

インターネット調査の適用可能性と限界

— データ科学の視点からの考察 —

大 隅 昇

Internet Surveys: A Review of Several Experimental Results
— Applying Data Science Approach to the Exploration of Internet Surveys —

Noboru OHSUMI

Rapid development in the survey methods available on the World Wide Web (WWW) is having a major impact on conventional survey data collection methods. The wide range of opinions has given rise to an ongoing debate regarding the future role of Internet surveys (in particular, Web surveys) based on the role that self-administration will play in research. We started by arranging a practical procedure for electronic data collection on the Web surveys experimentally designed from the viewpoint of "data science."

Aiming to verify the applicability, possibilities, and limitations of Web survey methods, we conducted three experimental surveys during the period from 1997 to 2000. They were designed to enable comparison with each other and with traditional methods such as face-to-face interviews and online surveys using conventional sampling procedures. These surveys provided informative results about the characteristics of Web surveys.

In the first survey, consisting of 12 continual surveys of a single panel of registrants, we examined the relationship between the response rates and the questionnaire's design, volume and content, as well as response rate differences among the 12 surveys and the discrepancies in repeated surveys. In the second experimental survey, we carried out Web surveys at about the same time on three different sites together with non-internet surveys using conventional sampling methods. Our experimental design enabled objective comparison of the surveys by using as much identical questionnaire design as possible. Our experimental surveys showed that Web survey results are similar to each other while distinctively differing from those of conventional surveys. In the third experimental survey, we simultaneously carried out a series of comparative surveys in order to examine the general characteristics of Web surveys found during the second experimental survey. Except that the number of sites used was two instead of three, the third experimental surveys were carried out in the same way as in the second trial. We confirmed the results that the same characteristics were evident again in the second survey. We also found that how the registrants of the surveys (named "resources") were selected and whether the interval between solicitation and survey was short or long would be factors influencing the answers and response-rate. We also found that the respondents do not necessarily represent the resources.

In addition, as an addendum in this paper, we report partly the results of a fourth experimental survey which has been carried out in 2001 to 2002 and compare it with the findings of the previous three trials. In particular, we also analyze the itemized causes of "nonresponse" on the datasets obtained from the tracking procedure of tracing electronically each respondent on the WWW. The fourth survey consists of Web surveys on three separate sites while the other surveys were based on conventional sampling methods (e.g., face-to-face interviews and mail

surveys). While we use the same questionnaire design, content, and duration as those used in the past surveys, we also attempt to examine how the questionnaire design has influenced responses. Through these experimental surveys, an appropriate route to how to design a Web survey, evaluate its quality and avoid possible risks or perils in design is proposed from the concept of "data science."

1. ま え が き

最近、調査環境が大きく変化したと言われる。ここでいう「調査」とは、いわゆる社会調査(意識調査、態度調査等)、そして俗にアンケート調査と言われるような分野を指す。とくに、インターネットの普及により、電子的に調査データを取得する環境が整い、いわゆるインターネット調査が今後の調査の主流のように考える風潮すらある。

インターネット調査(Internet survey)とは、未だ明確な定義があるわけではないと筆者は考えるが、一般には、電子メール調査(e-mail survey)やWeb調査(Web-based survey)と呼称される調査法を言う。また、インターネット上で行うアンケート調査として、既に国内には無数のアンケート・サイトがWWW上に登場しており、その数はちょっと数えてもゆうに百数十を越える。その多くは、Webページ上に調査票を掲載して回答を求めるWeb調査型が多いのであるが、電子メール方式を併用する場合もある。

インターネット調査については、いまだ不透明な要素が多く、従来からの統計理論に支えられた科学的な調査法とは一線を画して考えるべきものであるとの指摘も多く、筆者も同じように考える一人である。しかしながら、論評や批判はあっても、これを研究対象として真剣に取り組む姿勢は、ほとんど見られなかった。

実は、これに関連して、数年前に文部科学省科学研究費の助成を受ける機会を得たことをきっかけに、実験調査を行う機会を得て、また調査機関・企業の協力により、一連の実験調査を行うことができた。これは、インターネット調査の調査法としての問題点がどこにあるかを検証する国内で始めて試みであり、その実験調査の結果は、既にいくつか報告され、また引用もされている^{〔16〕,〔20〕他}。

その一つとして、1999年に統計数理研究所の公開講演会において「インターネット調査とそれを巡る諸問題」をテーマに、次の四つの講演が行われた(1999年11月)。

- (1) 調査環境の変化と新しい調査法の抱える問題
林 知己夫(統計数理研究所)
- (2) 電子調査,その周辺の話題——電子データ取得法の現状と問題点——
大隅 昇(統計数理研究所)
- (3) マーケティングにおけるインターネット調査の現状と課題
横原 東(電通リサーチ研究開発部)
- (4) インターネット調査にみられる回答者像,その特性
吉村 宰(岡山大学教育学部)

この主旨は、その後、「統計数理」に要旨記事として掲載されている^{〔17〕,〔18〕,〔21〕}。ここで、筆者は、かなりのページを割いてインターネット調査の現状の問題点や実験調査から分かってきた特性等に触れた。また、2000年11月には、(社)輿論科学協会の創立記念講演会の演者として、やはりインターネット調査の実状について述べる機会を得た^{〔22〕}。本報告は、これらの場で述べたことを元にし、整理したものである。

いずれにしても、筆者はインターネット調査について非常に慎重な立場を取っており、また従来型の調査法との関係にこだわってきた。換言すると、従来型の調査方法からインターネットを観察したときに、あるいは両者を比較したとき、その間に一体どのような類似性や差異があるのかが、まず初めの関心事であった。現状のインターネット調査の使われ方には、少なくともこうした視点はなく、もっぱら「早く、迅速に」「簡単、平易に」「安く、安価に」実施できて、しかもそれなりのサンプル数を確保できる調査法として認識されてきた。しかし、ここに大きな誤り、誤解があるのではないか、というのが筆者の主張であり、また懐疑的に思ってきたところである。しかも、ここ数年、IT(情報技術)の環境が激変し、いわゆるITバブルとなって、あちこちでIT関連企業の経営状況が危うくなり、インターネット調査についても、その多くが、投資効果に対する収益性が深刻な問題となっている。しかし、マーケティング・リサーチの世界では、インターネッ

る。この調査の特徴は回答個票まで含む全情報が、開示されていることにある。そのことから、このサイトにアクセスする人が非常に多い。その他、FIND/SVPによる調査 (American Internet User Survey ; <http://findsvp.com/>、Clemente^[2]) やニールセンの調査 (Nielsen Internet Demographics Recontact Study) 等もある。もちろん、この他にも調査会社やマーケティング企業が独自に行っている無数のインターネット調査やそのサイトがある。また、かなり前から、いわゆる電子メール調査を初め、後述する電子的調査情報取得の研究に関連して、無数の関連研究が報告されている。また PC 側の情報としてマシンの何台あるか、サイトがいくつあるか、という調査 (たとえば、Internet Domain Survey ; <http://www.isc.org/ds/>) もあり、こうした結果は多くの場合広く開示されている。

一方、日本国内の事情をみると、まずインターネット・ユーザー数が常に問題になるが (つまり調査対象とすべき母集団とは何かを考える意味で)、その推定数さえ、『平成 13 年情報通信白書』(総務省, 2001) と『インターネット白書 2001』(インプレス, 2001) とを比べただけで、すでに千数百万人も違っており、これを利用すること自体に疑問がある。この他、Nielsen/NetRatings、ピデオリサーチネットコム等多数の類例の推計があるが、大同小異、いずれもその推計値には差異が大きい⁽¹²⁾。加えて、昨今の携帯電話の普及により、携帯電話上でのインターネット利用が日常的となったことから、こうした統計指標の信憑性が怪しくなっている。

また、近時点の情報として、『インターネット白書 2002』(インプレス, 2002) があるが、ここでは約 4,620 万人のインターネット・ユーザー数があると推計している。しかしそこで行われた調査方法も、推計値の一つの例にすぎないと思われるべきである。また、携帯電話の普及で、とくにインターネット環境の利用が可能となったことで、これを介してのインターネット・ユーザー数が急増し、これも合わせると、6,000 万人を越えるのではないかという推計もある。

