接(CAPI)が回答モードとして順次提供されている。このように順次回答モードを与えていく方式が最善である。回答者に回答モードの選択機会を与えると回収率は下がるのでおすすめしない。

2-1-3. コストを減らす

イギリスの Understanding Society Innovation Panel で、CAPI のみの場合と Web と CAPI の Mixed-Mode の場合を比較する大規模な無作為実験が行われた(Bianchi, et al., 2017)。

その結果、Mixed-Mode グループは CAPI のみのグループよりも脱落が早いですが、時間の経過とともにその差はなくなり、累積的な違いは両者の間でなかったという。そして、コストに関しては Mixed-Mode のほうが節約することができた。

Understanding Society Innovation Panel

エセックス大学の社会経済研究所(ISER)が中心の研究プロジェクト”Understanding Society”に付随する確率パネル。この研究プロジェクトに関連する方法論的な実験研究のためのテストベットとして研究者が利用できる。1500世帯が登録されており、インタビューは基本的に面接方式(CAPI)で行われている。

2-1-4. より良い測定の可能性

Mixed-Mode 調査にすればなんでも良い測定になるわけではなく、それは場合による。例えばセンシティブな内容を聞く質問は自記式のほうがよ

り本音に迫れる可能性がある。また、移動経路を尋ねる場合には、モバイル端末を通じてGPS信号を使って取得したほうがより精度の高い測定が可能となる。

2-2. ” Mix” することは ” Design” すること

de Leeuw は、Mixed-Mode 調査のための調査票デザインの重要性について、「データ収集モードを Mix することで、カバレッジエラーやノンレスポンスエラーを低減できるというメリットがあるし、メジャーメントエラーが改善する可能性もある。だから、Mixed-Mode 調査の設計(design)をしなければならない」と強調した。

Mixed-Mode 調査では、異なるモード間で「同等な調査票」を作成する必要がある。各モードで異なる質問形式の伝統を持っているため、そのまま Mix するとメジャーメントエラーを増大させてしまう。例えば、電話や面接などインタビュー形式の調査では「わからない」という選択肢は回答者に明示されていないが、Web 調査など自記式の調査の場合は明示されている。ウェブデザイナーが調査画面を作ろうとすると、「わからない」を省略して強制的に選択させるか、「わからない」を明示することになるだろう。そして、その「わからない」は恐らく他の選択肢から少し離して配置される。こうしたことは、経験的証拠に基づけばまったく賢明ではない。モード間で異なる質問形式にならないように、最適な等価性を目指す必要がある。

そして、調査票の等価性に関するもう一つの例として、グリッド(マトリクス)質問を取り上げた。de Leeuw は、グリッド形式の使用を避けるべきだと主張した。とくにオンライン調査でよく見られる形式だが、ストレートライニング(全ての項目で同じ選択肢を選ぶこと)や答えにくいという回答者の調査に対する不満の観点から欠点があるし、自分の回答が表形式で見比べやすく前後及び全体の回答に影響を及ぼすコンテキスト効果が発生すると指摘した。また、画面サイズが小さいモバイル端末では、長いグリッド質問は難しい。一方で、グリッド質問を単問に分解すると、「次へ」ボタンをクリックする回数が多くなり回答者の負担が増す。こうした問題点の解決策として、de Leeuw はグリッド形式に代わる「HSM(Horizontal Scrolling Matrix)形式」を提唱している(図表2)。この仕組みは、「一つの画面に表示する質問項目は

一つだけで、提示される選択肢はそのまま変えない。最初の質問項目で回答した（選択肢を選んだ）ら、画面の質問項目部分のみが自動的に次の質問項目に替わる」というものである。インターネットでよく見られるバナー広告（カルーセル広告）のようなもので、一定時間ごとに表示内容が替わっていく感じと似ている。

2-3. 決して退屈な瞬間はない

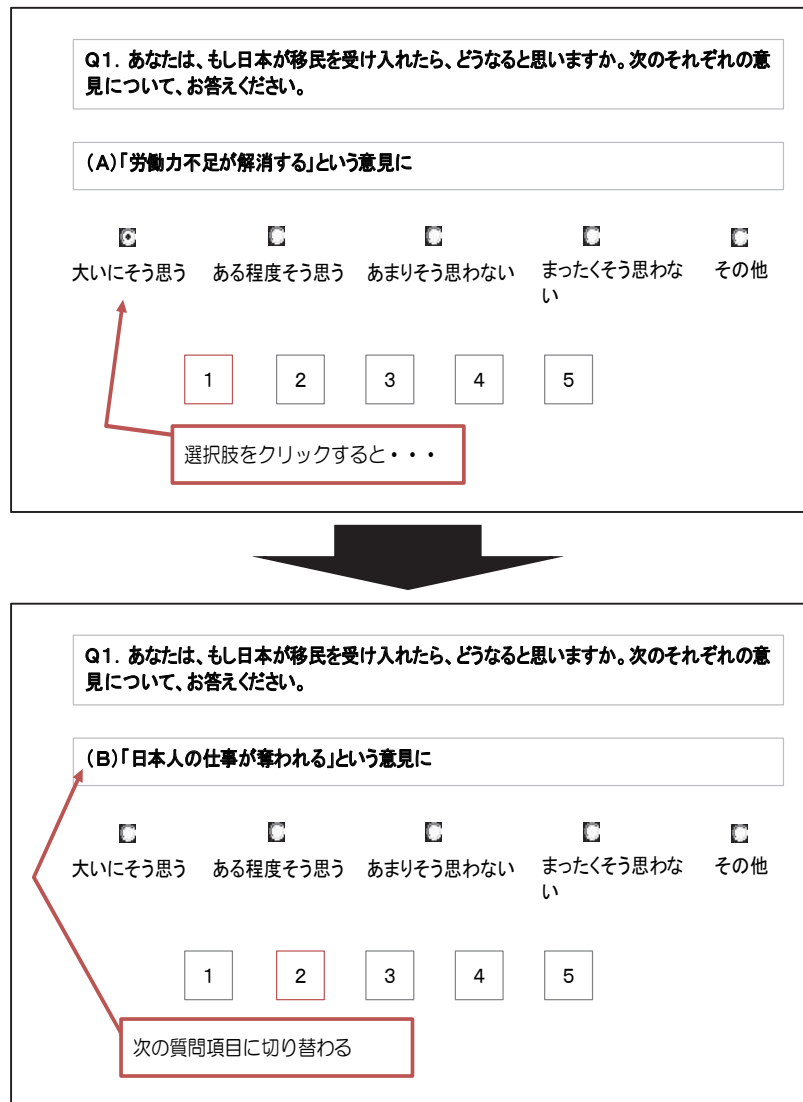
de Leeuw は、「社会や技術の変化に応じて、我々の調査の方法論や統計学もまた変化していく必要がある」と指摘する。これからの研究課題として、マルチデバイス調査や特定のデータ収集ツールとしての携帯電話への対応、そして、ビッグデータ、ソーシャルメディアなど非確率サンプルと確率サンプルのデータ融合など課題は多く、私たち調査者には「決して退屈な瞬間はない」と論じている。

2-4. Mixed-Mode 調査の論点

近年、日本でも Mixed-Mode 調査の実例が見られる。代表的な例としては、2015 年国勢調査においてインターネット→郵送→面接のシーケンシャル Mixed-Mode アプローチが採用された（東京都は例外的に紙とネットの同時依頼）。世論調査の事例としては、毎日新聞が郵送世論調査を 2016 年から郵送・オンライン同時依頼の Mixed-Mode に切り替えている（大隈・原田, 2017）。de Leeuw は複数の回答モードを同時に回答者に与えると回収率は下がると指摘している。毎日新聞のケースでも回収率は従来平均よりも約 3 ポイント低下しており、日本でもこの指摘は当てはまる。

日本の調査の回収率は、面接調査や電話調査で大きく低下しているが、郵送調査では工夫次第でなお 7 割以上を得ることが可能である。郵送調査が高回収率を維持できている間に、「調査票の等価性」などの Mixed-Mode の研究課題に対処しておく

図表 2 HSM 形式のイメージ



べきである。面接調査や電話調査が大きく劣化している現状では、郵送調査を基準としたベンチマーク調査とせざるを得ないからである。

一方、de Leeuw が提案している HSM 形式だが、たしかにグリッド形式の質問よりも回答負担は少ないように思える。オンライン調査で使用するにはよさそうだが、Mixed-Mode 調査での使用は、調査票の等価性の面で疑問がある。例えば、郵送とオンラインの Mixed-Mode 調査を想定すると、紙の調査票には当然一覧性があるため、コンテキスト効果の発生は避けられない。しかし、オンライン調査で HSM 形式を使用すると、それとは同等な調査票とならない可能性がある。バイアスを無くすることだけでなく、バイアスを等価にすることも重要な課題であろう。

3. Push2Web

Web を利用する調査法の研究発表が多いESRAのカンファレンスでは、Mick P. Couper と並んで Don A. Dillman も巨匠 (Legend) と呼ばれている。とくにヨーロッパの政府、学術研究機関が評価しているのは、Dillman が提唱した TDM (Tailored Design Method: 高回収率を得る郵送調査法) とこの郵送調査のノウハウを生かした Push2Web 調査である。Push2Web (Push to Web) は、伝統的な接触方法 (郵送法など) を用いて対象者がオンラインで Web アンケートに回答するよう促す手法である。セッションの冒頭、Dillman 本人から Web-Push メソッドを機能させる 6 つのポイントについて発表があった。

3-1. 対象者への最初の接触は郵送で

フィッシングやマルウェアの影響でインターネットリンクに対する恐怖心が高まっており、この不信感が Web を利用した調査に回答することの大きな障壁になっている。一方で、飛行機やホテルの予約から百科事典に至るまで、世界中の情報サービスの利用にはインターネット接続がほぼ必須となっている。Dillman は郵送調査の回収率を高める方法である TDM の開発で社会的交換理論を採用しており、ここに調査対象者との「信頼」関係を構築することが盛り込まれている。Push2Web においても、いきなり Web で調査依頼するのではなく、最初の接触は郵送で行い、手紙の中で「調査主体の住所」「電話番号」などの連絡先も知らせている。これにより高まった調査主体への信用が、Web での回答を誘発する。

3-2. 郵送ならインセンティブを入れることが可能

郵送した調査依頼の手紙でコンピューターの前に行くように調査対象者に求めても、調査対象者にとっては過酷な切り替えである。「手紙を読んでいる」という現状から「パソコンの所に行き電源を入れ、調査画面へアクセスする」という行為が伴う。謝礼を渡すことが、この行為の切り替えを促すという実験結果がある。

世帯調査での実験で、調査依頼に 5 ドルの謝礼を入れると対象者の 32% が Web から回答したが、謝礼なしの場合は 13% にとどまった。懸賞型の謝礼や後渡しの謝礼では効果的な代替にはならないが、先渡しの謝礼に加えて後渡しの謝礼を渡すこ

