

総調査誤差から「調査の質」を考える

大隅 昇

はじめに

我々は日常的に様々な調査報告を目にする機会が多い。身近には、マスメディアが行う世論調査、諸官公庁・自治体が行う統計調査、調査企業の行う消費者調査や市場動向調査、企業の行う顧客満足度調査、研究者が行う社会調査（意識調査、態度調査など）と色々な調査情報がある。

最近、情報通信技術の進歩で新たな調査用具（PC、スマートフォン、タブレット端末など）やコンピュータ支援型の調査方式（モード）、たとえばCATI、CAPI、IVR、ウェブ調査、モバイル調査などの登場で調査内容が多様化し、調査の信頼性や質の評価が複雑かつ難しくなっている。しかも調査実施環境は次第に悪化する傾向にあり、調査の質の維持が難しい状況にもある。

たとえば、調査の質を測る指標の1つである回答率は、ほとんどの調査で低下現象にある。マスメディアの行うRDD方式電話調査による世論調査の回答率が50%を下回ることもある。ウェブ調査では正確な回答率の算出が難しいことから参加率の利用が勧められているが、この参加率が開示される例は皆無に近く、筆者の知る範囲ではこれが一桁台のことも珍しくはない。

こうした事象はいずれも、標本調査で得られた推定値（例：内閣支持率、顧客満足度の割合、属性の割合、世帯収入など）と、標本を選んだ母集団特性値（真値）との間にずれ（違い）を生じる。このずれが誤差であり違いを生む要因

が誤差発生源である。調査を繰り返し行くと、推定値は変動を伴って変化する。つまり調査で得た回答に基づく推定値には、偏りに起因する誤差（系統誤差）とランダムに起こる不可避の変動誤差（偶然誤差）がある。これがもっとも簡単な誤差モデルの例であり誤差を最小限におさえることが調査の質の改善となると考える。しかし実際の調査では、こうした簡単なモデルでは説明できない。実際の調査実施過程で生じるあらゆる誤差成分を総調査誤差（TSE：total survey error）として詳しく調べ総合的に誤差特性を考察し調査の質の向上を図る枠組みが必要でありこれを総調査誤差パラダイムという。

総調査誤差（以下、TSEという）とは、調査測定が調査者の意図したように機能せずに生じた、あらゆるずれである誤差の全体をいう。つまり、調査実施過程のさまざまな場面で生じる可能性のあるすべての誤差の累積を意味する。狭義には、前述のように、真値と調査から得た推定値とのずれの全体をいう。

総調査誤差パラダイム（以下、TSEパラダイムという）または総調査誤差アプローチとは、主要な推定値のTSEを可能な限り低減するために、利用可能な調査資源や経費を最適配分するという概念である。理想的には、あらゆる誤差の原因を特定し、その誤差がなくなるように調査資源を配分し、所与の経費内でしかも決められた期限内に調査目的を達成することである。そしてこのことを通じて調査の質の改善を総合的

に図ることが目標である [Biemer & Lyberg (2003), Smith (2011), Groves 他 (2009)]。

一般に、調査誤差が小さいほど (大きいほど) 調査の質が高い (低い) と考えている。そして調査の質と投入可能な調査資源や経費は表裏の関係にある。適切な経費で優れた調査の質を確保するための要件の検討、経費に見合った質が確保できたかの客観的な評価が重要なことは自明である。

しかし現実には、投入可能な調査経費や調査資源には制約がある。制約条件の下で、質の改善を最優先とするか、質をある水準に維持しながら費用低減を図ることなのか、あるいは質の点で犠牲をはらっても情報取得を急ぐか、調査者は常にこれらを意識して調査設計に対応せねばならない。

つまり TSE パラダイムの目標は、当該調査の目的に応じて、調査実施過程のさまざまな場面で生じる調査誤差、たとえば標本誤差、非標本誤差に含まれる各誤差成分の特徴を詳しく調べ、その重要度を知り、どの誤差成分をどの程度低減するか、調査全体の質を最大化するための調査設計の合理的な戦略を用意することにある。

TSE パラダイムの概念は多様であり、現在も発展中の研究分野である。これを調査方法論研究の基軸として書かれた良書 [Groves 他 (2009), Biemer & Lyberg (2003), Weisberg (2005)] や、ウェブ調査を詳しく解説した学術書もある [Bethlehem & Biffignandi (2011), Tourangeau 他 (2013), Callegaro 他 (2014) など]。

現在の TSE パラダイムに至るまでの調査誤差研究は、古くから著名な統計家らにより、さまざまな視点から行われてきた。実際に関連情報を検索すると膨大な研究報告がある。とくに、Biemer 他編 (2017) の第 1 章 “The Roots and Evolution of the Total Survey Error Concept”