しかも、実態はたえず動いており、こうした状況下で全体のインターネット・ユーザー数を正確に捕捉する手段はない。つまりは、調査を行う際に拠り所となる母集団に相当する情報が既に曖昧なのである。しかしそもそも、インターネットの仕組みそのものを考えたとき、母集団云々という議論がおかしいと言う意見もある。この点で既に従来型調査と比較する意味がは

たしてあるのか、という素朴な疑問も生まれるのである。このことをもって、そのような性格を備えるからこそインターネット調査は信頼できないとの意見が出る一方、いや、それだからこそ新しい調査法として別の視点から検討すべきであり、利用法も異なるのであるから比較することに既に意味がない、という意見の対立が生じる。

現在、インターネット・ユーザーの特性について、一般的に言われていることは、年齢は次第に若年層・高齢層の両裾に広がっていること、性別も女性が増えて、いまや家庭の主婦にまで及んでいること、また、職業も特定の業種 (技術職など) から多様に広がる傾向にあること等が指摘されている。また、ネットワーク技術が進歩し、誰もがコンピュータを持つようになり、インターネット・ユーザーが国民全体を代表するようになりそうだと、というようなことがいわれるようになってきている。しかし、こうした議論や意見には、明らかな論拠がなく、現時点では筆者には納得できない事が多すぎる。

3. インターネット調査に対する欧米の考え方の差異

ところで、筆者の理解では、日本は欧米とはそもそも調査環境が大いに異なり、そのことがインターネット調査をどう考えるかという根本的な部分で食い違いを生じていると思われることである。日本のように、従来型の調査、つまり、住民基本台帳や有権者選挙人名簿に則ってサンプリングを行ったり、調査対象を捕捉するような調査が可能な環境は、世界的にみて珍しいことであろう。欧米、とくに米国ではこのようなことができないので、当然、別のことを考えるわけで、そのような研究のカテゴリーの一つが、CASIC (Computer-Assisted Survey Information Collection) であり CADAC (Computer-Assisted Data Collection) である (Couper^[6], Saris^[10])。これらは電子的にデータ・情報をどう取得するかの方法論の研究であり、行政統計資料の電子的収集、電子政府移行も含めて、あらゆるものがこういう枠組みの中で研究されている。インターネット調査の場合にも、フェイス・トゥ・フェイスではなくて、マシン・トゥ・マシンを前提とする調査法であるとの認識から議論が始まっている。いわゆる RDD 方式による電話調査も、そのような枠組みの中で議論され、それが CATI (Computer-Assisted

Telephone Interviewing) のような具体的な調査方式として定着してきたと考えられる。これが、米国における研究の方向だと筆者は受けとめているわけだが、一方、日本の場合は、これとはかなり異なることに留意せねばならない。日本では、今までは統計的な調査方法に忠実な、すぐれた調査環境を利用できたことから、電子的に調査データを取得するという CASIC のような発想やそれに関する研究が遅れていたと言わざるを得ないのである。

実際、CA、つまり「コンピューター・アシステッド (Computer-Assisted)」の付くシステムが欧米には無数にある。その中の一つに CATI (Computer-Assisted Telephone Interviewing) があり CAPI (Computer-Assisted Personal Interviewing) がある (de Leeuw^[7])。また、日本国内ではほとんど知られることのなかった「ディスク・バイ・メール方式」(disk-by-mail) という、フロッピーディスクに調査票を入れ、郵送で (つまり、snail で) 情報をやり取りする調査方式まで利用されていた。また、早い時期に、eメール調査が利用されてもいた (de Leeuw^[7])。こうした広い視点、しかも長い時間軸の中で、展開されてきた調査法の歴史の延長線上にインターネット調査がある。とくに、電子化された (RDD 方式に基づく) 電話調査の延長線上に、インターネット調査を位置付けてきたことがある。

4. インターネット調査とその分類

インターネット調査とは、一般的には、「電子メール調査・eメール調査」(e-mail survey) や Web 調査 (Web-based survey あるいは Web survey) がある。同時に、電子調査 (electronic survey) という言い方もよく引用されるが、この範疇で考えれば、電子的媒体を使うほとんどの方法はここに入る (たとえば、従来からあるファクシミリ調査)。また、インターネット調査と従来の調査法とは、本質的にその仕組みが異なるものである。従って、前述のように、従来の調査と比べること自体、本質的に意味がないとの意見もある。実際に、筆者の知る限りは、いわゆる統計学の分野の専門家が、インターネット調査に深く関わってきた例は少なく、ときに「インターネット調査は駅前調査である」と言ってはばからない人もいる。一方、調査関連分野、マーケティング・リサーチ分野では、インターネット調査と称した (しかし、実態はとても調査とは

言えないような) 行為を無節操に行っている場合も散見されるのが現実である。

インターネット調査の特性を知るとき、あるいは検証する際に、留意せねばならない事項は無数にある。つまり、それくらい分からないことが多い、あるいは多かったということである。筆者らが、1997年から2000年にかけて行ってきた実験調査の結果から (第1次調査～第3次調査まで)、どのような事象が生じ、またどんな事に留意すべきかについては、折りをみて報告してきた (報告書^[19]等を参照されたい)。ここでは紙面の制約もあるので、主要なキーワードを述べるとしよう。

まず、インターネット調査で必ず問題となるのが、回答者の捕捉方法である。いわゆるデータ取得は「セルフ・セレクション」(self-selection; 自己参加型) が一般的である。後で示す例の多くはセルフ・セレクションで行った調査結果である。セルフ・セレクションは回答者が自主的に手を挙げて答えるということであり (「この指とまれ」方式)、また、多くは統計的な意味でのサンプリング操作を伴わない。したがって、登録者集団 (筆者等は、これを独自に「リソース」と名付けた) を、いかに構築するかということに完全に依存すること、またリソースの特性をどう評価するのかが問題となることがある。たとえば、その鮮度 (リソース構築の時期が新しいか、古いか) の見極めといったことがキーになる。実は、インターネット調査を左右する答えの一つがこのリソース構築であることが、実験調査から見えてくるが、こうした実査現場における情報は一般にはほとんど知られることがないことに、現実の問題がある。

では、このことを念頭に、インターネット調査にはどのような種類があるのかが次の問題となる。もちろん、インターネット調査方式の分類は、いろいろな分け方が考えられ、また実際に様々の研究がある (たとえば、Couper^[5], Jones^[9], Dillman^[8])。異論もあろうかと思うが、今までの実験調査の経験から得た日本国内の事情を考慮し、Web 調査の形態を「回答者の捕捉のあり方」から分類すると、パネル型、リソース型、オープン型の3つのタイプに分けられる (図1)。オープン型が果して調査といえるのかどうかは問題ではあるので、ここでは主にリソース型をどう考えるかが要点になる。なお、リソース型を、リソース内オープン方式、属性絞り込み方式、リソース内サンプリング方式と分けたが、リソース内サンプリング方式 (構築したリソースからサンプリングを行うという方式) は筆者等が始