とは有効である。

Push2Web で最初に郵送で依頼することを推奨するのは、インセンティブを同封して回答意欲を高められるからである。

3-3. 応答速度の向上のため複数回の督促が必要

複数回にわたる督促の重要性は、電話や郵送といった全ての調査モードにおいて共通である。しかし、どのモードでも度重なる督促は拒否や無視されやすくなる。もし異なる連絡先情報 (接触手段) が利用可能なら、督促は違うモードで行うのがより効果的である。また、電子メールによる督促は回答者を Web に移動させるのにとくに役立つ。

Dillman は複数のモードで督促を行った例として、大学院生を対象にした調査を挙げた。調査の流れは以下のとおり。

- 1 日目 … 2 ドルの謝礼とともにウェブで回答するよう依頼状を郵送
- 3 日目 … 電子メールで 1 回目の督促
- 8 日目 … 電子メールで 2 回目の督促
- 16 日目 … 紙の調査票を郵送
- 21 日目 … 電子メールで 3 回目の督促

こうした手順で、10 時間で 21%、5 日間で 40% の回収率を達成している。最終的な回収率は 77% だった。後から郵送で紙の調査票を送付したが、これには Web 回答も増やす効果があった。紙の調査票の送付後、回収率は 12 ポイント増加したが、そのうちの半分が Web 回答である。

3-4. 異なる回答モードを順次提供して回収率を高める

郵送でインターネットを通じて回答するのを送り、異なる回答モード (郵送、電話、個別面接) を提供する主な目的は、回収率を大幅に向上させることである。例えば 2016 年のカナダ国勢調査では、全世帯の 68% がインターネットで回答、20% が郵送、10% が面接で回答している。

3-5. 異なる回答モードを提供することによりノンレスポンスエラーを低減させる

繰り返し行われた世帯調査で、紙で回答する人には低学歴で収入が低く、高齢者層が多いことがわかっている。もし電話が利用可能なら、フォローアップコールを行うことによってこうした層の

回答を促し、ノンレスポンスエラーを低減させられるかもしれない。国勢調査での個別訪問も同様の効果をもたらしている。

3-6. 異なるモード間でデータを結合する場合、「聴覚」対「視覚」のコミュニケーションが回答にどう影響するか理解し、その知識を利用して全ての調査モードで共通の測定値を得る必要がある

これは、同じ質問構造と同じ言い回しを維持することを意味している。しかし、スマートフォンの普及によって、これはとくに難しい課題になっている。

「視覚」(郵送やWeb調査)については、質問の文言以上に評価しなければならないことがある。以下の4つが質問の意味に関わってくる。我々は、これらすべてを評価する必要がある。

- <1> 数字 (例: 1, 2, 3)
- <2> 記号 (例: →)
- <3> グラフィカルな表示
(サイズ, 形, 一貫性など)
- <4> 言葉

また、その他の例としては、グリッド質問を使うか、個別質問に分解して聞くか、ドロップダウンメニューを使用するか、選択肢を全て表示するか、によっても違いが出てくる。

ベストな質問を提示したとしても、スマートフォンからの回答を得ることは、より大きな問題になっている。スマートフォンでは、調査のお願いは数語読んでスワイプされてしまう。電子メールの正当性を容易に確認することはできない(スマホ画面では受信メールリストでタイトルと送信者名は見られるが、パソコンで通常確認できる①具体的なアドレスなどを含むヘッダー情報、②迷惑メール警告、③ウイルスチェック結果が表示されないことを指摘していると思われる)。対象者は起きてから寝るまでスマホを操作しているにもかかわらず、回答の可能性は高くない。

3-7. ムーアの法則再訪

CPUの性能が2年ごとに倍になるというムーアの法則はまだ終わっていない。CPUの代わりに、情報を格納する記憶媒体、コンピューター間のネットワークシステム、タスクを実行するためのソフトウェア、センサー、などが発展している。こ

れらの変化が調査に与える影響はかなり大きいと思われる。技術の変化に対する人間の反応は、技術よりもゆっくり変化する。我々の調査デザインの難問は、回答者の先に行くことなく、遅れてもいないデータ収集方法を開発することである。この懸念は、近い将来、スマホを含むMixed-Modeの時代に、我々のもとに訪れるだろう。

3-8. 電話によるPush2Web

ギリシャのテッサロニキ・アリストテレス大学のEvangelia KartsounidouとIoannis Andreadisは、2015年のギリシャ国民投票調査において、電話調査でウェブ調査の回答者をリクルートする実験を行った。いったいどのくらいの人が電話でメールアドレスを提供してくれるのか。また、提供してくれた人のうち、どのくらいの人がオンライン調査に回答してくれるのか。そして督促が回収率に与える影響も調べた。

通常の電話調査の場合、1件あたり平均45分かかるが、メールアドレスの聴取のみであれば1件あたり平均2分に短縮することができた。そして18.69%の人がメールアドレスを提供し、そのうち37.07%がオンラインアンケートに回答した。全部で6回の督促を行ったが、効果があったのは最初の2回だけだった。通常の電話調査をPush2Web方式に切り替えることで、リソースの39%を節約することができた。

3-9. Push2Webの回収率を最大化するには

イギリスの社会調査研究機関であるNatCenにおいて主に分析とメソドロジー開発で活躍した後Ipsos MORIへ移籍したGerry Nicolaasは、Push2Webで回収率を最大化する戦略について発表した(図表3はNicolaasが示したスライドを筆者が翻訳したもの)。

Nicolaasは、実験結果よりPush2Webの回収率にプラスの影響を与えるのは以下の項目であると強く確信している。

- <1> 個人宛てにすること
- <2> コミュニケーションの形式, 長さ, 複雑さ
- <3> 謝礼の使用
- <4> 複雑な質問や認知的に難しい質問は避けること
- <5> デバイスに依存しないこと
- <6> Webで答えない対象者に代替モードを提供すること

そして、Push2Web の今後の研究課題として以下の項目を挙げている。

- 〈a〉 インセンティブの種類
- 〈b〉 郵便物の外観と種類
- 〈c〉 郵便物の中身の視覚的デザイン
- 〈d〉 督促の回数や間隔など
- 〈e〉 オフラインからオンラインへ移行する労力の軽減方法

3-10. Push2Web 調査の論点

カナダでは、2016 年国勢調査で Push2Web を用いて、インターネットを通じた回答が約 7 割という驚異的な成果をあげた。総じて回収率が低い世論調査でこの水準に到達するのは難しいかもしれないが、Push2Web はインターネットによる回答を高める方法として効果的な手法である。しかし、Dillman が公開している実験結果をみると、TDM で行われた郵送調査よりも Push2Web で行われた調査のほうが平均して 10%ほど回収率は低くなっている (Dillman, 2017a)。なぜ回収率が下がる調査手法を研究するのか。この疑問に対して Dillman は「将来のカルチャーの変化とコスト対策について考えているからだ。インターネットをうまく利用できれば、さらに安く速くできる。まさにいまなら郵送のみでやったほうが回収率は高いが、インターネットをうまく利用する方法がないか探るのが目的なのだ」と答えている (松田, 2009)。

Dillman は回収率が下がるというデメリットに対し、TDM と Push2Web の回収票の属性構成比 (性別や年代などの構成比) を比較して、ほとんど違いがないことを説明している (Dillman, 2017b)。しかし、そうした属性面だけでは捉えられない偏りが発生する可能性が残るため、属性面以外で違いが発生していないかの検証も求められる。

Dillman は今回の発表の中で「私がかつて考えてきたアイデアはとてもよいものだった

が、変わる必要があった。私はこれまでのキャリアの中で変わること慣れてきた。だから、今日述べたアイデアも変わる予定でいる」と述べている。Push2Web は完成された手法ではない、という趣旨である。Dillman が次にどう変わるのかを引き続き注目し、彼のアイデアを支持する研究や批判する研究成果からも学ぶ姿勢が期待される。

4. 確率パネル

社会調査やマーケティング・リサーチの分野において確率パネルの利用が拡大しているとのことで、その構築方法と品質に関するセッションがあった。そこで示された事例をいくつか紹介する。

4-1. オランダ・LISS パネルの事例

オランダ・ティルブルフ大学にある研究機関 CentERdata の Joris Mulder は、LISS (Longitudinal Internet Studies for the Social sciences) パネルによる研究成果を発表した。オンラインパネルには 5000 世帯、8000 人 (16 歳以上) が登録している。毎月 30 分程度の分量の調査をしており、インセンティブは 1 時間あたり 15 ユーロである。

まず、パネルメンバーの募集方法の最適化実験について報告された。最適化する要素としては、①電話 (CATI) か面接 (CAPI) か、②謝礼の額、③謝礼を渡すタイミング (前渡しか、後渡しか)、④手紙とパンフレットの内容、⑤パネル参加依頼のタイミング、の 5 点が検討された。その結果、手紙の内容やパネル参加依頼のタイミング (④⑤)、パンフレットを読み上げるか見せるか (①) は回収

図表 3 Push2Web の回収率最大化戦略

郵便物を開ける動機付け	郵便物を読む動機付け	調査に参加する動機付け	オンラインに移動する動機付け	調査を完了させる動機付け
個人化	個人化	郵便物の目的についての明確性	調査画面にログインするための明確な指示	ランディングページに明確な指示と本物に見えること
郵便物のタイプ: 封書かはがきか	読みやすさ: 長さ、フォント、文法	調査依頼の明確さ	ログインするための最小限の労力	モバイルのためのデザイン: 短さ、混乱とテキストを減らす
開封前の郵便物の見た目: ログ	見た目: 重要性、プロ仕様	参加への説得力のある理由	複数のアクセス方法: 任意のインターネット対応デバイス	中断しやすい質問形式を避ける
複数の郵便物を使う ・ 郵送回数を最適化する ・ 郵送の時間間隔を最適化する ・ 多様でありながら一貫した郵送物				認知負荷を軽減する

率に影響しないことがわかった。しかし、謝礼の額②では回収率に差があり、0ユーロと10ユーロで差は15ポイントと効果に大きな違いが出たが、10ユーロと20ユーロ、20ユーロと50ユーロでは約3ポイント程度の差しかなかった。また、謝礼を渡すタイミング③については前渡しの方がよかった(Scherpenzeel & Toepoel, 2012)。

この結果を受けてパネル募集はCATIとCAPIとともに使い、督促はCAPIまたはPAPIで行うことにし、謝礼は前渡しの10ユーロとし、パネルに参加した場合には追加謝礼10ユーロを渡すことにした。募集の手順は、まず前渡し謝礼を同封した依頼状を送り、CATIまたはCAPIによるインタビューでパネル参加のための質問もする。ネット接続が既にある世帯ならば、そこで登録を承諾してもらえれば追加の謝礼を支払う、という流れだ。ネット接続がない世帯に対しては、SimPCと呼ばれる最低限の機能だけを備えたPCとADSL接続を提供してパネルに登録してもらうことで、ネット非利用世帯をパネルに取り込んでいる。こうした手順で募集を行った結果、9844世帯のうちリーチできたのは90%で、パネル用質問への回答が完了した人は75%いた。パネルへの参加を了承したのは63%だが、実際に登録した人は48%だった。