に、著者らの主観的な要約との断りがあるものの、調査方法論研究における「TSE 概念の発展における重要事象」年表 (1902 年～2015 年) としてうまく要約されている。これをみると、TSE パラダイムは調査誤差研究に関するさまざまな主張を TSE の概念のもとに体系化することが目標であり、現在も発展途上にある分野であることがわかる。

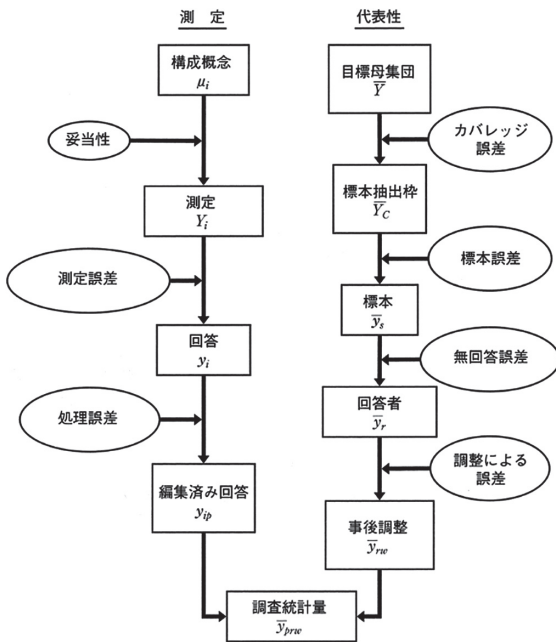
さらに、TSE の研究を促進する機構として総調査誤差国際ワークショップ (ITSEW: the International Total Survey Error Workshop) (<http://niss.org/itsew>) が 2005 年から毎年開催されている。ITSEW の趣旨は、TSE の考え方の普及活動や誤差要因を合同で調査することを目的とした研究を奨励することとある。このように大半が欧米の研究であり、筆者の知る限り、国内で TSE の観点から議論された例はほとんどみられない (稀な例に吉村 (2017) がある)。

調査実施過程における誤差発生源

調査実施過程における誤差発生源の影響を調べるには、TSE の誤差成分を分類して考える必要がある。しかし上述のように研究の方向は多岐にわたり、現時点で誤差分類を一意的に記述することが難しい。多くの研究に共通することは、調査実施過程のどの段階でどのような誤差発生源があるかを観測し、なるべく定量的に評価しようと試みていることである。

調査実施過程の一連の流れの中でどのような誤差が生じるか、これについての記述は多くの文献にみられる [Biemer & Lyberg (2003), Smith (2011), Groves 他 (2009)]。たとえば Groves ら (2009) は、調査実施過程と誤差発生源との関係を図 1 のように要約しているが、実際の調査の進捗段階にそった分かりやすい説明となっている。連続する調査過程の手順間にあ

図1 調査実施過程と誤差発生源の関係



資料：「調査法ハンドブック」（2011）より引用

る誤差という楕円は、隣り合う手順間で生じるずれを示している。このずれを最小限に抑える調査設計や推定方法を選ぶことで、目的とする調査変数の推定値の誤差を最小化できると考える。

では図1に示された誤差成分（カバレッジ誤差、標本誤差、無回答誤差、調整による誤差、測定誤差、処理誤差）は、具体的にどのような誤差であろうか。またこれら誤差成分の相互の関連をどう考えたらよいのか。TSEの誤差成分を類型化する試みは、多くの研究者によって行われてきた。これらを抜粋、要約しよう。

総調査誤差における誤差成分の類型化

前述のようにTSEパラダイムの厳密な定義があるわけではなく、TSEの類型化や、それを巡る多様な議論（とくに、弱点、長所、議論が不十分な部分）がある。また、各誤差成分を