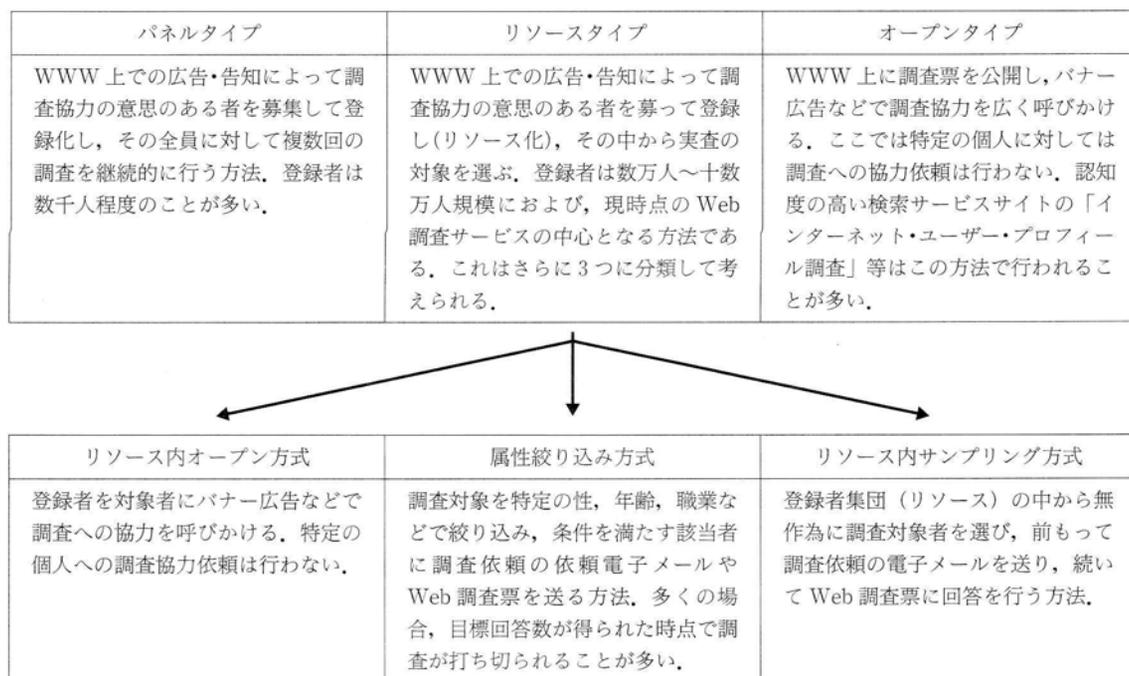


図 1. Web 調査の分類

めて試みた方式であり、現在のところ、これは他では使われていない。一般に多く使われている方式は、絞りこみ方式やリソース内オープン方式である。ここでは、回答者の絞りこみを行い、たとえば男性 20 代が 100 人集まったら終わり(打ち切り)、というような方式を言うが、こうした調査対象者捕捉法を用いるものを調査というべきか、熟考を要する。

ここで、リソースの構築ひとつを考えても、一時的に時間を止めなければ数字(登録者数)が確定しない。しかしどこかで時間を止めると、その後にメールのアドレス変更等が生じる等、様々のトラブルが生じる。大まかに言えば、筆者等が行ったことは、リソースを決め(一時的に登録者集団を固定し)、それからサンプリング(無作為抽出)を行うが、その際に登録者の合意を取って、その合意者に調査票を発信するという方式をとるとのことである。

インターネット調査の最大の利点とされてきたことは、「速い(迅速)」、「簡単(できる)」、「経費が安い」といったことであるが、以上の操作を忠実にすることを考えただけでも、そう容易ではなく、さらに、調査票の回収時あるいは回収後に生じる様々な事象を考える

と、決して迅速にできるものでもない。また質の高い調査データを取得するためには、インターネット調査システムを構築するための基盤整備が必須であり、すなわち調査経費が安価であるということも疑わしい。

たとえば、今までの実験調査で体験した諸事象の幾つかを列記してみる。最終的にデータを確定するまでにどんな検討事項が出てくるかという、まず登録者集団(リソース)から必要な計画標本数をサンプリングするが、既にここでアドレス変更があり、未着がある、したがって、有効に発信された数はいくつあったか分からないことがある、どうかすると途中で登録解消が入ってくることもある。また、調査サイトのシステム設計にもよるが(最近では技術改善により、こうした事象は少ないのだが)、回答が重複して届くとか、未登録者が介入することもある。つまり、われわれが従来やってきた調査のように、回収率すら正確には定義できない状況が生じる。このため、回収率ではなく、有効回答数(率)、有効参加数(率)等を考えねばならない。もちろん、サイトの設計によってこれらの状況は異なるのだが、こうした情報はほとんど開示されることがない。つまり、重複回答の確認、代理回答・不正回答・

虚偽回答の確認、(計画サンプル、つまり発信対象者)の中の未回答数の確認、接触数(調査票を見ただけで回答しない数)、回答数の最終確定方法と有効回答数の確認、等々挙げると際限がない。また、調査内容・結果に影響を与える要素として、インセンティブ(謝礼)、匿名性や多名性等も検討事項となる。

この他、調査計画全体が、コンピュータ上で行われることから、とくに管理サーバマシン上のトラブルという、インターネット調査特有の現象も無視できない。たとえば、サーバマシンの突然のダウン、調査票設計ミスから回答取得がこぼれてしまう現象、回答者側のマシンとサーバの不具合による回線切断現象、等々、これまた様々な現象を体験することとなる。さらに最近では、登録者の個人情報の守秘(プライバシー保護問題)、ハッカーなどの攻撃から保護するためのセキュリティ管理、セキュリティ・システムの実装や暗号化処理技術への対応などへの配慮も必須の要件となる。

要は、これらの事象や現象が、現実には表に現れることなく、インターネット調査の陽の部分、利点のみが強調されることに問題がある。少なくとも調査法としての信頼を得るためには、負の部分にも言及し、それらをどう解決するか議論がなされねばならないが、現状はきわめて貧しい状況にある。筆者等が、複数の調査機関・会社とのコラボレーションを計画し、また調査現場における実験調査を計画した理由の一つは、ここらを検証し可能な範囲で情報開示を行うことを考えたからである。実際、実験調査に参加いただいた各機関には、こうした情報の取得を含めて、きわめて好意的にご協力をいただいたことは幸運であった。いずれこうした情報の開示を行うことのための、準備を進めているところである。

この他の問題として、母集団をどう考えるか(現実にはほとんど不可能である)、回答の代表性等々いろいろなことが言われている。それに対して、どのようなアイテムをどう扱わなければならないか、ということの実験調査を通じて筆者なりに整理してきたが、紙面の都合でここではこれ以上は取り上げない^{[20],[25]}など参照)。インターネット調査を調査法の一つとして利用するというからには、こうした視点で、その特性を客観的に検証することが肝要とだけ指摘しておく。

5. 研究の発端と背景

以上で、いわゆるインターネット調査の現実がいかなるものかの概要は示せたと思うので、次に、こうした研究を始めた端緒、そしてなぜそのようなことが必要と考えるに至ったかを、述べてみたい。このことは、データ科学の精神に沿って(少なくとも筆者の考える方向として)、進めてきた実証研究であり、またその経緯の概略を述べるのがデータ科学の考え方を立証する一つの方法と考えたからである。

実は、筆者が当初考えていたことは、電子調査やインターネット調査の研究が主ではなく、いわゆる自由回答の取得法や、そこで得たテキスト型データ(textual data)の解析方法の研究であった。これに関する研究は、フランスの知己であり友人である Ludovic Lebart 氏他と、ここ数十年近く行ってきたテキスト型データ解析の方法論の研究に関連している。こうした研究について、10年程前にはほとんど関心を示す者はなかったが、ここ4,5年前から自由回答設問の解析に関して頻繁に問い合わせがくるようになった。インターネット調査を行うと電子的手段を通じて自由回答がたくさん取れる、回答者のナマの声が聞ける、またテキスト・マイニング・ツールを用いることで素晴らしい結果が得られるというのである。実際、ここ数年の間に、テキスト・マイニングを標榜する多数のソフトウェアが登場している(大隅^{[23],[24],[26]}参照)。しかし、Web調査で取得したというデータを貰って分析を行ってみると、どうも芳しい結果が得られない、データ取得方法に問題があるのではないかと、との疑問を抱くようになった。とくに、インターネット調査を使って、廉価に、クオリティが高い自由回答データが豊富に得られるというような風説に納得できない、本当にそうなのかということから、では、インターネット調査とはそもそも何であるのか、それを知るためにも実際に実験調査をやろうということになった。

目的はまず、自由回答の取得方法の実験にあったとはいえ、まずはインターネット調査がどう機能するかということ、またそのための実験調査をどうデザインするか、ということが要点となる。このとき、実際の調査現場と同じ状況下で、ある一定期間調査を継続して行い、また複数回行うことを考えた。また、ほとんど同時的に、異なる複数のサイトで、調査票も設問構成も同じとし、また従来型調査で実績のある質問を採用し、従来型の調査と比べるという方式を採用した。同

時に他の従来型調査法に則った調査(オムニバス調査, オンライン調査)を併用することで, 調査方法間の比較も可能となるように設計した。この他, 追跡調査(複数の調査回に参加した回答者を, 確かにインターネット調査に参加したかどうかの確証を郵送と電話で行う), トラッキング調査(回答者が調査票への回答を開始してから終了するまでどう回答するか時間軸に沿って電子的に追跡する方法), 調査票のページネーションを変えとき, 回答行動がどう変わるか, 等々, 様々の要素を考慮した, 一連の実験調査を総合的に進めた。

調査は, まず1997年に第1次調査(プリテスト)として, このときにはRON (Recruit Research on the Net)サイトの協力を得て, 併せて12回連続実施した。その後, 第2次調査, 第3次調査を計画し, 最終的には5社の参加協力を得て, 国内初めての複数の調査サイトにおける, 同時的, 継続的, かつ調査内容を同じとした実験調査(Web調査および従来型調査)を行うことができた。調査の詳細のすべてをここに述べることはできないので, その概略を示すが, このうち, 第1次, 第2次調査については既に報告書があり, また第3次調査については, その概略が報告されている^{[20], [21]}。