時間とともにアクティブメンバーの数が少なくなっていくので、メンバー数を維持するためにオランダ統計局の協力を得て、2009年、2011年、2013年、2017年に統計資料に基づいてパネルの追加募集(補正)を行っている。また、休眠メンバー(3カ月間回答していない人)を再びアクティブ化するため、与える謝礼の額(10ユーロ、20ユーロ、30ユーロ)や条件(前渡ししか復帰したら渡すか)を変える実験を行った。その結果、復帰したら10ユーロ与える場合が最も効果が高く、40%の人が再アクティブ化した。

4-2. アメリカ・American Trendパネルの事例

政党寄りの調査機関が多いアメリカでノンパルチザンの調査機関として信頼を集めているピューリサーチセンターから、Nick Bertoniが発表した。ピューリサーチセンターの確率パネル「American Trend Panel(ATP)」は、デュアルフレームRDD調査によってモニターを採用している。ATPでは現状、インターネットではなく郵送法で調査を行っている。インターネット非利用者層と利用者層では属性などの違いが大きいからである。

今回発表は、現行の郵送調査をWeb調査に置き換えたことによる影響を確認している。郵送法だと、どうしても調査期間が長くかかってしまう。調査票の印刷や郵送の時間、返送後のスキャンやデータ入力の手間をなくして、迅速でタイムリーな調査を実施したいという思いがある。また、資料の印刷後に内容を変更したり追加したりする柔軟性がなく、スキップロジックやランダムイズが使えないので、アンケート設計が制限されるという欠点もある。

置き換えは、パネリスト(574人)に手紙を送って呼びかける形で行われた。置き換えを承諾してもらえるならメールアドレスを提供してもらい、ネットを利用していない人に対してはタブレット端末とネット接続を提供した。タブレット端末がベンダーから出荷され、パネリストが受け取った後、フォローアップコールをして使用方法を示している。

その結果、66%のパネリストが置き換えに同意した。ところが、タブレット端末を受け取った後に拒否した人が12%も出るなど、最終的に置き換えできたのは41%にとどまった。

どのような属性の人たちが置き換えに同意したのか。年齢別にみると30-49歳は65%、50-64歳は52%の人が置き換えしたが、65歳以上では32%しか置き換えできなかった。人種別では黒人や非ヒスパニック系で置き換え率が高めだったほか、支持政党別では共和支持層よりも民主支持層で置き換え率が高めだった。また、ネット利用者層では52%が置き換えしたが、ネット非利用者層では32%にとどまった。性別や学歴、収入、有権者登録の有無、政治的イデオロギーなどでは、置き換え率に有意な差はなかった。

この置き換えによってパネル構成が大きく変化した。パネリストに占める65歳以上の比率が30%から27%に減少している。また、ネット非利用者層は6%から2%に減少し、学歴が高卒以下の層も18%から15%に減少、年収3万ドル以下の層は24%から21%に減少した。性別や人種、支持政党、有権者登録の有無、政治的イデオロギーなどでは有意な差は出なかった。

こうした置き換え後のパネル構成比を補正するため、デュアルフレームRDD調査(固定電話25%/携帯電話75%)による追加モニターの募集を行っている。ネット非利用者にはモバイルインターネット接続付きのタブレット端末を提供する。通常

の依頼では協力が得られにくい集団に対しては異なる謝礼を提供することにした(5\$/10\$/20\$)。結果、調査に回答した3905人のうち1628人(42%)がパネルに参加することを承諾した。

この追加募集後のパネルの構成比をみると、65歳以上の構成比はさらに減少してしまった。また、ネット非利用者の比率に変化はなく、低学歴層はやや増加した。年収は正しい方向に変化したものの、大きな変動ではなかった。

モード変換は大きな課題だが、時間や費用の削減など得られる効率にはその価値がある。ただ、変換を行う場合は、サンプル構成の不均衡を修正する準備が必要である。

4-3. オーストラリア・Life in Australiaの事例

オーストラリア国立大学・Social Research CentreのDarren Pennayは、オーストラリアの確率ベースオンラインパネル「Life in Australia」の募集実験について報告した。

募集調査はデュアルフレームRDD調査(携帯電話60%)で実施され、携帯電話サンプルに対してはSMSの予告メッセージを送付している。また、住所が判明した固定電話サンプルに対しては、事前予告の手紙を送付した。募集調査を完了した人に対しては10\$の謝礼を渡した。この募集調査が完了した人たちに対してパネル登録のための「インフォメーションパック」を郵送して、電話またはオンラインでパネル登録調査に回答してもらった。このパックには10\$のインセンティブを同封している。

今回は、募集調査のやり方を変えることによってパネルへの登録率に差が出るかどうかの実験結果が発表された。

条件①:「最初に尋ねる」か「最後に尋ねる」か
パネルへの参加意向を確認するタイミングを、調査の最初にするか、最後にするかで登録率に差が出るか

条件②:「直接的」か「間接的」か
募集調査の中でパネル登録の意向まで確認してしまう「直接的」な勧誘と、募集調査の中では郵送物を送付する案内までにとどめ、郵送物を見てから登録するか決めてもらう「間接的」な勧誘とで登録率に差が出るか

この2つの条件に対して2×2の4グループ(各700s)に分けて比較実験を行っている(図表4)。募集調査の完了率をみると、「最後に尋ねる」と「間接的」の場合が最も高く20.1%(RR3)だったが、そのうちモニター登録した人は66.3%(RR3で13.4%)とふるわなかった。募集調査を完了した人の中での登録率が最も高かったのは「最初に尋ねる」と「直接的」の場合で87.5%だったが、サンプル全体での登録率で見ると、「最初に尋ねる」と「間接的」の場合が14.6%(RR3)で最も高かった。

図表4 各グループの完了率と登録率

	「最初に尋ねる」 &「直接的」	「最初に尋ねる」 &「間接的」	「最後に尋ねる」 &「直接的」	「最後に尋ねる」 &「間接的」
サンプル数	701	706	705	705
募集調査完了数	80	97	79	104
パネル登録数	70	78	63	69
完了率(RR3)	15.1%	17.6%	15.4%	20.1%
完了した人の中での登録率	87.5%	80.4%	79.7%	66.3%
登録率(RR3)	13.2%	14.6%	12.3%	13.4%

そこで、本番の募集調査では、①全員に対して「最初に尋ねる」、②対象者に応じて「直接的」か「間接的」かで対応したところ、募集調査の完了率は20.2%に、全体の登録率も15.7%に改善することができた。

作成したパネルの精度を確認するため、高品質な人口統計ベンチマークと比較した。比較したのは自宅の所有権、家庭構成といった二次デモグラフィック13項目と、生活満足度や喫煙状況といった具体的な測定6項目。ベンチマークからの平均誤差はABS調査やデュアルフレームRDD調査と比べても遜色ない結果となった。

RR3

アメリカ世論調査協会(AAPOR)は回収率(Response Rate)の計算式を6種類定義しており、その第3番の定義に従って算出した回収率。

ABS

Address Based Samplingの略。郵便集配リストなどの住所情報を基に対象者(対象世帯)を抽出する手法。

4-4. 各種パネルの精度比較実験

ドイツのマンハイム大学・School of Social SciencesのAnnelies Blomらは、確率と非確率の両パネルと標本調査を比較検証した結果を発表し

た。選挙など政治的現象や社会的現象を研究するために使用する場合、確率ベースのパネル調査、非確率ベースのパネル調査および確率抽出の面接調査では、どれが最も適しているのか。この実験では、レイキングウェイトによって精度が向上するのか、時間経過による変化をとらえるのに適しているのはどれかについても調べられた。

比較に使用されたデータは以下のとおり。

〈1〉2つの確率オフラインサンプル(面接調査)

- German General Social Survey (ALLBUS)
- European Social Survey (ESS)

〈2〉2つの確率オンラインパネルサンプル

- German Internet Panel (GIP)
マンハイム大学のSFB884が開始。確率抽出による面接調査でパネルを採用し、ネット非利用者には必要な機器を提供
- German Internet Panel onliner sample (GIPon)
GIPのうちネット利用者層に限ったサブグループ

〈3〉8つの非確率オンラインパネルサンプル

- 商業用のオンラインパネル。いずれも自己登録型のもので、ネット非利用者層は含まれない。全てのサンプルは割り当て法を使用。

性別、年齢、学歴といった社会的人口構成比や政治的態度(選挙での投票先など)について、小規模国勢調査(Mikrozensus)などをベンチマークとして精度を評価した。その際、Average Absolute Relative Bias (AARB, 質問項目ごとに誤差の絶対値の平均をとったもの)を指標としている。

$$AARB = \sum_{k=1}^K \left(\frac{|y_{sk} - y_{bk}|}{y_{bk}} \right) / K$$

y_{sk} サンプルsの各カテゴリkの値

y_{bk} ベンチマーク統計量bの対応する値

K 評価されたカテゴリの総数

まず社会的人口構成比のAARBを比較すると、面接調査では平均0.093、確率オンラインサンプルは平均0.085なのに対し、非確率オンラインサンプルは平均0.140とやや大きく、確率サンプルよりも偏りが拡大していた。確率サンプルのうちAARBが最大のものに比べても、各非確率サンプル

のAARBのほうが大きかった。また、8つの非確率サンプルについて、コストと精度の間には関連性はみられなかった。

政治的態度のAARBを比較すると、確率オンラインサンプルでは平均0.518、非確率オンラインサンプルは平均0.713で、こちらも確率サンプルのほうが精度は高かった。同様に、非確率サンプルのコストと精度には関連性はなかった。面接調査については、調査時期がオンラインサンプルと異なり選挙時期により近いため比較から除外した。

レイキングウェイトによる補正については、社会的人口構成比のAARBは改善することができるものの、政治的態度のAARBはほとんど改善しなかった。非確率サンプルを用いる研究者は独自の補正方法を導出する必要がある。

確率オンラインパネルおよび非確率オンラインパネルについて、1年に3回の調査を実施し、モニターの保持率の差を比較した。確率オンラインサンプルでは2回目で89%、3回目でも87%と高い保持率だったが、非確率オンラインサンプルでは1つを除いて44%から68%と総じて低かった。非確率サンプルのコストと保持率の間に関連性はなく、最もコストが安いもので保持率が最も高くなった(3回目で83%)。

4-5. 確率パネルの論点

パネルではなく毎回フレッシュな確率サンプルで調査を行えば理想であるが、回収率の低下は続いており、コスト面からも運用が難しくなる可能性もある。そのため、欧米では有望な代替手段の一つとして確率オンラインパネルが構築されている。だが、確率オンラインパネルの構築が解決策となるかどうかは未知数である。

今回報告された確率オンラインパネルの募集実験結果をみると、パネルへの登録率は高くない。しかし、通常の高確率ベースの調査結果や人口統計などと比較すると、回答者の属性構成はそれほど偏りがなく、非確率オンラインパネルを使った調査より精度は高い。発表ではほとんど触れられなかったが、パネルの管理や維持にかかるコストと精度を勘案したコスト対効果の精査が必要である。パネルメンバーの脱落が生じるため、追加募集が必要であるし、ネット非利用者に対する端末とネット接続の提供にかかるコストも重い。調査の精度面の検証だけではなく、そうした実務面の情報開示も今後進むことを期待したい。