排反的に分けることへの懸念を指摘する研究者もいる（実際、現実にかかる誤差事象をどの誤差区分に入れるか確定できない場合がある）。

周知のように従来の標本調査では、伝統的に調査誤差を標本誤差と非標本誤差に分けて考えてきた。この2つを併せて、たとえば浅井（1987）は全誤差とし、あらゆる誤差発生源で生起する誤差の合計としている。「統計学辞典」ではこれを総合誤差として扱っている〔竹内他（編）1989〕。また、「社会調査ハンドブック」では、節「調査の誤差」を設けて説明している〔林（編）2002〕。ここでは誤差を調査法の違いで生じる回答の違いであるとし、誤差は調査のすべての段階で生じること、そして標本抽出による推定誤差と標本抽出・推定の方法で生じる偏りと、非標本抽出誤差とに分けている。また調査の各段階で起こる誤差を、「調査する側」（正当な sampling と歪んだ sampling）と「調査される側」（うそ、ゆれ、でたらめ、思い違い、不注意）と分けている。また、「正当な sampling」に標本誤差と非標本誤差をふくめ、「歪んだ sampling」に「クオータ・サンプリング、電話帳、RDD、インターネット」が含まれるとしている（誤差分類の例示は国内の文献では珍しい）。これらにみるように、少なくとも国内においては、母集団から標本抽出で得た標本を用いることから生じる**標本誤差**（標本の大きさに依存の許容誤差推定、抽出法の違いなど）と**非標本誤差**（調査過程で生じるその他のさまざまな誤差）に分けて考えることが基本であり、これでほぼ対応できていた。

通常、標本誤差は調査者側で制御・管理が可能で、誤差の定量的評価が比較的容易にできるが、この低減のみでは誤差を過小評価する可能性がある。つまり非標本誤差の諸要因の影響が看過できないこともかなり昔から認識されて

いた。しかし非標本誤差は制御が面倒な要因(回答者、調査方式、調査員など)の影響下にあり一般に評価が難しい。国内の調査報告における調査誤差の記述は上に挙げたような非標本誤差の詳細な分析や説明はほとんどみられない。

国内外を問わず調査環境が複雑になったこと、とくに、新しい調査技法や調査方法論の進歩、調査方式が多様化し、調査方式を場面に応じて使い分ける必要性が増したこと(例:混合方式の適用)などから新たな誤差発生源が生まれ、非標本誤差を含めて調査誤差全体を総合的に評価することが必要になった。そしてその研究の大きな潮流が、欧米におけるTSEパラダイムとして進展してきたといえる。

誤差分類のもっとも単純な例は $TSE = \text{標本誤差} + \text{非標本誤差}$ とすることである。しかし最近の調査環境の変化を念頭にTSEパラダイムの観点から誤差成分を類型化する試みが多数ある。たとえば、以下のような提案がある。

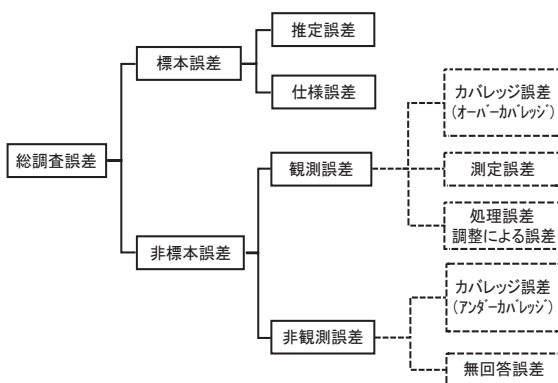
- Biemer & Lyberg (2003)、Biemer (2010) の場合: $TSE = \text{標本誤差} + \text{非標本誤差}$ とし、さらに非標本誤差を仕様誤差、枠誤差、無回答誤差、測定誤差、処理誤差と5つに分ける。
- Groves (1989)、Couper (2003, 私信) の場合: $TSE = \text{観測誤差 (測定誤差)} + \text{非観測誤差} + \text{処理誤差}$ とする。観測誤差(測定誤差)と括って「ある調査変数の推定値と母集団特性値(真値)あるいは根拠ある参照値との差」とする。Grovesは誤差成分を平均二乗誤差(MSE)を用いて評価することを重視し調査誤差の類型化を考えている。そして、非観測誤差として、カバレッジ誤差、標本誤差、無回答誤差とする。ここで標本誤差を非観測誤差に含めていることに注意しよう。
- Smith (2011) の場合: SmithはTSEの“Total”は完全性を意味し、よってTSEは可能な限

り包括的でなければならないという。実際に細かい分類図を作りTSE全体を詳しく説明している。まず $TSE = \text{標本誤差} + \text{非標本誤差}$ とする。非標本誤差を観測誤差(データ収集に関わる誤差)と非観測誤差(観測・測定に直接は関係しない誤差)に分ける。非観測誤差にカバレッジ誤差と無回答誤差を含める。また処理誤差、分析時誤差を設ける。それぞれの誤差成分の具体的な内容を、例を交えて説明している。

- Bethlehem & Biffignandi (2012)、Bethlehem (2010) の場合: オンライン調査をはじめコンピュータ支援型の調査方式の普及で、従来の調査とは異なる誤差発生源がみられるようになった。こうした事情を考慮して、Bethlehemは誤差成分を図2のようにかなり細かく分類している。図2では上に挙げたいくつかの分類と微妙に異なる部分がある。たとえば、カバレッジをアンダーカバレッジとオーバーカバレッジに分けたこと、測定誤差を観測誤差の一部に含めるなどである。