5.1 実験調査の概略

1) 第1次調査

- リクルートリサーチ社, RONにおける実験調査
- 1997年3月24日~4月8日に応募の登録者を対象
- 計画サンプルは, 登録パネル全体とした(調査期間中サンプルを固定)
- 1997年5月~11月にかけて連続的に12回実施(ほぼ同一対象者)

2) 第2次調査

- 国内のWeb調査形態の分類・類型をサーベイで確認
- 実験調査の協力参加呼びかけ
- Web調査と通常調査の併用を計画(通常調査は, オムニバス, オンラインを利用)
- 実施期間は, 1999年1月末~4月初旬
- Web調査(3サイト×4回=12回; 第1回と第4回は同一登録者を利用)
- オムニバス調査(1サイト, 3回)
- オンライン調査(1サイト, 2回)

3) 第3次調査

- さらに, 回答者行動を詳細に検討するために, 類似調査を継続的に実施
実施期間は2000年4月~6月
- 第一次調査(電通リサーチ分)における第1回, 第4回の両調査参加者の「追跡調査」(郵送調査, eメール調査を併用)
- 新たなサイトの参加を得て, Web調査実施(2サイト×2回=4回)
- 1社では「トラッキング調査」を組み入れた
- オムニバス調査(1サイト, 2回)

(注1) 実験調査に参加いただいた企業は以下の通りである(五十音順)。

(株) エー・アイ・ピー(AIP: Asia Internet Plaza) および博報堂, NTTナビスペース(株), (株)電通リサーチ, リクルートリサーチ(株), (株)マーケティング・サービス

(注2) この一連の実験調査の後, 2002年には第4次実験調査を計画・実施した。この原稿の最終稿の執筆中に, 分析結果が得られたので, その一部を後ろに補遺として加えた。

まず, 第1次調査は一つのサイトで1997年に集中的に行った。これはいわゆるパネル型で, 事前に登録確認(登録と調査への協力の合意)を取り, その登録者に対して一定期間, 集中的に行った。第2次調査からは少し調査設計を変えて, リソース型サンプリング方式とする等の工夫を行った。同時にオムニバス調査, オンライン調査を行っている。なお, ここでいうオンライン調査とは, 通常の調査のように住民基本台帳からサンプリングし世帯を確定し, その世帯の承諾を取ったうえで, 回答に用いるある特別な機器を設置し, 電話回線を使って回答を集めるという方式である。第3次調査では, 第2次と同じようなWeb調査に併せて, 追跡調査とトラッキング(回答者行動の電子的追跡)を行っている。

インターネット調査の一つの特徴は, 調査票のデザインの豊富さにあると言われている。実際に, 回答選択肢の作り方にしても, ラジオボタン, チェックボックス, コンボボックス, プルダウンメニュー等多様な方式が使える。また, イメージ(静止画, 動画), 音声などマルチメディア的な機能を用いることもできる。こうした多様性に関心があり, 当初の調査票の設計では, こうした多彩な機能の利用を試みたのであるが, 結果をとして, こうした諸要素が回答結果に及ぼ

ず影響がよく見えないこと、とくに回答率（換言すると無回答率）にどう響くかの定量的な評価が困難ということが見えてきて、第2次調査からは、調査票を単純にして回答しやすくした。つまり、通常の質問紙に近い形とし、しかし、インターネットでなければできないいろいろな手当や工夫は行った。

たとえば、ラジオボタン、チェックボックス、コンボボックス等を、設問の種類、1項選択か多項選択かなどで使い分けること、自由回答であれば、テキスト・エリアを用意して、その中に書いてもらう等である。ここは普通の紙による調査票と違うといえば違う。また、シングル・アンサーの場合、マトリックス形式の場合と、それぞれに応じた形式を用意し、その中に置いたボタンをクリックすると、裏側で電子的に CGI (common gateway interface) を使って回答が自動的に取得できるという仕組みである。こうした処理が可能であることが電子的手段を用いるメリットであるが、同時に取りこぼしといったトラブルも生じる（インターネット調査特有の調査不能の一例である）。調査票の配色もきわめて単純にして、かぎりなく紙の調査票に近づける努力を払った。

実は、こうした調査票の設計が回答行動、回答率等にどう影響するかといった研究は、欧米とくに米国では非常に進んでいる。マニュアルや多くの事例報告があつて、たとえばGVU調査を閲覧するとどのような設問フォームでどのような結果となったといった例が多数示されている。また、筆者らも、2002年から、欧米の研究者の協同研究の呼びかけに応じて、調査票のデザイン、設問の設計レイアウトなどが回答に及ぼす影響の評価実験を進めている（米国、ドイツ、スロベニア他）。

次に、具体的な設問構成とその設計方針であるが、ここでの要点は二つある。一つは従来型の調査、たとえば国民生活意識調査や社会意識調査（内閣府）、日本人の国民性調査（統計数理研究所）など、従来、定期的あるいは長期的に使われていて、ある程度回答の傾向がわかっているような設問を用いる場合、第二は、インターネットに関係する環境や事項を聞く独自に作成の設問と、大きく二つに分けて繰り返し行っている。

5.2 調査結果にみられる特徴の概要

以上が、筆者等が行ってきたインターネット調査の概要であるが、では実査から一体どのようなことがみえてきたかを、数値例をもって示すことにしよう。

まず、実験調査から見えてきたことから、結論を先に示せば、一般に通説としていわれていること、類推されてきたことはほぼ検証されるということである。その一つが、回答者と登録者集団（リソース）の関係である。つまり、「回答者はリソースを代表していない」ということがある。どう代表していないかを図で見る。多数の結果があるのだが、ここでは2つを示そう。図2-1は、1999年の第2次調査第3回実施分について、リソースをある時点で止めて、そこからサンプリングして得た計画サンプルについてみた時の、性別、年齢別、地域別の属性の構成である。（注：後の図3にみるようにリソースと計画サンプルは、当然良く一致するので、計画サンプルを見ることはそのままリソースの特徴を見ることとなる）。性別では圧倒的に男性が多い、年齢が若い層に多い、また地域が特定の場所に片寄って集中するといった、一般に言われてきたインターネット調査特有の傾向がある。地理的距離のないのがインターネットの特徴だといわれているが、それにもかかわらず、一部地域に集中する傾向があり、すでに地域代表性が保持されていない（また、東京23区内について、区別にリソースをブレイクダウンすると、都区市内でも地域差があることも分かっている）。

次に、2000年調査になると（図2-2）、多くの他のインターネット調査でそういう傾向が指摘されているように、確かに女性の比率が高くなり、また男女ともに年齢層が上に広がってくる。しかし、地域差はそのままである。この例は、リソース自体の分布と一部は計画サンプルの分布を示している。ここでは、第3次調査における2社のサイトについて、一つ（電通リサーチ社分）はリソースと、そこからサンプリングした計画サンプルを比較し、他の一つ（AIP社分）は、サンプリング操作を行っていないので、リソースそのものについて、それぞれの属性についてグラフとしてある。

実は、ある例であるが、比較検証のために、インターネット調査を専門とする複数のサイト（5社）でほぼ同時に、いわゆるリソース打ち切り型で行った調査によると（いずれも300名と固定して打ち切り）、男女比他の属性の構成比のバラつきがまちまちという結果を得ており、言い換えれば、とても調査とはいえないような結果を見ている。しかし、いわゆるインターネット調査の現場では、多くの場合はこうした妙な方式が採用され、その結果がまかり通っている。リソース打ち切り型のサイトには、こうしたある種の偏った傾向が現れることはわかっているが、商用上は、廉価に行