5. オンライン調査でのモバイルデバイスへの対応

スマートフォンなどモバイルデバイスの普及により、オンライン調査におけるモバイルデバイスへの対応が不可欠となっている。Dillman たちもこの課題が重要になることを指摘している（3-6, 3-7 節参照）が、いち早くモバイル対応を研究している調査者の間では、モバイルを付随的な課題として扱うのではなくメインとする「モバイルファースト」が叫ばれている。

5-1. モバイルファースト

モバイル対応の調査研究で最先端にいる Ray Poynter と共同で NewMR を運用するオーストラリア・クイーンズランド大学の Sue York も発表でモバイルファーストを強調する（Poynter, Williams, & York, 2014）。

York は、オンライン調査へのモバイルデバイス対応における課題として、以下の7項目を挙げた。

- 〈1〉 デバイスの画面サイズの小ささ
- 〈2〉 プラットフォーム、OS が異なる
- 〈3〉 デバイスが異なる
- 〈4〉 調査に時間がかかる
- 〈5〉 調査の途中打ち切りが多い
- 〈6〉 質問を飛ばすことが多い
- 〈7〉 モードまたはデバイスエフェクト

一方、メリットは、以下の5項目などを指摘した。

- 〈a〉 カバレッジの改善
- 〈b〉 位置情報などの付加情報が得られる
- 〈c〉 写真や動画の利用など調査の選択肢が広がる
- 〈d〉 より対象者に接触しやすくなる
- 〈e〉 買い物中や旅行中の調査が可能

オンライン調査へのモバイルデバイス対応は遅れている。Research Now 社のデータによると（図表5）、モバイルに最適化された調査はこの3年間

モバイルファースト

Google のプロダクトディレクターである Luke Wroblewski が提唱した（Wroblewski, 2011）。「Web サイトやアプリケーションは、モバイル向けに設計、構築されている必要がある」という考え方で、現在、マーケティング・リサーチ業界のモバイルアンケートデザインにおける動向となっている。

でほとんど増えていないことがわかる。

ではなぜモバイルファーストなのか。モバイルの制約を受け入れることで、質問の数や長さ、指示、異なるデザインなど、機能の選択と集中が行われる。また、カバレッジの改善やパッシブデータの取得など成長の新しい機会が提供される。こうした新しい能力を手に入れることによって、回収率の低下や離脱、回答者の負担などへの対応が可能になる可能性がある。モバイルファーストを追求することで、研究と調査デザインを再考する機会が得られる。

図表5 モバイルへの対応状況

	2014年	2015年	2016年
モバイル未対応	30%	33%	29%
モバイルも可能	27%	23%	23%
モバイルフレンドリー	30%	30%	33%
モバイル最適化	13%	15%	15%

このセッションでは、モバイルファーストを含めモバイルデバイスへの対応事例も報告された。

5-2. German Internet Panel の事例

ドイツ・マンハイム大学の Christian Bruch らは、確率オンラインパネル「German Internet Panel」において、調査画面をモバイルファーストデザインに切り替えた実験結果を発表している。具体的には、モバイル向けにサイトを設計して、タブレットや PC でアクセスした場合にはそちらに転送されるようにしている。スマートフォン利用者とその他の端末利用者との間で測定誤差を最小限に抑えるためである。

この対処をする前の調査では、スマートフォンでアクセスした場合、PC 向けの画面がスマートフォンの画面サイズに合わせて縮小される形で表示されていた。モバイルファーストデザインへの切り替えて、以下の5項目の変更が為されている。

- 〈1〉 文字が縮小されることなく大きく表示される
- 〈2〉 ボタン同士の間隔を広くとる
- 〈3〉 「続行」ボタンのサイズを大きくする
- 〈4〉 ラジオボタンの周囲のスペースを広くとる
- 〈5〉 グリッド質問や水平スケールは使用しない

こうしたモバイルファーストデザインに変更したことで、調査にスマートフォンで回答する人が増えるのかどうか、デザインが変更されたことを、対象者に通知するか否か、また通知する場合でも通知方法を変えることで、どの程度差が出るのかを比較実験して調べている。実験では、調査対象者が以下のように a, b, c の 3 群に無作為に分けられた。

- 〈a〉 デザイン変更を通知しない通常の調査依頼メールを送信する
- 〈b〉 デザイン変更をメールの本文に記載した調査依頼メールを送信する
- 〈c〉 デザイン変更をメール本文だけでなく件名にも入れた調査依頼メールを送信する

その結果、スマートフォンによる回答は a 群で 27.6% だったのに対し、b 群は 32.2%、c 群では 40.2% と変更の情報を増やすほど増加した。変更情報の影響だけではなく、新デザインがスマートフォンでの回答を容易にしたこともスマートフォン回答者の増加に影響を及ぼしたと推察される。

5-3. オランダ・労働力調査の事例

オランダ・ユトレヒト大学の Vera Toepoel, Peter Lugtig は、回答デバイスの違いによる調査脱落を低減させるための研究成果を発表した。

オランダの労働力調査では調査票が長すぎるのが問題となっている。2015 年の調査で回答デバイス別の調査脱落率をみると、デスクトップ PC では 15%、タブレット端末では 21% なのに対し、スマートフォンでは 41% に上った。

この問題を解決するため、調査票を複数のパーツに分けて調査する方法の有効性を確認している。オランダの確率ベースパネルである LISS パネルから抽出した調査対象者を、①通常の調査、②3 分割、③10 分割の 3 群に分け、調査脱落率の違いや回答データの品質などを比較した。調査テーマは政治や価値観が中心で、回答にかかる時間は 30 分程度の分量である。インターネットに接続可能な携帯電話を持つパネルメンバーは、3 群に無作為に割りあてられている。

その結果、分割数が多いほど調査開始率は高く、調査票を分割していない①群よりも 10 分割してある③群のほうが調査開始率は 10% ほど高かった。しかし、③群では途中脱落が約 10% あり、最終的

な調査完了率の差はほとんどなかった。3 群のいずれの調査完了率も 77% 前後であった。ただし、データに欠損のある票は、③群が最も少なく 11.2% で、①群が最も多かった (13.8%)。なお、携帯電話による回答の割合は、③群が①群に比べてやや多めとなっている。

データの品質についてはどうか。全ての質問項目で同じ選択肢を選ぶ「ストレートライニング」の発生率に差はほとんどなく、「DK」を選択する割合にもほぼ違いはなかった (いずれも③群が若干よかった)。分析の結果、データの品質には差はないと結論づけられている。

5-4. グリッド質問形式の比較実験

ミシガン大学の Mick P. Couper らは、グリッド質問に関する比較実験の成果を発表した。調査対象者の回答行動やデータの品質を、以下のように「デバイス」「質問形式」「スケールの方向 (選択肢配置)」「次へボタンの表示位置」の 4 条件で比較している。

- 〈1〉 デバイス…PC か、スマートフォンか
- 〈2〉 形式…グリッド形式か、項目ごとか
- 〈3〉 スケールの方向…水平方向か、垂直方向か
- 〈4〉 「次へ」ボタンの表示位置…常に画面に表示するか、ページの最後にだけ表示するか

この実験ではオプトイン・オンラインパネルが使用された。調査対象者は、上記の 4 つの条件に対して図表 6 のように 7 群にランダムに分けられている。なお、調査対象者が条件と異なるデバイスでアクセスすると、ブロックされるように設定されている。質問をグリッド形式ではなく項目ごとに区切る形式の場合は、項目ごとに「次へ」ボタンが配置されるのではなく、全項目を 1 ページに表示してスクロールさせる形のため「次へ」ボタンは最後に 1 つ配置されるだけである。なお、質問を飛ばすことは可能な設定とした。スマートフォン用の画面は、スマートフォン向けに最適化されたものを使用している。

PC とスマートフォンでの回答にかかった時間を比較すると、PC のほうが短くなった。「次へ」ボタンの表示位置については、「常に表示」の場合、回答完了時間は短くなり、より無回答項目が多くなってしまふ。グリッド質問と項目ごとの形式の比較では、スマートフォンの場合、グリッド形式

図表6 実験グループの設定

グループ名	PC-G	PC-H	PC-V	SOa-G	SOa-H	SOa-V	SOc-V
N	214	214	214	208	209	209	208
デバイス	PC			スマホ			
形式	グリッド	項目ごと		グリッド	項目ごと		
スケールの方向	水平		垂直	水平		垂直	
次へボタン	最後だけ			最後だけ			常に表示

のほうが一部で無回答項目が多かった。スケールの方向について比べると、PCでもスマートフォンでも水平方向のほうが回答完了時間は短くなった。また、スマートフォンの場合、無回答項目が水平方向のほうで多めになる質問項目が一部あった。

総合すると、スマートフォンで「次へ」ボタンを常に表示させるのは回答品質の面で問題である。しかし、そのほかの点については一貫した違いは見られず、スマートフォンの回答品質は言われているほど悪くなかった。グリッド形式の質問についても、きちんとスマートフォン用に最適化されていれば、いくつか批判されているほど悪くはない。

5-5. スライダーバーについて

オランダ・ティルブルフ大学の Angelica Maineri らは、Web 調査で使われる「スライダーバー」の設定の違いで回答行動に違いが出るのか、回答に使用するデバイスによっても違いがあるのかどうかを実験し、その成果を発表した。

比較のために使用したデータは、イタリア・トレント大学の学生を対象にした Web 調査で、2015 年に行われた「生徒の満足度調査」と 2016 年の「時間とスペースの使用調査」のものである。以下の 2 つの実験を行い、検証している。

実験1：スライダーバーの数字ラベルの有無

生徒の満足度調査で、スライダーバーのハンドルの上に数字ラベル (0 から 100) を表示する場合と表示しない場合で回答に違いが出るか実験した。ラベルがある場合は、回答はきりの良い数字 (1 桁目が 0 や 5 の数字) が増える傾向がみられた。ただし、スマートフォン回答者に

では、PC やタブレット端末での回答者に比べるとその傾向は弱かった。この違いは、画面サイズやフォントサイズが原因だと推測される。

タブレット端末での回答者はスマートフォン回答者ほどではないが、PC 回答者よりもこの傾向はやや弱めである。スマートフォン、タブレット端末と PC での入力方法の違い (スクリーンタッチか、マウスか) も影響している可能性がある。

また、極値 (左端または右端の値) を選ぶ傾向や中央値を選ぶ傾向については、質問項目によって異なっており、はっきりした傾向はあらわれていない。

実験2：スライダーバーのハンドルの初期位置

時間とスペースの使用調査では、スライダーバーのハンドルの初期位置を、中央にするか、右端にするか、左端にするかで回答に違いがあるか分析した。初期位置が左端に設定された回答者は、右端に設定された回答者よりも回答数値が低くなる (回答数値が左に寄る) 傾向がみられた。また、回答デバイス別にみると、スマートフォン回答者のほうがこの傾向がより強く出現している。