このように、調査方法論研究の進展の中で、また研究者が焦点をどこに絞るかに応じて類型化の分類基準も変化する。

図2 総調査誤差(TSE)の分類の例



資料:「ウェブ調査の科学」(2019) から引用・再編集

主な誤差成分の特徴

図2の各誤差成分は調査実施過程の中でどのように現れるのだろうか。**標本誤差**は推定誤差と仕様誤差がある。**推定誤差**は標本抽出による標本を調査することで生じるずれである。**仕様誤差**は、調査目的（企画策定時の構成概念）と調査実施者の設計内容との間の理解違いから生じるずれである。いずれも調査設計時に制御が可能な誤差であり、調査実施前に、調査目的に合った標本は入手可能か、実際に利用できる枠はあるか、どの程度のずれが生じるかを検討し誤差低減の目安を立てることが肝要である。

非標本誤差の誤差成分は細かい。まずカバレッジ（網羅）を考えてみよう。カバレッジ問題は、図3のように目標母集団と標本抽出枠（以下、枠という）との間に生じるずれに関わることである〔大隅他（2011）、大隅他（2019）〕。

日本には住民基本台帳や選挙人名簿といった優れた枠がある。欧米では、人口登録簿があるごく少数の国を除いてこうしたリストはないことが多い。しかもこれらは標本抽出用として整備されたものではない。周知のように居住者を正確に網羅しているわけではない。ある市の居

住者全員を目標母集団とし、住民基本台帳を枠と考えたとき、全居住者を網羅できないことは明らかである。たとえば、その市に「住んでいないのに名簿に記載がある」「住んでいるのに名簿に記載がない」人がいるからである。つまりカバレッジ問題が生じる。

選挙人名簿を枠として、若年層（18才未満）までを含む標本要素を抽出したくともできない。また、一部の人たち（選挙権のない人、服役中の人など）が除外される。よって日本人全体を代表する一般母集団からの枠は作れない。

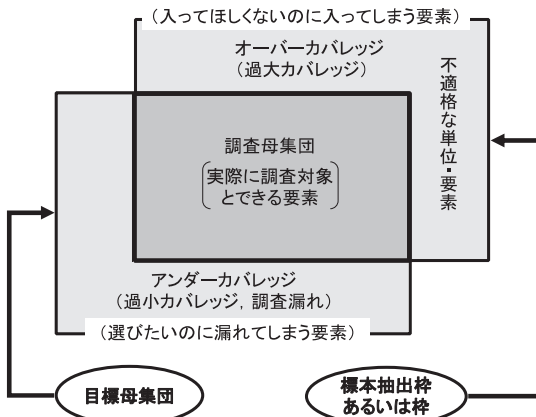
このように、カバレッジ問題は実際の調査実施場面で日常的にみられる事象で、実際に調査設計時に留意されてきたことだが、TSEの一成分としてカバレッジ誤差が表だって議論されることはほとんどない。

カバレッジ誤差には、アンダーカバレッジ（過小カバレッジ、調査漏れ）誤差とオーバーカバレッジ（過大カバレッジ）誤差がある。アンダーカバレッジ誤差とは目標母集団の一部が選択できないこと、つまり選びたい要素が漏れることで起こるずれである（図3）。

目標母集団から世帯を標本抽出したいときに、
 枠として固定電話番号簿を用いても、RDD方式で選んだとしても、電話の非保有世帯が対象から漏れる。最近では携帯電話やスマートフォンの普及で固定電話の利用率の減少傾向が顕著なので調査漏れとなりやすい。個人の抽出でスマートフォンでインターネットにアクセスする人と、そうではない人（携帯や固定）との間にもずれがあること、カバレッジの違いがあることも報告されている。

ところで、ボランティア・パネルを用いたウェブ調査で、パネル登録者数が大きいから代表性があるとか、パネル登録者の人口統計学的変数の分布、たとえば性年齢区分分布が、国勢調

図3 目標母集団、杣、カバレッジの関係



資料：『ウェブ調査の科学』（2019）から引用

査のそれに類似しているから代表性がある（一般母集団に近い）といった意見をみることもある。これをカバレッジの観点からみると不確かな情報であることは明らかである。標本の大きさの大小で推定値の特性を評価できるのは確率標本を用いたときだけである。また目標母集団としたいインターネット母集団が特定できないボランティア・パネルは非確率的パネルであり、回答者数が多いからといって誤差推定や推定値の精度の正確な評価はできない。