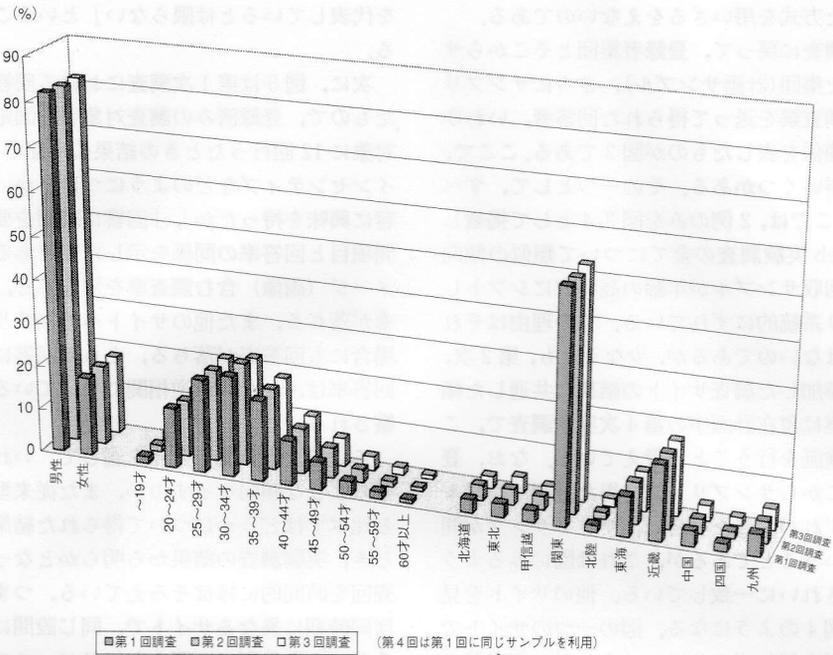


図2-1. 計画サンプルの属性構成 (1999年, 電通リサーチの場合)

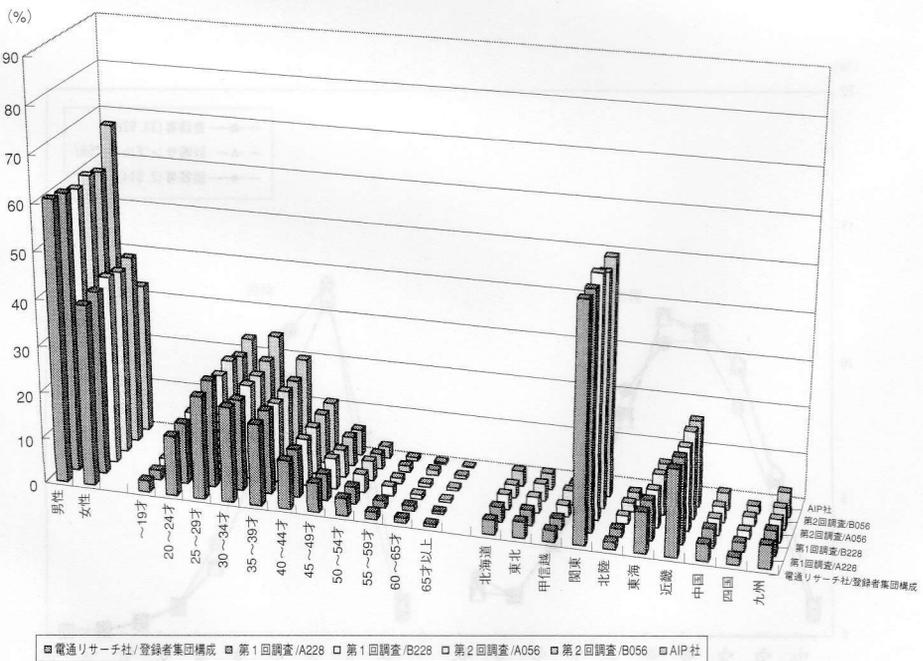


図2-2. リソース, 計画サンプルの属性構成 (2000年, 電通リサーチ, AIPの場合)

うにはこうした方式を用いざるをえないのである。

われわれの調査に戻って、登録者集団とそこからサンプリングした集団(計画サンプル)、さらにサンプリングした人に調査票を送って得られた回答者、いわゆる回答状況の関係を表したものが図3である。ここで、特徴的な傾向がいくつかある。その一つとして、すべての場合に(ここでは、2例のみを図3,4として掲載したが一連のWeb実験調査の全てについて類似の傾向が見られる)、回収サンプルが年齢の高い方にシフトしている。つまり系統的にずれている。この理由はそれほど明らかではないのであるが、少なくとも、第2次、第3次調査に参加した調査サイトの結果に共通した傾向ではある。実は現在計画中の第4次実験調査で、これのさらなる検証を行うことを考えている。なお、登録者集団とそこからサンプリングで得た計画サンプルの属性情報にずれが生じた場合は、サンプリングが間違っているということであるが、これは図にみるように当然ながらきれいに一致している。他のサイトを見ても同様で、図4のようになる。図の一つのサイトではサンプリングを行わずにリソースの全員に調査票を送ったが、ここでも回収した結果がずれている。こうして、今までの実験の範囲内では、「回答者がリソース

を代表しているとは限らない」ということが見えてくる。

次に、図5は第1次調査における回答率の推移を見たもので、登録済みの調査対象者を固定したパネルを対象に12回行ったときの結果の一部である。図には、インセンティブをどのようにつけたか、また「調査内容に興味を持ったか」、「回答に時間を要したか」の設問項目と回答率の関係を示した図である。たとえば、イメージ(画像)含む調査票を用いると、とたんに回答率が落ちる、また他のサイトへの接続リンクを設けた場合にも回答率が落ちる、さらに回答に要する時間と回答率は、きれいに逆相関になっていることなどが観察される。

ここで、実験調査全体を通じて、いわゆる回答率がどのような傾向を示すのか、また従来型調査の回収率と比べてはどうかについて得られた結果を見る。このとき、実験調査の結果から明らかとなったことは、調査回を時間的にほぼそろえている、つまり回答者はほぼ同時期に異なるサイトで、同じ設問に対して回答するという条件下で、明らかにリソースやパネルの鮮度が落ちるにつれ、回答率も減少する傾向にあることが分かる。その顕著な例が実験調査全体の回答率の結果

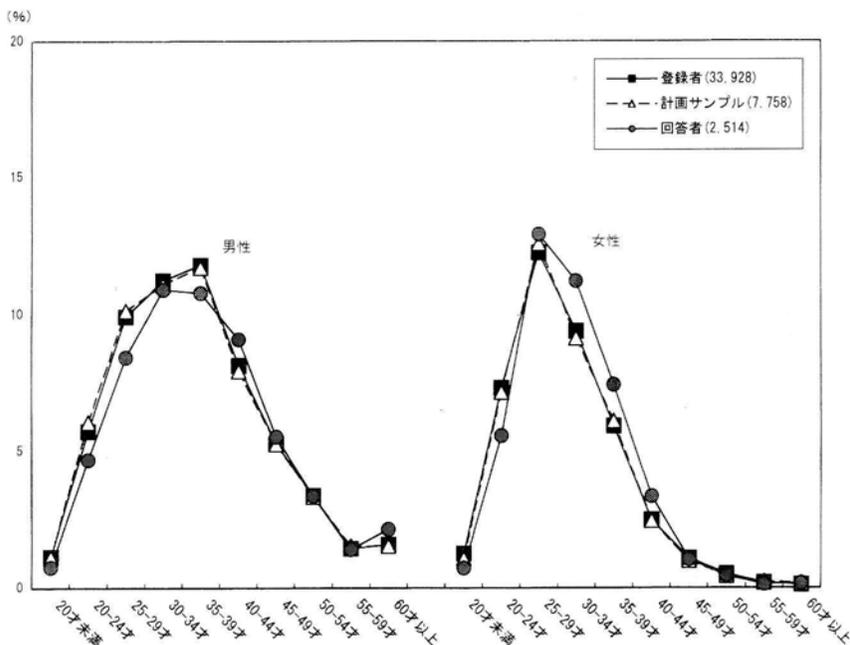


図3. リソース、計画サンプル、回収サンプルの性別・年齢区分別の構成(2000年、電通リサーチ)