いずれの実験でもスマートフォンで回答した人のほうが、回答行動に強い影響を受けている。マルチデバイス調査においては、デバイス間の回答特性に配慮した慎重な設計が必要である。

5-6. モバイルデバイスへの対応の論点

オンライン調査において、スマートフォンなどモバイルデバイスへの対応が必須となっている状

況の中、海外でもその対応があまり進んでいないのは、率直に言って意外だった。

調査におけるモバイルデバイスへの対応において、モバイルファーストは取り入れるべき考え方であると思う一方、モバイルデバイス用に設計された調査票を、そのままPC向けに表示することが本当によいのかについては検討する必要がある。指先で操作するスマートフォンと、マウスやキーボードで操作するPCとでは、回答者が「回答しやすい」形式には違いがあって当然だからである。

モバイルデバイスと一口にいっても、さまざまな種類の端末が存在している。代表的なスマートフォンである iPhone を例に考えてみても、最新機種種の iPhoneX の画面サイズは 5.8 インチでアスペクト比は 19:9 だが、iPhone6s の画面サイズは 4.7 インチ、アスペクト比 16:9 である。同じモバイル用の調査画面を表示したとしても、端末によって一覧できる範囲が異なる。調査画面の設計においては、このような端末差への配慮も必要となる。

6. 非確率オンライン調査のデータ品質

調査対象者を母集団の中から無作為に選ぶ確率抽出ではなく、割当など確率によらずに選ぶものを「非確率」と表す。標本抽出の場合でも、調査パネルの設計の場合でも意味は同じである。このカンファレンスでは、非確率オンライン調査のデータ品質に関する発表が行われた。その一部を取り上げる。

6-1. スイスでの比較実験

スイス・FORS の Nicolas Pekari らは、非確率サンプルによる調査結果の品質について調べた結果を報告した。3 種類の非確率オンライン調査データとベンチマークとなる参照データを比較し、データの品質を検証している。検証に用いる非確率サンプルとして、単価がそれぞれ①10 ユーロ、②6 ユーロ、③3.5 ユーロの 3 種類のオプトインパネルを使用し、デモグラフィック（性別、年齢、世帯規模、言語）と政治変数（投票率、スイス国民党への投票）を比べた。

連邦統計局の人口登録データをデモグラフィックの基準値として使用し、確率サンプルデータとして、2015 年国政選挙後調査（Mixed-Mode、単価 60 ユーロ）、と 2015 年パネル調査（Web のみ、

単価 40 ユーロ）を用いている。

また、比較の指標として 2 変量分布の Mean absolute bias (MAB) (Dutwin & Buskirk, 2017) を使用している。例えば、図表 7 の場合、性別と年齢別との MAB は「差の絶対値」の部分の 4 つの数値を平均した 10%ポイントとなる。

点推定値をみると、「投票率」「スイス国民党への投票」が、3 つの非確率サンプルの間で大きく異なる値となった。ただ、最もコストの高い①のオプトインパネルの推定値が参照調査の値に最も近い。デモグラフィックについて MAB を比べると、非確率サンプルについてはいずれも 2 つの確率サンプル調査に比べて大きな値になっている。しかし、非確率サンプルの中では①のパネルが最もましな結果だった。「投票率」「スイス国民党への投票」の MAB についても同様に、確率サンプルよりも非確率サンプルのほうがかなり大きな値になっている。

結果として、非確率サンプルの中では単価 10 ユーロのものが最も品質が高く、統計モデルを作成するには十分といえる。しかし、非確率サンプルは確率サンプルに比べると大きな偏りがあり、不確実性が高い。パネル構築の透明性が低く、パネルの再現性も欠如している。

6-2. オーストラリアでの比較実験

オーストラリア国立大学・Social Research Centre の Darren Pennay らは、オンラインパネル調査の品質を「確率」と「非確率」の複数の調査を用いて比較実験した。オーストラリアでは現在、86%の世帯がインターネット接続を持っており、オンライン調査の人気の高まりが続いている。しかし、非確率オンラインパネルには長所と短所がある。長所としては、コストが安く、迅速に実施できることや、捕捉が難しい目標母集団に到達できることなどが挙げられる。短所は、カバレッジがないことや、非確率抽出であるために標準偏差や信頼区間の計算ができず、関心のある目標母集団に一般化できないことである。オンラインパネル調査

図表 7 MAB の計算例

	調査X		基準		差の絶対値	
	男性	女性	男性	女性	男性	女性
5 0 歳未満	60%	40%	50%	50%	10%	10%
5 0 歳以上	40%	60%	50%	50%	10%	10%

の品質を調べるため、3つの確率サンプル調査と5つの非確率サンプル調査の比較実験を実施した。

比較に使用した確率サンプル調査は以下の3種類である。

- 〈1〉 ABS 調査 (Mixed-mode)
- 〈2〉 デュアルフレーム RDD 調査
- 〈3〉 デュアルフレーム RDD 調査から構築したオンラインパネル調査 (Mixed-mode)

非確率サンプルは5つのオプトインパネルを使用し、それぞれのパネルから600人の回答を得て「全国の代表」になる調査を実施するよう依頼した。

調査結果にはウェイト付けを行っている。確率調査については、デザインウェイト（例えば、電話回線数や世帯内人数などによる重み付け）をかけた。また、確率、非確率ともに性別、年齢、学歴、地域などを調整するレイキングウェイトをかけた。

こうしたウェイトをかけた上で、二次的人口統計（市民権、投票登録、家計構成など）、健康特性（健康状態、喫煙状況、飲酒状況など）を比較した。比較の指標には、誤差の絶対値の平均を用いている。

その結果、二次的人口統計においては、確率、非確率とも合理的に正確な推定値を得ることができた。しかし、健康特性に関しては、確率サンプルのほうが非確率サンプルの2倍正確だった。また、健康特性についてウェイト付けが精度にどう影響しているか確認すると、確率サンプル〈1〉と〈2〉については精度が向上していたが、確率サンプル〈3〉および全ての非確率サンプルについては精度が悪化していた。

非確率サンプルについては、以下のような偏りがあることが既に知られている。(Callegaro, The current and future state of online research in Australia, SRC-ANU Workshop, July, 2016)

- 〈a〉 ネット利用率が高い
- 〈b〉 テクノロジーにより興味を持っている
- 〈c〉 アーリーアダプターに偏っている
- 〈d〉 エコに関心が高い
- 〈e〉 高学歴、高収入
- 〈f〉 プライバシーに敏感

このような偏りを持ったデータに対して、限られた変数によるウェイト付けを行うと別の変数の誤差を拡大する可能性があり、非確率サンプルを扱う場合は注意が必要である。また、継続的な追跡調査を行う場合は、パネルを変えてはいけない。ウェブベースのパネルは代表的なサンプルを提供せず、異なるパネルは異なる結果を生み出す。

近年工夫されているキャリブレーション、サンプルブレンド、およびサンプルマッチングについては、状況によっては、非確率調査における誤差を低減する有望な技術となる可能性がある。

6-3. 非確率オンライン調査のデータ品質の論点

今回のカンファレンスでは、非確率オンライン調査の精度は、確率ベースの調査データよりも悪い、という報告ばかりだった。非確率サンプルのデータを扱う際には、この点は十分に認識しておく必要があり、安易な利用は慎むべきである。しかし、非確率サンプルのコストの安さやスピードは魅力的である。非確率サンプルが持つ誤差を低減させる技術の確立が大きな課題であるが、確率ベースの調査の継続が困難となる時に備え、研究する価値は十分あるのではないかと。

7. 混合デバイス調査

7-1. モバイルウェブオンリー層

ドイツ・ダルムシュタット工科大学の Anke Metzler らは、「モバイルウェブオンリー層の状況」について発表した。2015年時点でヨーロッパ諸国の60%の人がスマートフォンを利用しているという報告があった（ピューリサーチセンター、2016年）。Web調査の回答にスマートフォンを利用する人は増えている。スマートフォン回答者のうち、インターネットへのアクセスにスマートフォンしか使わない人たちを「モバイルウェブオンリー層」と呼ぶ。モバイルウェブオンリー層とはどういった人たちで、欧州各国でどの程度いるのか、欧州委員会が実施する世論調査 Eurobarometer（欧州各国500～1500人の18歳以上を対象に実施された面接または電話調査）の2012年から2015年のデータからみても。

2015年調査で、インターネット利用者のうちモバイルウェブオンリー層は5%だった。2012年からのトレンドをみると、1%→3%→4%→5%と増加傾向が続いているが、まだ低水準だ。ネット利用頻度

についてみると、「毎日使う」という人は、スマホと他のデバイスを併用している人に比べると少ない。また、モバイルウェブオンリー層のインターネット活動状況をみると、メールの送受信やオンラインバンキング、SNS、オンラインショッピングなどの利用率が、他のウェブ利用者よりも少ない傾向がある。モバイルウェブオンリー層の属性をみると、他のウェブ利用者比べて女性や若年層が多く、低学歴、管理職層が少ない傾向がある。

Web 調査において、モバイルウェブオンリー層が今後も増加し、ほかの層と属性や特性が異なれば、各層の調査協力率の違いに応じてかなりのバイアスを引き起こす可能性がある。

7-2. スペインの確率ベースオンラインパネルでの事例

スペイン・IESA/CSIC の Sara Pasadas らは、確率ベースオンラインパネルを用いて、回答に使用されたデバイス別に回答行動の違いを研究した。

回答に使用したデバイスをみると、「常にデスクトップPCまたはラップトップPC」は73.2%を占めている。「常にタブレット端末」という人は11.4%、「常にスマートフォン」は14.0%で、「デバイスを変える」は1.4%だった。依然として回答に使用されるデバイスとして主要なのはデスクトップPCやラップトップPCであるが、およそ4人に1人はモバイル端末を使用している。

回答完了に要したセッション数は、1セッションが87.7%、2セッションが12.0%、3セッションが0.3%となっている。回答完了に複数のセッションを要した人の主に使用したデバイスをみると、「デスクトップまたはラップトップ」が73.6%、「タブレット」11.8%、「携帯電話」11.7%、「タブレット」(Phone と Tablet を合わせた造語で画面サイズが大きめのスマートフォン) 2.9%で、モバイルデバイスは26.4%を占めている。

モバイルデバイスを使った回答者の属性をみると、女性や若年層、低学歴層が多めの傾向がある。回答行動をみると、モバイル回答者は途中脱落が多めであり、回答に要するセッション数も多めで所要時間も長い。また、モバイル回答者のうち、小型デバイス(6インチ以下)を使う人が多くなっており、その層では途中脱落がより多くなる傾向がみられる。

7-3. イギリスの事例

イギリス・サウサンプトン大学の Olga Maslovskaya らによる発表。英国では、インターネットを毎日使う成人は82%で、スマートフォン所有率は71%だ。そのため、様々なデバイスに対応したオンライン設計がなされ、調査への回答方法を選択できる形に移行している。例えば2021年の英国国勢調査ではオンラインで75%の世帯から回答を回収する計画だ。もはや、すべての回答者がオンライン調査においてデスクトップPCやラップトップPCを使用することは期待できない。