通常のRDD方式電話調査でも類似の現象が生じる。つまり、モバイルユーザ（スマートフォン、タブレット端末など）が正確に捕捉されるか、そしてモバイルユーザが加わることでカバレッジが改善されるか、回答の内容に差がないかどうかといった問題である〔Couper 他報告としてBiemer 他編（2017）の第7章〕。

オーバーカバレッジ誤差とは（図3）、調査対象に含まれていても目標母集団に含まれない要素がある場合に生じる誤差、枠の不具合から生じる誤差のことで**枠誤差**ともいう。ある商品の製造会社が会員登録で集めたりリスト（枠）を用意し、その商品の評価をウェブ調査で行ったとする。自由参加で行うとその商品の非利用者も枠に入るだろう。これがオーバーカバレッジである。事前に商品の利用者かどうかを尋ねる、あるいはユーザ登録簿と照合し不適格な回答者を除外すればオーバーカバレッジは改善される。

電話調査では、RDD方式で生成した番号（概念的リスト）を枠としたとき、その枠に実際に存在しない番号がある場合にオーバーカバレッジが生じる。ウェブ調査では、適切な枠を使用しても（かりに作れたとしても）オーバーカバレッジ誤差が発生することがある。ボランティア・パネルに基づくウェブ調査の場合、オーバーカバレッジはさらに深刻な問題となる。この

場合、誰が調査に回答するかは制御できない。

測定誤差はもっとも一般的な誤差成分であり、回答者の回答が真の回答と異なる場合に生じる誤差の総称である。測定誤差の主因は、用いる調査方式、回答者、調査用具（調査票、PCなど）、調査員などである。調査方式効果、調査員効果、調査票設計（質問紙と電子調査票、デザイン、文字や色彩、グリッド、分岐質問など）、質問文の影響（ワーディング、文脈効果、回答選択肢・尺度の作り方）、記憶効果などが関連するが、これらについて多くの研究報告がある。

回答者が質問を理解していない、真の回答を知らないあるいは回答したくない、意図的に誤った回答をするなどが誤差を生む。**調査員効果**として、話し方、過剰なブローニングや回答誘導、外見、態度、指示書違反、誤記入、データ改ざん、誤記・誤入力、PCの誤操作などがある。**調査方式効果**として、調査員関与の調査（電話、面接）と自記式（郵送、ウェブ）との回答の違い、守秘事項や**微妙な内容の質問**（違法薬物使用経験、犯罪歴などを問う質問）の回答で起こる偏りもこれに含まれる。

測定誤差はいずれの調査方式でも観察されるがオンライン調査では従来の調査方式（郵送、面接、電話）では生じなかった事象が多々みられる。調査における測定器は**調査票**である。オンライン調査（ウェブ、モバイル）では、単純な調査票から高度に洗練された双方向的な様式の調査票まで、多様なデザイン（ボタン、チェックボックス、グリッド、プルダウン・メニュー、ドロップダウン・ボックスなど）、マルチメディア（画像、動画、アニメーションなど）の利用が可能である。この可能なこと、設計の自由度が高いことが回答行動の捕捉を複雑にし、無回答を増やす原因となる。多くの実験調査で、調査票設計はなるべく簡潔に、機能を単純化し、

グリッドや分岐・パイピングは多用しない、自由回答質問は控えるなどが誤差低減には必要とされている [大隅他 (2017), 大隅他 (2019)]。

回答者が質問に答える回答過程は複雑であり回答行動は測定誤差に影響する。調査測定は人の心理的、認知的側面に関わることから、心理・社会心理系の研究者らが関与した調査研究も多い（とくに態度や意見を問う調査）。実際にこうした分野で検証されてきた誤差生起に影響する回答行動が多々報告されている。たとえば、**労働最小化行動**（一種の手抜き回答）、**非識別化**（特定の回答選択肢だけを選ぶストレーティングなど）、**黙従傾向**、**社会的に望ましい回答の偏り**、そして**回答の順序効果**（初頭効果や新近性効果）、**DK効果**などである。これらは誤差の発生量（比率）として具体的に定量化できるものから、それが難しいものまでさまざまである。

ウェブ調査では、かりに回答が得られてもその内容について深刻な問題も生じる。いい加減に答える謝礼目当ての**プロの回答者**や回答時間が短い**速度違反者**などを捕捉し、怪しい回答を除外するなどの手当が必要である。こうした回答の不具合は、正しい回答が得られない（強いあるいは弱い）労働最小化行動、ストレーティング、初頭効果などの現象となって現れる。