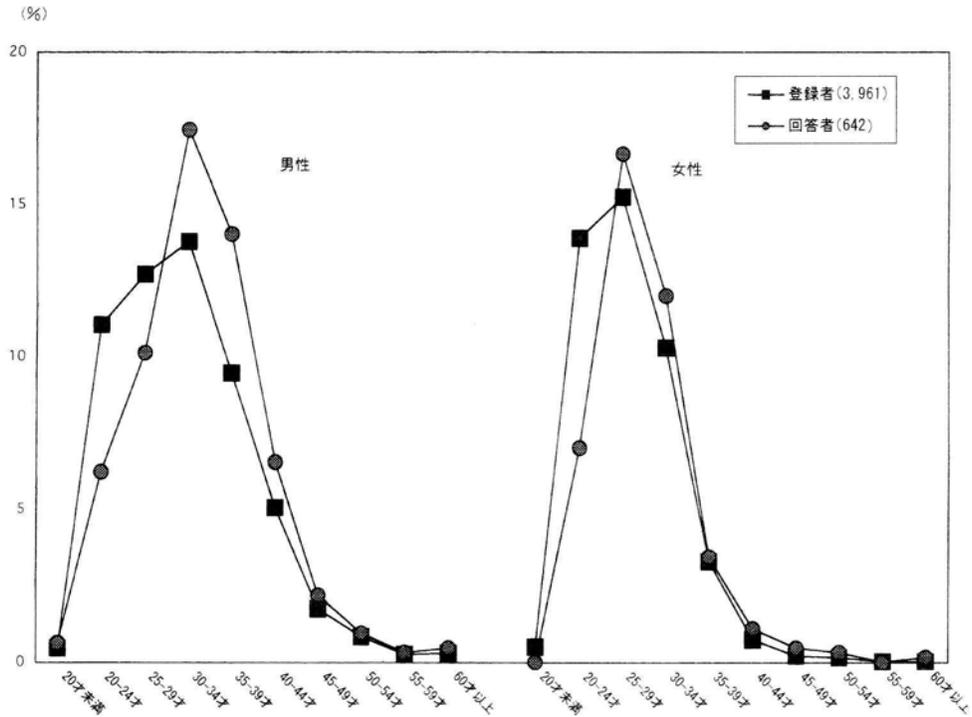


図4. リソースと回収サンプルの性別・年齢区分別の構成 (1999年, リクルートリサーチの場合)

を比べてみたときで、図6のようになる。ここで1999年の第3回目、電通リサーチのWeb調査で回答率が高くなっているが、これはこの調査直前にリソースを更新しているからである。また、第2次調査の場合と第3次調査では総じて回答率が低い。とくに第3次調査のWeb調査のAIP分については、そのリソースは構築してからかなり時間が経過しており、陳腐化がかなり影響している。他の場合も、そのリソースの作成時を考慮して検討すると、回答率には、リソースの鮮度が大きく関係するということが分かってくる。また、従来型調査の回収率と比べて、Web調査では総じて回答率がかなり低いことも見えてくる。

4.3 調査結果からみた回答者像

つぎに、従来型調査で用いられてきた設問で、その回答傾向が知られた質問について、インターネット調査と他の調査(オムニバス, オンライン)による結果を比較してみた。実はここでは、単純集計をみるだけで、インターネット調査と他の調査とで類似や差異が顕著に観察される。

たとえば、インターネット調査の場合、「現在の生活満足度」については、満足が比較的多目に出る。日本人の国民性調査の「人は信頼できるか」の設問では、「信頼する」がすこし下る。良く知られた「2人の課長」の設問については、「仕事以外の面倒はみない」方の課長の支持がかなり多くなる。また、これも良く知られた設問である「人のくらし方」では、「趣味にあうくらし方」(金や名誉は考えずに、自分の趣味にあったくらし方をする)という意見が、オムニバス調査や国民性調査に比して圧倒的に多くなり、一方、「その日その日をのんきに」(その日その日をのんきに、クヨクヨしないでくらす)が、かなり少なくなる。これらの関係を総括的に図示したのが図7-1, 7-2である。図7-1は、比較のために、国民性(1998年)を横軸に、国民性(1993年, 1988年), オムニバス, オンラインの各調査を縦軸とした等比グラフである。これに対して、図7-2は国民性調査・オムニバス調査とインターネット調査すべてを比較したものである。

ここで図中の対角線上に打点があれば、横軸、縦軸においた調査回間に類似性があることになる。図7-1

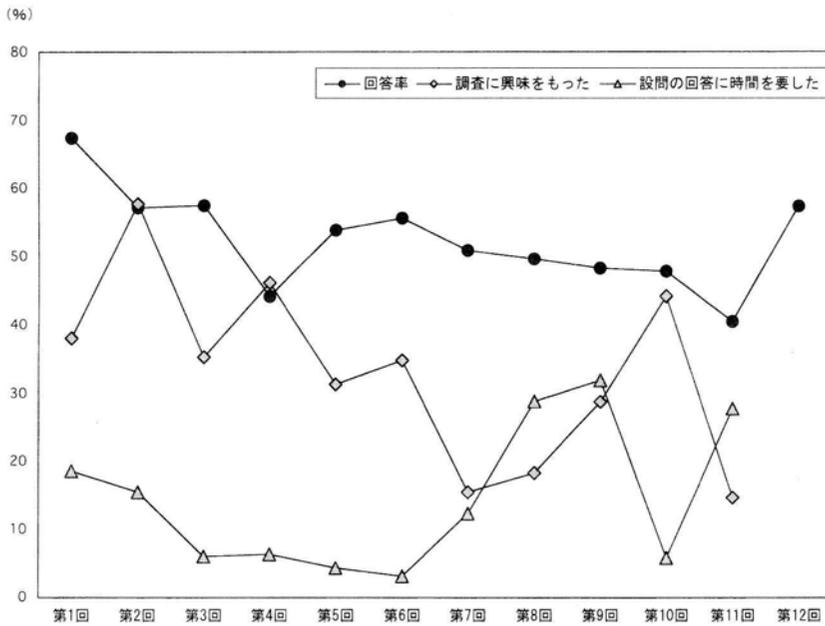


図5. 第一次調査における回答率他の推移 (1997年, リクルートリサーチ)

(注) 第1回 (抽選でハワイ旅行), 第4回は回答者資格を条件選択, 第11回は他社HPへのリンク設定, 第12回 (全員に謝礼)

は説明するまでもなく、よく似ている。一方、図7-2は、そうは見えない。つまり、国民性調査とインターネット調査との結果には類似性がない。またどの項目がどういう出方となったかでインターネット調査の回答者の特徴が見られる。しかしながら、インターネット調査の結果が、サイト間の差異が非常に少なくまとまっていることがわかる(つまり、回答者は異なるのに、インターネット調査ということではきわめて類似していることが分かる)。

また、世論調査の常套設問である、内閣支持率などについても、従来型調査や新聞社等の報道機関が発表してきた結果と非常に異なる結果となる。現在の小泉内閣については、第4次調査に組み入れることとなっているので、いずれ結果を開示出来るのだが、実査当時の小淵内閣と森内閣についての支持は、インターネット調査では、調査の当初から森内閣の支持率が10%を切るというような数字が出てくる。また、政治に対する考え方、保守か革新かを10段階で尋ねた設問では、やや予想に反して、革新側に偏るとはいえず、むしろ中庸から保守側が多いといった傾向を示す。ここらは指摘するだけに留めるが、要は、インターネット

調査における回答者行動には、一般の従来型調査法で得た回答者とは異なるものがある、ある種の回答者像が垣間見えてくることである。日本人のキャラクターをどう考えるか、インターネット・ユーザーのキャラクターや回答者像をどう考えるかは大変にむずかしい問題であり、さらなる検証を要することでもある。ここではいくつかの例に見るような違いが出たということだけを指摘しておきたい。

5.4 インターネット関連の設問への回答傾向

次に、インターネット関連設問にみられる傾向について若干記述したい。インターネット関連の調査は、実は無数に行われているのだが、いずれもが断片的であり、しかも単発的な調査が多く、筆者等のような、同時的に複数のサイトで継続的に実施したケースというのは例がない。その意味では、ここに示す幾つかの例示は、インターネット調査の回答者の類似や差異を知る例となるのであるが、紙面の都合で、調査結果で見られた特徴の記述に留める。なお、その一部は吉村による報告に記されている⁽¹⁸⁾。