しかし、英国における混合デバイスオンライン調査での回答者属性に関する研究は見当たらない。そこで、英国でのオンライン調査に異なるデバイスを使用して回答している人たちの属性を調べた。使用したデータは以下の5つ。

- <1> Innovation Panel of Understanding Society Waves 7-8(IP7 and IP8)
- <2> Community Life Survey (CLS) 2014-2015
- <3> European Social Survey (ESSMM)
- <4> 1958 National Child Development Survey (NCDS) (注：参加者は全員55歳)
- <5> Second Longitudinal Survey of Young People in England (LSYPE2) Wave4 (注：参加者は16-17歳)

モバイルデバイスでの回答が有意に多いのは、「若者や女性」「有職者」「高所得者」「世帯人数が多い」「子どもがいる」といった層である。「学歴」については有意差がみられなかった。こうした結果は他国の先行研究の結果にほぼ一致している。ただ、モバイル回答者のサンプル数が、いくつかの調査については少なかった(モバイル端末の使用を推奨されていなかったかブロックされていたため)ことを留意しなければならない。

7-4. イギリスのパネル調査における事例

Kantar Public の Emily Bell らは、イギリスのパネル調査 Second Longitudinal Survey of Young People in England (LSYPE2) のデータを使い、回答に使用するデバイスによって回答傾向が異なるかどうかを調べた。このパネル調査は当初面接調査で行われていたが、第4波(2016年)から Mixed-Mode に切り替えられた。分析は、この第4波のデータを使用している(回答者は16-17歳)。

回答に使用したデバイスを見ると、4人に1人(28%)がスマートフォンである。スマートフォン使用率は女性やパートタイム職の人で高めである。逆に使用率が低いのは、高学歴層や世帯収入が平均より高い層、両親と同居している層である。

データの品質についてみると、途中脱落はスマートフォンやタブレット端末使用者で多くなっている。ストレートライニングについては、スマートフォン利用者で高めになる質問項目もあれば、そうならない質問項目もあり、はっきりとした傾向はみられなかった。回答にかかった時間は、単純にみるとスマートフォン利用者で長くなったが、回答時間は学歴と相関があり、それを調整すると回答デバイスによる違いはなくなった。

回答内容を見ると、スマートフォン回答者はリスク行動(飲酒、喫煙、薬物)を認める比率が高くなっている。しかし、第4波でのスマートフォン回答者は、1年前に実施された第3波(面接調査)の回答でも、リスク行動を認める人が多い傾向があるため、この違いは回答デバイスによる差ではないといえる。データの正確性については、調査の中で回答した「達成したGCSE(一般中等教育修了試験)の数」を、別の管理データとマッチングして検証した。その結果、データの正確性と回答デバイスとの間に関連性はみられなかった。

今回の実験結果では、回答にモバイルデバイスを使用することが、データの品質に影響を与えるという証拠はほとんどなかった。しかし、調査がデバイスに適合していないと、回収率や代表性に影響を及ぼす可能性がある。長くて複雑な継続調査でも、今後、回答にスマートフォンを使用することを心配する必要はないし、スマートフォンを使用しないように回答者にアドバイスする必要はない。ただ、この結果がすべての調査に当てはまるとは限らないことを付け加えておきたい。

7-5. 混合デバイス調査の論点

いまオンライン調査を行えば、何らかの制限を加えない限り、自ずから混合デバイス調査となる。調査画面をモバイルデバイスに最適化する必要性については5章で紹介したが、それとは別に回答に使用するデバイス自体が異なることによる回答行動の違いも考慮する必要がある。例えば、スマートフォンであれば、調査の回答中に友人からメッセージが届くかもしれないし、上司から突然電話がかかってくるかもしれない。PCで調査に回答

する層よりも途中中断や脱落が多いのには、そうしたデバイス自体の特性の違いが要因になっている可能性がある。

今回のカンファレンスでは、「スマートフォンなどモバイルデバイスによる回答が品質の低いものだと一概には言えない」という報告が目立った。恐らくデバイスの違いよりも、調査の質問文、選択肢の内容や見せ方(画面表示)などの影響のほうが大きいのではないかと、引き続きバイアスの要因研究が必要である。

8. モバイルデバイスでの新しい測定機能

8-1. グリッド質問をモバイルフレンドリーに

GfK Custom ResearchのFrances Barlasらは、グリッド形式をモバイルにうまく適用する方法を提唱している。グリッド形式(マトリクス形式とも言われる)の質問はオンライン調査でよく使われる形式の一つである。ブランドや行動のような、複数の要素を測定するのに使われる。しかし、PC用のグリッド形式質問をスマートフォンでそのまま表示すると、1画面には表示しきれない。平均して2.5個の選択肢しか表示されず、垂直方向と水平方向のスクロールが必要になる。他の選択肢を表示するため右にスクロールすると、左にある項目が見えなくなり、どの項目に回答しているかわからなくなる。さらに下にスクロールすると状況はより悪化し、ラジオボタンの海で迷子になる。

こうした状況を回避するため、回答者に画面を横向きにするよう依頼することを試してみた。質問文の最後に、画面を回転すると回答が容易なることを知らせる以下のメモを追加した。

「スマートフォンやタブレットでこのアンケートを受けていますか?画面を水平にすると読みやすくなります。」

確率パネル「KnowledgePanel」を使用して、この横向き要請のメモの有無を無作為に割り当てて実験した。図表8は、画面を横向きにする回答者の比率をメモの有無とグリッド形式の回答経験(回答済数)で比較したものである。

メモを表示することは一定の効果を上げたものの、回答者の6割が画面の向きを変えていない。横向き要請のメモを追加する方式に代わる画面デザインの工夫が必要なことが判明した。

代替デザインとして「アコーディオングリッド」が考案された(図表9)。KnowledgePanelを使用して、従来型のグリッド形式とアコーディオングリッド形式を比較するテスト調査を行った。

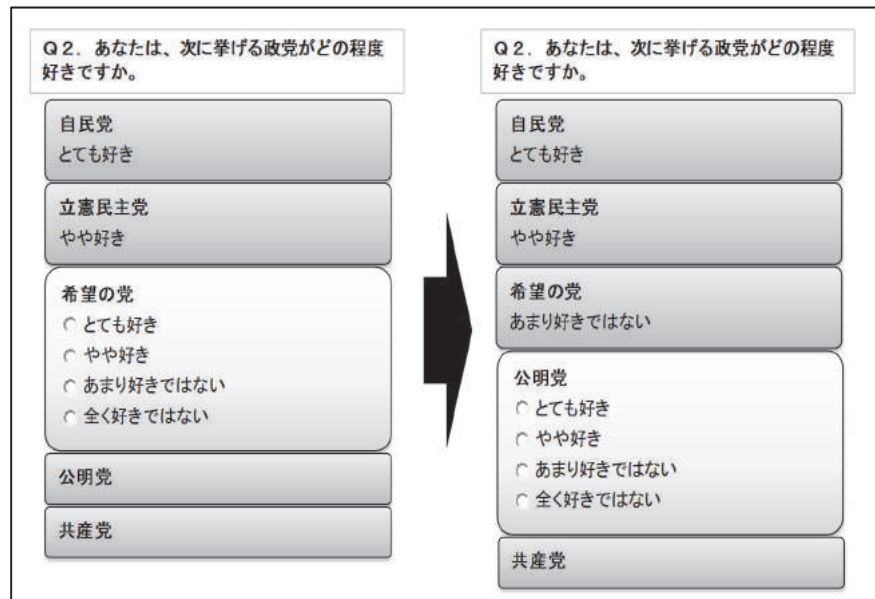
両者で質問の回答比率が異なるか確認すると、質問項目によって有意差のあるものがわずかしかなく、この差も体系的なものではなく概ね従来方式と新方式の結果は近かった。各項目の相対的な順位については、両者で近い結果にはなったものの一致はしなかった。

対象者の調査に対する満足度はアコーディオングリッドを使った回答群のほうが高くなった。また、アコーディオングリッドはスマートフォン利用者から使いやすさについて高い評価を受けた。全体としてアコーディオングリッド形式は、従来のグリッド形式よりも、スマートフォンに優しい実用的な代替品といえる。

図表8 実験調査の結果

	横向き率	
	メモあり	メモなし
調査開始時	11.9%	10.7%
1グリッド後	33.4%	20.4%
2グリッド後	38.3%	21.0%
3グリッド後	43.4%	23.6%

図表9 アコーディオングリッドのイメージ



8-2. よりよい測定

GfK Custom Research のRandall K. Thomasらが発表した。Thomasは、インターネット調査により2000年アメリカ大統領選挙の結果を的確に予測したハリスインタラクティブで、当時からメジャーメントエラーを改善する研究をしている。ミシガン大学のCouper(5-4参照)と並びThomasのメジャーメント関連の研究成果は、多くの関係者が注目している。今回は、グリッド質問で提示する選択肢の形式を変える比較実験を行っている。グリッド形式で提示する選択肢の形式によって、回答結果が異なる可能性がある。一つは、選択肢全てにラベルを付ける形か、両端にのみラベルを付ける形か、という違いだ。また、「バイポーラスケール」(例:好き-嫌い)か、「ユニポーラスケール」(例:好き-好きではない)かによっても違うかもしれない。さらには、提示されるカテゴリの数によっても違うだろう。過去のいくつかの実験で、全ての選択肢にラベルを付けた5カテゴリのユニポーラスケールが最適な妥当性

をもたらすことを確認している。

モバイル機器の登場は、Web調査に大きな影響を与えている。2010年には、米国のGfKのオンライン調査にスマートフォンで回答する人は3%以下で、ほとんどはデスクトップPCやノートPCで回答していた。しかし2年前には、スマートフォンでオンライン調査に回答した人は約18%に増えた。今日では、米国のオンライン調査の30%から40%がスマートフォンで回答されており、一部の集団ではより高い割合となっている(例えば29歳以下など)。スマートフォンで回答する対象者に対して、全てラベルを付けた5カテゴリの回答選択肢を提示すると、左右にスクロールしなければ表示できない場合がある。今回の実験調査の目的は、スマートフォンに対応するため選択肢のカテゴリ数を少なくすることが、調査の推定値や妥当性にどう影響するかを調べることである。

実験には、確率パネル「KnowledgePanel」を使用し、以下の2条件について回答者をランダムに割り当てた。

条件1: 選択肢のカテゴリ数が3つ, 4つ, 5つ
の3種類

条件2: ユニポーラスケールか, バイポーラ
ースケールの2種類 (いずれも全ての
選択肢にラベルを付ける)

実験で使用した質問項目は, 以下のとおり.