観測誤差の**無回答誤差**も重要な誤差である。これは、回答者が必要な情報を提供しない、あるいは収集した情報が無効であるときに生じる。無回答は用いる調査方式に関係なくほぼすべての調査で発生するが程度や内容には違いがある。また調査方式によって無回答率に差があり無回答の理由も異なる。無回答は大別して、質問のすべてで回答が得られない**調査不能**（完全項目無回答）と、質問項目の一部だけ回答を得られない**項目無回答**がある。

無回答誤差は回答者行動に関わるので、上述の測定誤差の諸要因と重なる部分が多い。このことから、前述のように測定誤差に無回答誤差を括って**観測誤差**と考えることもある。周知のように、調査不能があれば回収標本の大きさは計画標本のそれよりも小さくなるから推定値の精度が低下する。無回答の重要な問題は（他の誤差以上に）、母集団特性値（真値）の推定値に大きな偏りを生むことである。たとえば、回収標本の人口統計学的変数の分布に大きなずれがあれば過大あるいは過小推定が生じるおそれがある。ウェブ調査などで回収標本の属性分布が計画標本と大きくずれたときなどにもこれが生じる。無回答にはさまざまな原因があり、それぞれ推定値への影響が異なるので、これらを区別して扱うことが重要である。無回答の主因は、**非接触**、**回答拒否**、**回答取得不能**である。

無回答誤差を低減するには、まず回答者の理解、信頼を得ることであり、回答負担を軽減することである。**督促**を行う、**謝礼**を用意するなどの対策も、誤差低減が期待できる。とくに自記式調査（郵送、ウェブ）では調査依頼時（事前）の謝礼が効果的であることが多くの研究で報告されている。一般に、調査員が関与する調査（面接、電話）にくらべ、自記式調査（郵送、ウェブ）の無回答率が高い傾向にあることも多くの研究で指摘されている。

枠として電子メールアドレス・リストを使い電子調査票を送って未着となると**非接触**となる。メールアドレスのミスや変更、スパムメール監視ツールで受理遮断、回線不備などで生じる。これは、郵送調査での調査依頼状の未着や電話調査の応答なしなどの非接触の変形である。どの調査方式（電話、面接、ウェブ）でも、調査対象者と接触ができて、いろいろな理由で**回答拒否**となることがある。ウェブ調査の場合、

事前に郵送で依頼状を送っても、無視される、回答を忘れるなどが起こる。こうした場合、電子メールや郵便葉書による督促や謝礼（ポイント提供）などが回答率改善に有効な場合もある。

回答取得不能は回答者が回答する意思があるのにそれができない場合に起こる。ウェブ調査の場合はこれが顕著で、インターネットの利用環境の不備（回線接続や速度の問題、ブラウザの不具合など）、不適切なナビゲーション、さらに利用者のデジタル・ディバイドの問題など様々な理由で回答取得不能が起こる。

社会的望ましさによる回答の偏りは、調査員関与の調査（電話、面接）で起こりやすいが、自記式（郵送、ウェブ）では少ない傾向にある。とくに自記式調査では、多くの研究報告によると、微妙な内容の調査では、他の調査方式（電話、面接）よりも無回答が少ない傾向にある。

このほか、**処理誤差**（収集データの入力、編集・加工・作表集計処理、コーディングやポストコーディングなどの誤り）や**調整による誤差**（加重調整や補定など調査の事後処理で生じる誤差）がある。

以上の説明のように誤差成分は多岐にわたり非常に細かい。こうした誤差の誤差発生源を丁寧に調べ、重要度を勘案しながら誤差低減を行うことが調査の質の改善につながる。

調査結果の評価

誤差発生による影響の程度を定量的に図る指標、調査報告時に用いる指標（比率など）がいろいろある。たとえば、**回答率**は調査の質を測る代表的な指標である（実は、あまり適切な用い方がされていない）。AAPOR（米国世論調査学会）では、たとえば、用いる調査方式や回答取得の細かい場面（回答数、無回答数、拒否、中断、接触不能など）に応じて6種類の回答率

算出のルールを示している。これ以外にも、**協力率**、**拒否率**、**接触率**といった指標を用いることを勧めている。ウェブ調査では（適格な枠が作れないため）回答率の正確な算出が難しいので、**参加率**あるいは**完了率**を用いることが勧められている。国内の調査で、こうしたきめ細かい評価指標が調査報告に記述された例はほとんどみられないようである。