設問内容は、インターネットに関する情報入手源、イ

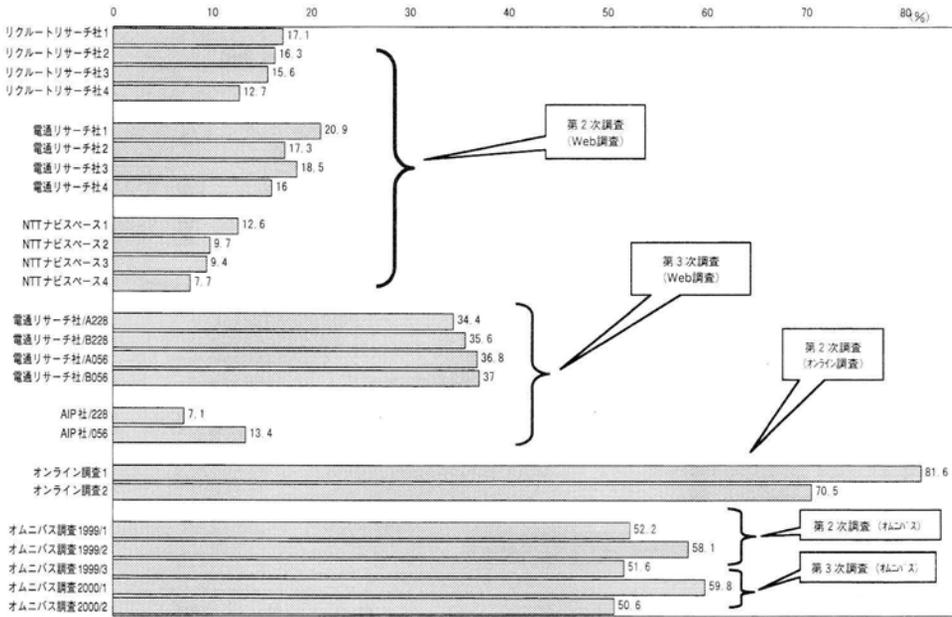


図6. 各調査次・調査回における回答率、回収率の一覧

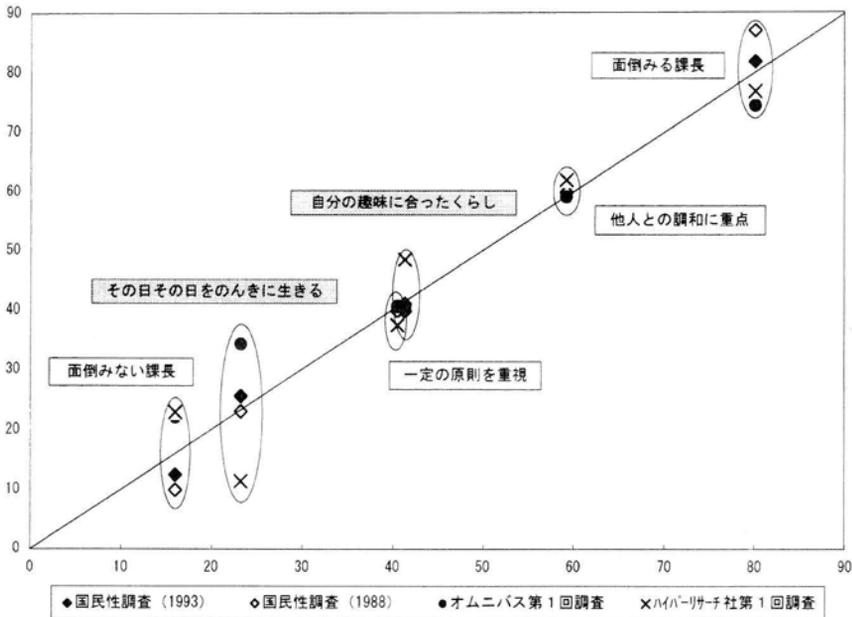


図7-1. 国民性 (1998年) 対 国民性 (1993年, 1988年), オムニバス, オンラインの関係
 (注) 横軸に日本人の国民性調査の回答比率(%), 縦軸にはオムニバス調査, オンライン調査における回答比率(%)を比較した。

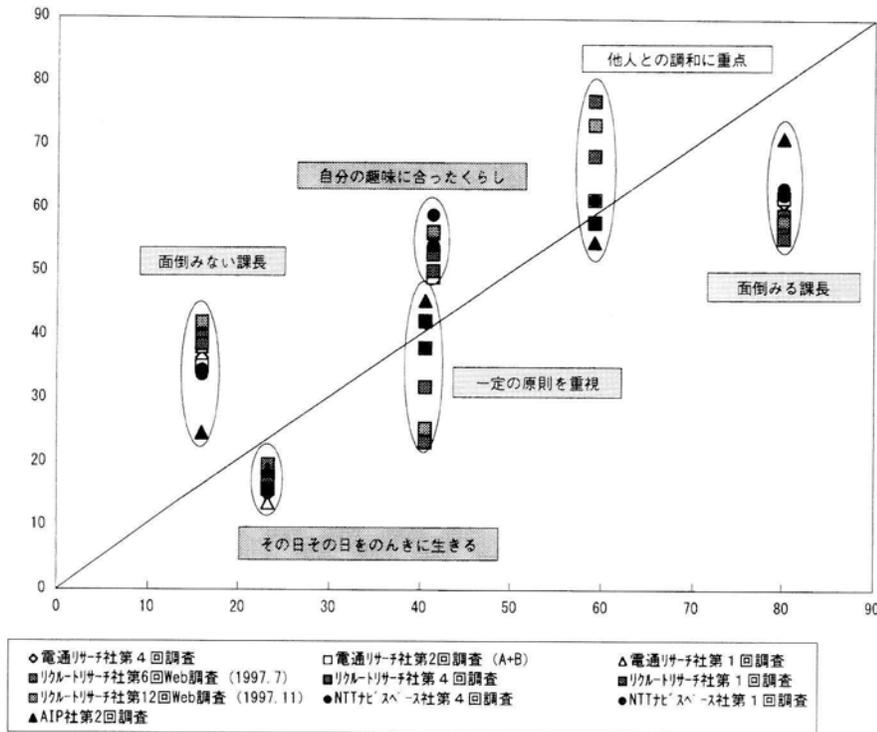


図7-2. 日本人の国民性(1998年)と各サイトWeb調査との関係
 (注) 横軸に日本人の国民性調査の回答比率(%), 縦軸にはWeb調査における回答比率(%)を比較した。

インターネット歴, インターネットの利用場所と利用時間帯, 電子メールをやり取りする人数, 情報機器類の装備, 生活に占めるインターネットの位置付け, インターネットを使った行動(商品注文, 資料請求など), インターネットを通じて相手に教えることに抵抗がある情報, 提供してよいと思う情報, プライバシー関連の意見, 法規制への意見, インターネット調査に対する印象(参加の意思, 協力, その理由等), なりすましへの意見, 本音で答えやすいかへの意見, アンケートモニターサイトへの登録状況, ネットワーク社会と現実の社会の差異の認識, メディア接触, 等々である。

たとえば簡単なことでは, インターネット歴については, 3~4年のところが多いが, これは性別と調査サイトによって差異がある。また, リソース型と打ち切り型とでも違いがみられる。年齢, サイト間でバラツキがみられる例は, 交信相手数である。大まかに言うと, 携帯電話の場合と傾向が異なるようで, 少なくとも, 一連の実験調査の範囲内ではあるが, 年齢が高く

なるにしたがって交信人数が多くなるように感じられる。

所有する電子メールアドレス数については, 各サイトの傾向がきわめて類似し, 1~2個が多いが, 男性と女性で大きな違いが出るのが特徴である(女性の方が少ない)。さらにインターネットに関する情報の入手経路(TV, ラジオ, 新聞, PC専門誌, インターネット専門誌, 電子メール, ホームページ, 一般書籍等々)については, サイト間の差が出るのではないかと予想したが, ほとんどサイトによる差はない。また, すこし設問文を工夫した生活に占めるインターネットの位置付け(「ネットワークは自分らしきを出せる」「趣味の一つに過ぎない」「ネットワークがないと淋しい, つまらない, 物足りない」など)についても予想に反しサイト間の差はみられない。

少し立ち入って, 調査への協力条件についても尋ねたが, インターネット・ユーザーは一般の場合と少し違い, 意見がはっきりしており「調査の主題・目的が

はっきりしていること、興味あること」「回答の時間の余裕があること」「調査の趣旨・目的が理解できること」などへの意向が多く、また「結果の公表開示」を求める意見も少なくない、となっている。

情報機器類の装備については、携帯電話、携帯情報端末(PDA)、コンピュータ(デスクトップ型PC、ノート型PC)などの利用が多いことが当然予想されるが、筆者の関心は、従来型調査と比べてどういう出方をするか、どれに重きを置いているかにある。これについては、携帯電話よりはPHSや情報端末であり、ハイビジョンよりはデジタルビデオというふう一般の調査の回答者とは若干違う特徴が出る。