Q. あなたは以下のことが, どの程度好きで
すか?

- A. コーヒーを飲む
- B. 自分の服を買う
- C. 科学記事や本を読む
- D. ゲームをする
- E. お茶を飲む
- F. 牛肉や牛肉をつかった製品を食べる
- G. 宗教的なサービスに参加する
- H. チョコを食べる
- I. たばこを吸う
- J. 少なくとも15分間激しく運動する

実験条件1 (選択肢のカテゴリ数) と2 (ユニ
またはバイポーラ) に対応して使用した選択肢
は以下のとおり.

<ユニポーラスケール>

3段階: 好きではない, 好き, 非常に好き

4段階: 好きではない, 好き, 強く好き,
非常に好き

5段階: 好きではない, どちらかといえば好き,
好き, 強く好き, 非常に好き

<バイポーラスケール>

3段階: 嫌い, 中間, 好き

4段階: 強く嫌い, どちらかといえば嫌い,
どちらかといえば好き, 強く好き

5段階: 強く嫌い, どちらかといえば嫌い, 中間,
どちらかといえば好き, 強く好き

回答にかかった時間を比較すると, ユニポーラ
ーかバイポーラによる違いはなかった. カテゴリ
数が多くなるほど回答時間が増える傾向があ
り, スマートフォン使用者のほうが回答時間は長
めになる傾向がみられた. 回答比率の平均値をみ
ると, どのカテゴリ数でもバイポーラスケール
のほうがユニポーラスケールよりも大きくなっ
た (DK が少ないことを意味する). 妥当性につい
ては, 基準値との間の決定係数の平均値をみるこ

とで確認したところ, ユニポーラの4カテゴリ
または5カテゴリでやや高めとなった.

以下のような政治的な質問項目でも同様に違い
を確認した.

Q. 以下に挙げる問題について, あなたはど
の程度賛成しますか.

- A. 政府は, 最も貧しい市民を世話するために,
より多くの費用をかけなければならぬ
- B. この国では, 貧富の格差を縮小すること
が重要だ
- C. 政府は, 必要なだけ多くの資金を借りる
ことが許されるべきである
- D. 政府はその支出に厳格な制限を設けるべきだ
- E. 私たちの指導者は, 米国の力を拡大すべきだ
- F. ほとんどの国税は大幅に削減されるべきだ
- G. 国家間の自由貿易は, この国にとって良
いことだ
- H. 殺人を計画し, それを実行する人は死に
至らなければならない
- I. 違法移民は, この国にとって大きな脅威だ
- J. 宗教的信仰は, 私たちの国が直面するで
あろう問題に対して, 有用な解決策を提
供することができる
- K. 人々がもっと宗教的になれば, この国は
改善するだろう
- L. 地球温暖化は人間の活動のせいでは起こっ
ている

実験条件1 (選択肢のカテゴリ数) と2 (ユニ
またはバイポーラ) に対応して使用した選択肢
は以下のとおり.

<ユニポーラスケール>

3段階: 賛成しない, 賛成, とても強く賛成

4段階: 賛成しない, 賛成, 強く賛成,
とても強く賛成

5段階: 賛成しない, どちらかといえば賛成,
賛成, 強く賛成, とても強く賛成

<バイポーラスケール>

3段階: 反対, 中間, 賛成

4段階: 強く反対, どちらかといえば反対,
どちらかといえば賛成, 強く賛成

5段階: 強く反対, どちらかといえば反対,
中間, どちらかといえば賛成,
強く賛成

回答にかかった時間については、ユニポーラーとバイポーラーで違いはなく、カテゴリ数でも違いはなかった。しかし、回答デバイスでは差があり、スマートフォン利用者は回答時間が長い傾向があった。回答比率の平均値については、バイポーラースケールのほうが大きくなった。妥当性については支持政党と各項目への賛成度合いとの関連性から判定したが、違いはみられなかった。

選択肢カテゴリ数が少ないほうが、回答を完了するまでに要する時間は短い。また、ユニポーラースケールを使用する場合は、選択肢カテゴリ数が多いフォーマットで妥当性が高いことがわかった。モバイルデバイスにおいては、選択肢カテゴリ数が少ないスケールでは、やや回答時間が短くなり、回答完了が簡単になる。ユニポーラーの3つまたは4つのカテゴリ数は、最適なレベルの妥当性をもたらす。スマートフォンの「親指」環境（親指で画面をタップすることを指す）および画面の制約条件ではうまく機能するかもしれない。

8-3. モバイルデバイスでの新しい測定機能の論点

Barlas はオンライン調査におけるグリッド質問の代替形式としてアコーディオングリッドを提案した。得られる回答データも従来形式と大きな違いはなく、モバイル端末との親和性が高い形式ではないかと感じた。スマートフォンやタブレットの画面は縦長の画面サイズのため、縦方向に配列された質問領域を同じく縦方向に伸縮させることに違和感はない（図表9参照）。ただ、PC画面は横長であるため、質問する項目数や選択肢数が多い質問の場合には質問領域の伸縮でうまく画面に収まり切らず、画面スクロールや回答負担が大きくなる可能性がある。

Thomas が示したスマートフォンに最適な選択肢のスケール形式やその数についても重要な示唆を与えてくれる。質問を考える際にはどうしても調査者の意図と目的に沿った形式にこだわりがちであるが、スマートフォンなど画面の小さなデバイスへの親和性も考えた設計が必要となることを留意すべきである。

9. パッシブモバイルデータの収集

9-1. IPアドレスや地理（位置）情報の収集

ドイツ・マンハイム大学の Barbara Felderer らは、質問で聴取した地理情報を回答デバイスの

IP アドレスとリンクして分析することの同意を得ることが回収率にどのような影響を及ぼすかを調べた。近年、調査データを補強する補助データに関心が集まっている。例えば、パラデータや地域情報を含むビッグデータなどである。しかし、補助データの収集やリンクすることに同意を求めることはほとんどない（例えば、Web 調査における IP 情報の自動収集など）。文献(Couper & Singer, 2013)によると、パラデータの収集についての同意は難しい。パラデータ収集の意味を回答者に理解しやすく説明できないため、結果的に「知らせる」ことを断念しなければならないからである。

我々は郵便番号を含む回答者の現在地情報を質問で聴取する形で提供してもらうのにあわせ、IP アドレスの記録と地理情報を IP アドレスにリンクする Javascript プログラムの実行に、どの程度の人が同意するのか調べた。この調査には確率ベースのオンラインパネル「German Internet Panel (GIP)」を使用した。

現在地を確認する質問文は以下のとおり。

Q. 私たちは、今日のアンケートにあなたが回答している場所についてもっと知りたいと思っています。あなたはいま、ドイツのどの州、都市にいますか？

- ・町または地域名：自由回答
- ・郵便番号（もしわかれば）：自由回答 +DK（チェックボックス）
- ・ドイツの州名：選択肢から1つ選ぶ（プルダウン形式）+ドイツにいない（チェックボックス）

自動収集の同意を求める文章は以下のとおり。

■私たちは、今日のアンケートを回答している場所を自動的に追跡したいと考えています。この情報は、方法論的な問題にとって非常に価値があり、後続の調査を改善するのに役立ちます。この目的のために、私たちはインターネットプログラミングで非常によく使われる Javascript を使用したいと考えています。もちろん、情報は科学的目的のために収集され、匿名で扱います。アンケートにご同意いただければ幸いです。連邦および州のデータ保護法が遵守されます。

Q. あなたの現在位置を追跡することを許可してくれますか？

- ・はい
- ・いいえ

この調査の結果、95.9%の人が、意味のある町の名前を回答し、90.4%の人が、郵便番号を回答した。96.9%が少なくとも一つの位置情報を回答している。一方、Javascriptプログラムの実行に同意した人は62.1%だった。質問に対して位置情報を回答した人のうち、3分の2の人がプログラムによる自動収集にも同意した。逆に、質問に回答しなかった人で自動収集に同意した人は5人しかいなかった。

位置情報の自動収集に同意するよりも、はるかに多くの人が質問に対して位置情報を提供した。自動収集に同意した人は、「高齢者層」「低学歴層」「より好意的な人」「よりオープンな人」が多い。

質問に対して位置情報を提供する層と、自動収集に同意する層との間には高い相関がある。しかし、位置情報を手動で提供する人の特性はやや異なり、「より外向的な人」や「アンケート中に家にいる可能性が高い人」が多くなっている。

9-2. パッシブモバイルデータ収集の論点

モバイルデバイスが備えているGPSやカメラ、センサーを使用することで、調査者はより豊かな情報を取得できる可能性がある。しかし、情報を「自動的に取得」されることへの抵抗感は強い。Feldererらの実験調査で、位置情報を自らの手（自発的）であれば97%が回答するのに対し、自動収集には6割程度の人しか同意しなかったことからわかる。この実験結果は、①同意無しでの自動収集は調査主体の信用問題となる可能性をはらむ、②同意を求めれば回収率が低下して調査精度が落ちる、という調査運用と精度に絡む難しい課題を突き付けている。

センサーなどから「豊かな情報」を得られるようになるということは、プライバシー情報など、より取り扱いに注意が必要となる情報に触れる機会も高まるということである。法的な問題への対処と、データを蓄積するシステムのセキュリティや匿名化処理など技術的な問題への対処が、実務的および学術的な課題として注目を集めている。

10. ビッグデータと調査

ビッグデータと調査を組み合わせることで生まれる様々な可能性についても発表された。

10-1. ビッグデータで測定誤差を低減

ニューヨーク大学/ミシガン大学の Daniela Hochfellner らは、調査データとビッグデータを組み合わせて調査誤差などを低減する試みについて発表した。

オンライン調査には、応答負担やアイテムノンレスポンス、ブレイクオフなどの問題点がある。ビッグデータを利用（Google Trends など、あらかじめパッケージ化されたものや Facebook、Twitter などの API を利用）することで、こうした問題を解決できる可能性がある。

「The Human Project」は、ニューヨーク大学による調査とビッグデータを組み合わせた研究事例である。ニューヨーク市民約4000世帯、10000人を対象に追跡調査を実施している。できるだけ多くのデータを収集するため、Webアプリなどを活用してソーシャルメディアデータをリンクさせている。ソーシャルメディアなどのAPIへのリクエストを行うことで、参加者の同意があれば追加データを収集することが可能になる。例えば、Facebook API は Facebook のアプリ開発者であれば利用できるものである。回答者が Facebook アカウントを通じてアプリにログインするか、Facebook アカウントをアプリに接続する必要があるが、こうすることで回答者の Facebook プロフィールへのアクセスが可能になり、登録されている学歴や職歴、人脈、好きなもの、政治的スタンスといった情報を取得することができる。これらの情報と調査の回答内容を比較することで、測定誤差の検討が可能になる。

API からデータを取得するため、調査で聴取する必要がなくなり、回答者の負担を軽減することもできる。また、非回答や調査中断の埋め合わせや、調査での回答内容の正確性を確認することで精度を向上させることが可能となる。

10-2. Google Trends データを使った小地域推定

Google の Jeffrey Oldham らは、Google Trends データ、いわゆる検索データを使用し、小地域推定を行う試みについて報告した。

検索データと Cookie 情報に含まれる地域情報を組み合わせ、スパイクとスラブの帰帰

(Spike-and-slab regression) を使用して予測モデルを作成している。調査データにその予測モデルを適用して、小地域内の調査結果を予測する。これを「survey amplification」と呼んでいる。