ウェブ調査は一般に他の調査方式に比べて回答率が低い傾向にある。一方、他の調査方式では測定が難しい**種々の指標**、たとえば調査票の閲覧率（接触・アクセスのみ）、勧誘率、累積回答率、到達率、中断率、脱落率、回答所要時間、調査票内のページ遷移情報（選択肢を選ぶ、ページを移る）などを回答行動の電子的追跡で**パラデータ**として測定できる〔大隅他（2017）〕。同じ自記式である郵送調査とくらべてはるかに情報量が多い。ウェブ調査やモバイル調査は、誤差要因が複雑でしかも増えているが、誤差評価つまり質の評価の点では長所が多いことを示している（しかし、実際には利用されていない）。

ここまで述べたことから、調査方式と誤差発生にはそれぞれ固有の特性があり、単純に優劣を比較することは慎重に対応すべきことがわかるであろう。**ボランティア・パネル**を使うウェブ調査は迅速・廉価・簡単に実施可能だが、調査の質に問題がある。一方、一般母集団を対象に**確率的パネル**を作り、インターネット利用者だけでなく（PCやタブレット端末を貸与した）非利用者までを含むウェブ調査を行えば、費用も手間もかかるが、調査の質は確保されるだろう。実はかつて国内にはこうした確率的ウェブ・パネルがあったが現在は皆無であるが、国外には現在もいくつか存在する〔大隅他（2019）〕。また、同じ自記式であるという理由で、ウェブ調査と郵送調査の比較実験調査が多数行われて

きたが、単純に比較を行うことは注意が必要である。とくに最近、郵送調査が好ましいという意見がみられるが、調査内容と目的によるのであり、十分な証拠固めが必要である。

むすび ―なぜTSEなのか―

TSEパラダイムの特徴は、調査実施過程で生じる種々の誤差を成分に分け、それぞれの誤差特性を具体的に精査し、できるだけ定量的な指標に表して比較することにある。また、誤差成分の難易度を勘案し、扱いの容易な誤差から誤差低減をはかることで調査の質を全体として改善することにある。前述のように、調査の質と投入資源・経費の均衡をどう図るかは非常に難しい課題である。理想と現実の乖離をどのように埋めるか、これの判断基準としてTSEパラダイムの概念を有効に活用することが肝要である。

調査実施環境だけでなく、調査データの形態や利用方法も変容している。ビッグデータなどの新しいデータ資源の登場は新たな誤差構造の出現につながるだろう。さまざまな分野の保有データ（リポジトリ、データアーカイブ、オープンデータ）や諸官公庁のマイクロデータなどのデータ集合、調査情報（調査設計、調査票、報告書など）の二次利用や共有化が分野横断的に進んでいる。こうした**調査データの開示**に伴い、調査の質を具体的かつ明らかに示すだけでなく、**調査倫理**や**データ保護・守秘性**、著作権、さらに**個人情報保護**に関する検討も求められる。データの利用形態の変化に伴い、調査の質の一貫した比較・評価のためにTSEパラダイムの適用が有効と考えられ、実際に欧米ではこうした分野の研究も進んでいる[Biemer他編(2017)]。

欧米の研究者らは、ほとんどが官庁統計またはそれに関連する機関・機構（センサス局をは

じめ各機関の統計部門など）とも密接な関係にあり、また市場調査や他分野との横断的な共同研究にも積極的である[大隅他(2011)の「監訳者まえがき」を参照]。TSEパラダイムも産官学の共同体制が整ったなかでの研究成果である。

日本国内の調査方法論分野における過去の調査誤差研究を振り返ってみると、確率標本を前提とする伝統的な調査設計に基づく標本誤差評価に軸足をおいたアプローチが多かったように感じる（もちろんこれは重要な要素である）。また、我々が非標本誤差にかかわる諸事象の実証研究に無関心であったわけではない（たとえば調査不能に関する研究など多々ある）。しかし、欧米のように体系的かつ組織的に調査誤差研究を進めてきたかという点で疑問がなくもない。情報通信技術などの急速な進歩で調査環境は確実に変容している。とくに調査方式の主流となったコンピュータ支援の調査方式（CATI、CAPI、IVR、ウェブ調査、モバイル調査）と従来方式（面接、郵送、留置、電話など）の比較研究や**混合方式の実用化**は不可欠だが十分とはいえない。

最近、この複数の調査方式を用いる混合方式への関心が高まっている。混合方式を用いる理由として、データ収集費の削減、目標母集団のカバレッジの向上、回答率の向上、測定誤差の低減などが挙げられる。調査方式の組み合わせ方、実施順序（同時並行的か、逐次的か）、投入経費の配分と投入時点、回答者の調査方式選択権の有無、など、さまざまな検討事項がある。そしてここでも、調査内容の適否を評価するためにTSEパラダイムに則った検証が不可欠である。たとえば、調査方式の変更はカバレッジや無回答に影響し、測定誤差の評価が必要となる。そして調査全体の誤差を最小化するか、調査方式間の違いを最小化するかなど、状況に応