2012年に各州でどれくらいの人々がオバマを支持しているのか。1000人を対象にした全国調査では、意味があるのは「全国」だけである。調査結果を地域別にみると非常に粗くなるという問題がある。1000人を州別に単純平均すれば1州あたり20人。DMA別(Designated Market Area - ニールセン・メディア・リサーチ社が視聴率などマーケット調査に使用するエリア区分。全米を210のエリアに分類している)、都市別(～10000都市)となるとさらに悪化する。

都市別の検索データを使って、「あなたは、きたる選挙でオバマ氏を支持しますか?」という質問の回答比率を予測するモデルを構築し、この問題を解決することを試みた。調査データに作成したモデルを適用した予測結果と、2012年大統領選の各州のオバマの得票率を比較すると、DC、ユタ、ハワイの3州については外れ値となったが、大部分の州の得票率は、予測値と概ね一致した。

調査とGoogle Trendsデータを組み合わせることで、プライバシーを保護しながら、サンプル外の予測を生成し、地域ターゲティングをサポートすることが可能となる。

10-3. ユーザーが作成したオンライン調査データの活用

Survey MonkeyのSarah Choは、自社が提供するオンライン上の調査機能を使ってなされた全世界の調査のパラデータをを使って、よりよい調査へ改善する工夫を発表した。

Survey Monkeyは世界最大のオンライン調査プラットフォームで、毎月5000万件以上の調査が世界で実施されている。ユーザーが作成した調査の傾向を分析することで、オンライン調査の改善方法についての洞察を得ることができる。

2010年1月以降のデータをみると、調査の質問数は漸減傾向がみられ、今日では平均6問程度となっている。回答者は1日の中でいつ調査に回答しているのか。アメリカのみで2016年9月から11月のデータを対象に時間別の回答率をみたところ、平日(月曜日～金曜日)はどの曜日と同じ

傾向で、午前10時と午後2時あたりにピークがあり、正午に少しへこみがあるM字型の分布を描いた。土曜日は午前10時あたりがピークとなり、その後は夜にかけてなだらかに減少していく分布に、日曜日については、午前中から夜にかけて増えていき午後9時がピークとなる分布となった。回答デバイス別にみると、平日の分布がモバイル端末と非モバイル端末で異なっており、非モバイル端末については先に示した分布となったが、モバイル端末については、午前中から増えていき午後8時にかけてピークとなる分布で日曜日に近い軌跡を描いている。

回答数が多い上位20カ国の2013年9月から2017年4月のデータについて、モバイル端末を使用しての回答がどの程度を占めているかみると、2013年9月では平均13%だったが増加傾向が続いており、2017年4月時点では平均44%と3倍以上になっている。世界を地域別(アジア、ヨーロッパ、北米、ラテンアメリカの4区分)にみても、いずれの地域においてもモバイル回答は増え続けている。

こうしたユーザーが作成したアンケートに関するデータ(ビッグデータ)を蓄積し、調査に関するノウハウを組み合わせることで、より良い調査の提案を可能にした。「Survey Monkey Genius」という機能で、ユーザーが作成したアンケートを採点し、より回収率を高めるために改善を図るものだ。

例えば、質問1を自由回答形式の質問に設定していた場合、この機能を使って採点すると、「あなたのいちばん最初の質問は自由回答形式になっています。この質問形式は、時には抵抗感を与えることがあります。複数の選択肢やドロップダウン形式の質問を使用することで、アンケートに回答者を簡単に誘導できます。」との指摘を与えてくれる(図表10)。

このように蓄積したデータを活用するためには、追跡が容易な形式にドキュメントを変更する必要がある。世界にはさらに多くのデータがあるが、それが何を表しているかを知っている場合にだけ、活用することができる。まずはよいデータプラクティスから始めよう。

10-4. ビッグデータと調査の論点

単体モードの調査で回収率の下落が続いているのをカバーすることが Mixed-Mode 調査のねらいの一つだが、それをさらに一段階先に進めたのが、調査とビッグデータとの組み合わせといえる。

ノンレスポンスやアイテムノンレスポンスによる誤差低減や回答データの正確性のチェックに利用して精度を向上させることができる可能性には大いに期待できる。また、ビッグデータを組み合わせた

小地域推定についても、増え続ける調査コストの負担の軽減を考えると、有用な代替手段となる可能性がある。

2018年10月25から27日の日程で、スペイン・バルセロナにおいて、ESRA が主催する「Bigsurv18- BIG DATA MEETS SURVEY SCIENCE-」(<https://www.bigsurv18.org/>) というカンファレンスが予定されている。調査データとビッグデータの組み合わせという新しい可能性について、多くの新しい知見が示されることを期待したい。

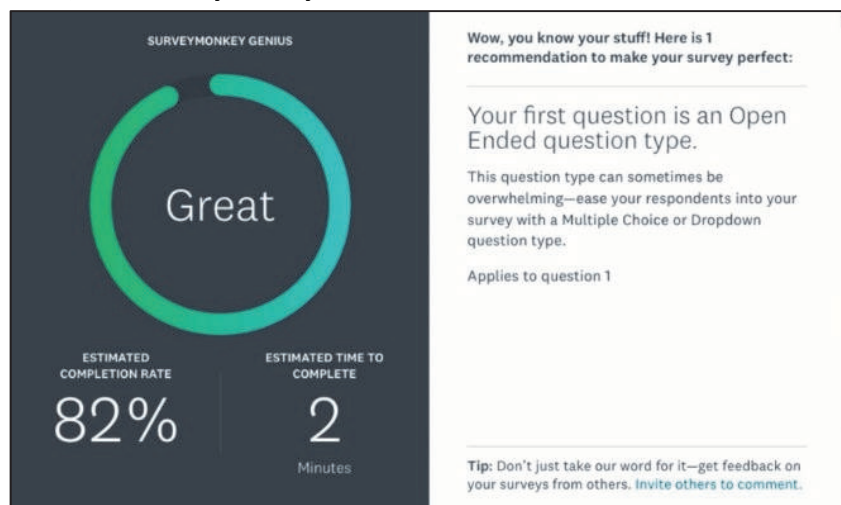
11. おわりに

今回のカンファレンスに参加して驚いたのは、オンライン調査に関係するセッションの数の豊富さだった。日本では類を見ない規模で、様々な報告が行われた。本稿では多くの報告の中から有用と思われるものを選んで紹介した。モバイルマーケティング調査の観点からは ESRA カンファレンス直後に書かれた Ray Poynter のブログ記事 (Poynter, 2017) も参考になるはずである。

面接調査や RDD 方式による電話調査など伝統的な調査手法がどんどん劣化を続ける中、日本でも Mixed-Mode 調査を含めてインターネットを調査に活用する研究や非確率サンプル調査に関する研究がもっと活発に行われることを願ってやまない。いずれにしても課題は多く困難な時代ではある。de Leuwue が言うように「退屈な時間」はとても訪れそうにない。

(朝日新聞社世論調査部・専門社会調査士)

図表 10 Survey Monkey Genius による評価結果の例



参考文献

- Couper, Mick P., Eleanor Singer. (2013). Informed Consent for Web Paradata Use. *Survey Research Methods*, Vol.7, No.1, pp.57-67.
- Bianchi, Annamaria, Silvia Biffignandi, Peter Lynn. (2017). Web-Face-to-Face Mixed-Mode Design in a Longitudinal Survey: Effects on Participation Rates, Sample Composition, and Costs. *Journal of Official Statistics*, Vol. 33, No. 2, pp.385-408.
- Dillman, Don A. (2017a). The promise and challenge of pushing respondents to the Web in mixed-mode surveys. *Survey Methodology*, Vol.43, No.1, pp.3-30. (<http://www.statcan.gc.ca/pub/12-001-x/2017001/article/14836-eng.pdf>)
- Dillman, Don A. (2017b). Why Do a Mixed-Mode Survey?. *In AAPOR Webinar*.
- Dutwin, David, Trent D. Buskirk. (2017). Apples to Oranges or Gala versus Golden Delicious?: Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples. *Public Opinion Quarterly*, 81(Special Issue), pp.213-239.
- 松田映二 (2009). アメリカの最新調査動向 グッバイ RDD, ウェルカム ABS. *Journalism*, 2009年8月号 (No.231), pp.44-54.
- 大隈慎吾, 原田和行 (2017). 郵送とインターネットの複合調査—毎日新聞社と埼玉大学の試み—. *政策と調査*, 第13号, pp.5-14.

Poynter, Ray (2017). Major update on mobile market research.

(<https://www.linkedin.com/pulse/major-update-mobile-market-research-ray-poynter>)

Poynter, Ray, Navin Williams, Sue York. (2014). *The Handbook of Mobile Market Research: Tools and Techniques for Market Researchers*, Wiley.

Scherpenzeel, Annette, Vera Toepoel. (2012). Recruiting a Probability Sample for an Online Panel: Effects of Contact Mode, Incentives, and Information. *Public Opinion Quarterly*, 76(3), pp.470-490.

Wroblewski, Luke. (2011). *Mobile First*. A BOOK APART NO.6.

文中で引用はしていないが、関連する文献・資料を以下に示した。

■Push2Webに関連する情報

Dillman, Don A., Jolene D. Smyth, Leah Melani Christian. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, 4th Edition*, Wiley.

Dillman, Don A., Feng Hao, Morgan M. Millar. (2016). Chapter13. Improving the Effectiveness of Online Data Collection by Mixing Survey Modes. In Fielding, Nugel, Raymond M. Lee and Grant Blank. *The Sage handbook of Online Research Methods, 2nd edition*. Sage Publications, London.

■HSM形式に関連する情報

Klausch, Thomas, Edith de Leeuw, Joop Hox, Anouk Roberts, and Anneke de Jongh. (2012). Matrix vs. Single Question Formats in Web Surveys: Results from a large scale. *General Online Research Conference*, March 5-7 2012, Mannheim, Germany.
(http://conftool.gor.de/conftool12/index.php?page=downloadPaper&filename=Klausch-Matrix_vs_Single_Question_Formats_in_Web_Surveys-37.pdf&form_id=37&form_version=final)

■確率パネルに関連する情報

Leenheer, Jorna, Annette C. Scherpenzeel. (2013). Does It Pay Off to Include Non-Internet Households in an Internet Panel?. *International Journal of Internet Science*, 8(1), pp.17-29.

■モバイルデバイス対応に関連する情報

Poynter, R. (2015). The Utilization of Mobile Technology and Approaches in Commercial Market Research. In: Toninelli, D, Pinter, R & de Pedraza, P (eds.) *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, pp.11-20, London: Ubiquity Press.

■非確率サンプルの品質に関連する情報

Baker, Reg, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, Roger Tourangeau.(2011). *Report of the AAPOR task force on nonprobability sampling*. American Association for Public Opinion Research.

Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly*, 75(4), pp.709-747.