じた戦略策定が必要となる [Tourangeau 他報告として Biemer 他編 (2017) の第 6 章]。

住民基本台帳などの利用制限から、商用調査では、便宜的標本、有意抽出標本を用いる非確率的アプローチや定性調査が多くなっている。ボランティア・パネルによるウェブ調査がコモディティ化し、さらに新たな調査方式 (オンライン・フォーカス・グループ、バーチャル面接) も登場している。

こうした調査環境の大きな変化の中で、すべてに通底することは、総調査誤差パラダイムの重要性を認識し、「いかに調査の質の向上・改善を図るか」ということにある。

<補注>

本稿に登場する用語句には、あまり馴染みのないものもあるかもしれない (たとえば、調査方式 (モード)、参加率、カバレッジ、労働最小化行動、非識別化、ストレートライニング、パラデータなど)。これらについては、大隅他 (2019) に付録として付けた用語集を参照されたい。

<参考文献>

- 1 AAPOR (2016), Standard Definitions - Final Dispositions of Case Codes and Outcome Rates for Surveys -, Revised 2016.
https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf
- 2 Bethlehem, J. (2010). Selection Bias in Web Surveys, *International Statistical Review*, **78** (2), 161-188.
- 3 Bethlehem, J. and Biffignandi, S. (2011). *Handbook of Web Surveys*, John Wiley & Sons.
- 4 Biemer, P.P. (2010). Total Survey Error: Design, Implementation, and Evaluation, *Public Opinion Quarterly*, **74** (5), 817-848.
- 5 Biemer P.P. and Lyberg, L.E. (2003). *Introduction to Survey Quality*, John Wiley & Sons.
- 6 Biemer P.P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L.E., Tucker, N.C., and West, B.T. (eds.) (2017). *Total Survey Error in Practice*, John Wiley & Sons.
- 7 Callegaro, M., Baker, R., Bethlehem, J., Göritz, A.S., Krosnick, J.A., Lavrakas, P.J. (eds.) (2014). *Online Panel Research: A Data Quality Perspective*, John Wiley & Sons.
- 8 Groves, R.M. (1989). *Survey Costs and Survey Errors*, John Wiley & Sons.
- 9 Groves, R.M., Fowler, F.J., Couper, M., Lepkowski, J.M., Singer, E., Tourangeau, R. (2009). *Survey Methodology*, second edition, John Wiley & Sons.
- 10 Groves, R.M. and Lyberg, L. (2010). Total Survey Error - Past, Present, and Future, *Public Opinion Quarterly*, **74** (5), 849-879.
- 11 Smith, T.W. (2011). Refining the Total Survey Error Perspective, *International Journal of Public Opinion Research*, **23** (4), 464-484.
- 12 Tourangeau, R., Conrad, F.G., and Couper, M.P. (2013). *The Science of Web Surveys*, Oxford University.
- 13 Weisberg, H.F. (2005). *The Total Survey Error Approach: A Guide to the New Science of Survey Research*, University of Chicago Press.
- 14 浅井晃 (1987, 1992): 調査の技術, 日科技連出版社.
- 15 大隅昇 (監訳) (2011), 調査法ハンドブック, 朝倉書店. (*) 文献 9 の 2004 年版の完訳.
- 16 大隅昇, 鳩真紀子 (2012), 「総調査誤差」をめぐってーロバート M. グローヴス, ラース・ライバーク論文「総調査誤差ー過去、現在、未来ー」を中心にー, 「よろん」日本世論調査協会報, 第110号, 18-31.
- 17 大隅 昇, 鳩真紀子, 井田潤治, 小野裕亮 (訳) (2019), 「ウェブ調査の科学ー調査計画から分析までー」, 朝倉書店. (*) 文献12の翻訳版.
- 18 大隅昇, 林文, 矢口博之, 箕原勝史 (2017). ウェブ調査におけるパラデータの有効利用と今後の課題, 社会と調査 (特集: パラデータの活用に向けて) No.18 (2017年3月), 50-61.
<https://wordminer.org/wp-content/uploads/2017/02/a5fc04447aa50d0b3b43e6b86225f01f.pdf>
- 19 竹内啓他 (編) (1989). 統計学辞典, 東洋経済新報社.
- 20 林知己夫 (編) (2002). 社会調査ハンドブック, 朝倉書店.
- 21 吉村治正 (2017). 社会調査における非標本誤差, 東信堂.

(おおすみ のぼる 統計数理研究所・名誉教授)