プライバシーや法規制に関する質問も行っている。これについては、前述のGVUの調査に、プライバシーに関する質問があるのだが、これをそのままでは使えないので、すこし手を加え、独自に20項目の質問を用意した。ここには、「わいせつ情報が公序良俗に反する場合でも、表現の自由は保障されるべき」、「インターネット上の誇大広告には法的規制が必要」などがあるが、これらについても、サイト間の差違が出ると当初は予想したが、男女間の差は幾つかの設問で差が出るものの、サイト間ではきわめて類似した傾向を示している。念のため数量化法III類(対応分析)で分析し比較確認してみても、やはりサイトによる差はなく、類似の結果を得た。

これは確証があるわけではないが、サイト間に差がないことの理由の一つとして、重複登録が関係しているように思われる。現に「このアンケートには他のサイトで答えた」という例が現れるが、こうした複数のサイトに登録している人が若干ではあるが存在する。現在でもアンケート・サイトはどんどんふえており、次の第4次調査の候補リストには既に100以上が列記されている。これらの複数のサイトにどう登録しているか、どの位の重なりがあるかをみるのが1つのポイントだが、この点については、第4次調査で検証を進める。

5.5 定性情報の利用 — 自由回答の分析 —

終わりに、この実験調査の研究のテーマの一つである、自由回答取得法とその取得結果の分析の方法論の研究について少しふれておきたい。設問の中には、多数の自由回答設問があるが、ここでは、日本人の国民性調査でも使われている「あなたにとって、大切だと思うものはなにか」への自由回答データを例としたい。

周知のように、最近の国民性調査では、その1位が「家族」である。しかし、2位以下の回答頻度とそのパターンの出方についても、どう現れるかをみてみた。ちなみに4回分の調査結果について、まず、単語の出現頻度から、「家族」の他には「自分自身」「仕事」「人間関係」といった語句が比較的多くあがっている。またここで注意すべきは、これらと並んで「環境」などがあがってくる。大切なことを自分の立場としてだけではなく、世の中全体の事としてとらえるグループもあることが分かる。さらに、4回の調査とも結果が非常に似ているということが分かっている。自由回答という定性情報であるにも関わらず、ここでも、ある種の類似性が検証されること、それは継続的に同一の設問を用いて、複数の調査サイトで継続して行くと、このように分ってくるのが重要であるという例としたい(詳細は^[23]、^[26]を参照)。

また、インターネット調査の場合、自由回答設問への書き込み率は高いと言われてきたが、確かにいずれの調査結果でもこれが高い傾向にある。また、通常を選択肢型設問に設けた「その他;具体的に」欄への記入率も多くの場合高く、70%以上になることも珍しくはない。インターネット・ユーザーはたくさん書き込む傾向にある、確かに郵送調査やオムニバス調査にくらべ、記入量が多いように見えることがあるのだが、しかし、通説として言われているほど、書き込みの内容が豊かだというようには思えない。言い方をかえれば、内容豊かな回答を得るためには、やはり設問の仕方、取得方法をきちんと考えないといけないということであり、ここでもやはり調査設計のあり方が問題とされる。また、さらなる調査による検証が必要でもある。

6. ま と め

まとめると、インターネット調査の実験調査から得られた情報の断片を書き連ねてきた。ここでとくに重要と思われる事項を要約すると以下ようになる。

(1) インターネット・ユーザー数を正確に捕捉する手段がない現状では、インターネット・ユーザーが日本人の全体を代表しているとはいいがたい。

(2) 次に、少なくとも今回の一連の実験調査の結果からは、登録者集団(リソース)は、インターネット・ユーザーを代表しているとは思われない。

(3) そして、回答者は登録者集団(リソース)を代表しているとは限らない。

回答者は、勿論インターネット・ユーザーの一部であるが、その元となる集団であるインターネット・ユーザーの全体を代表はしていないし、しかもリソースも代表してはいない、ということが見えてくる。また、従来型の設問についての比較検証から、回答に類似性のある場合と、そうではない場合が当然ある。しかし、少なくとも従来型調査と Web 調査の結果の出方にはある種の差異があり、しかも Web 調査の結果には調査サイトに関わりなく、多くの場合に高い類似性が見られる。

さらに、今後の課題として、これは調査法を考えたうえできわめて常識的なことだが、以下のようなことが検証されるべきである。

(1) インターネット利用環境のハード面、技術的要素だけで解決されないことが無数にある。しかも、これが従来とは異なる新たな調査変動や非標本誤差等を生む要因となっていることがある(たとえば、接続回線の中断による回答不能、調査票の設計不備による回答欠測、回答者の利用 PC の環境要素など)。この他、登録者情報の守秘や保護に関わるセキュリティ技術の諸要素から起こる問題や制約もある。

(2) しかも、インターネット調査はマシン・トゥ・マシンであって相手の顔がみえないことに問題がある。この点では、CATI などに類似する点もあるが、直接的に人を介しないという点でこれらと大いに異なることがある。

(3) さらに大きな問題として、リソースやパネルをどのように構築し、それをいかに保守維持するかの方法論が必要であるということ、とくに、リソースの鮮度をどう保持するかが一つの鍵となること。

(4) そして、従来の科学的調査法の利用で得たノウハウや知識を、インターネット調査にどう適用できて、あるいは、どこに適用が難しいか、あるいは新たな方法論の構築が、どう求められるかを体系的に研究する必要があるということ。

この他、多くの未解決の問題があることは、これまでの論述で明らかであろう。これらを解決するためには、データ科学のコンセプトの下に、調査の現場とのコラボレーションを通じた、息の長い継続的調査を行うことが重要である。また、繰り返しになるが、従来型の調査の実施環境が益々悪化する中で、新しい調査法としてのインターネット調査の役割や適用可能性を科学的に検証することの重要性に、少なくとも調査に関心のある研究者は目を向けるべきである。

7. さらなる検討事項

インターネット調査について、さらに、従来の調査法やそれに関連させて、以下に挙げるような事項についての集中的な研究も必要と考える。

(1) 調査不能とその関連事項

インターネット環境を利用するうえで、その固有の特性から、新たな意味での「調査不能」、ないしはそれに類した事象が生じることがあり、実際にそのような現象が多々見られることがある。従来型調査でも見られた諸事象(たとえば、代理回答)もあるが、インターネットの特性上、今までは見られなかった事象が多々見られる。たとえば、受け取った調査票をディスプレイ上で視認はしたが回答行動は起こさない(接触数(率)と名付けた)、回答者の回答意思はあってもハードウェア上の問題から回答送信やサーバ上での取得に失敗するなどである。この他、「なりすまし」や匿名性の問題、多数のハンドルネームやアドレスを所有して、相手に応じて使い分けること(筆者等はこれを「多名性」と名付けた)、従来とは異なるインセンティブとその方式(ポイント制が多い)に関わる問題等々、いろいろな事象が見られる。しかも、こうした事項の研究は、現状のインターネットの仕組みの中では、多くの場合はその実態を把握することも難しく、インターネット調査システムの設計自体のありようにも関わってくる。

このように、「インターネット調査における調査不能」(non-responses)をどう考えるかは重要であり、慎重に検証することが、今後の重要な研究課題の一つであると考えている(現実には、研究が非常に遅れている)。

(2) インターネット利用環境の急速な変化

ここ1, 2年間の間に、社会・経済共に様々のアクシデントがあり、インターネット・バブル崩壊や IT 革命への疑念の噴出等、あれこれと議論されるようになってきた。また、携帯電話とそのインターネット利用の急速な普及で、電話調査やインターネット調査の仕組みそのものも、見直しが余儀なくされている。とくに、ハッカーなどのスキルの高い者達が引き起こすトラブルのように、一般の人にとって非日常的と思われた問題だけでなく、最近はごく身近なところで、様々な不具合や問題が表面化している。インターネット上のプライバシーの問題、インターネットを用いた新たな形の犯罪、中傷誹謗事件等、枚挙にいとまがなく、様々な負の事象が次々と現れている。

これらを、インターネット調査との関連で考えると、今後は次第に、自由度の高い調査環境を得ることが困難となることが予想される。実際、様々な法規、たとえば、プロバイダー法、迷惑メール商法、あるいは国際間のサイバー犯罪条約、等が次々と登場している。また、情報公開法、個人情報保護法等も、従来型調査の実施環境に大きな影響を与えることは明らかで、翻ってインターネット調査のありようにも無関係ではない。こうした側面から、インターネット調査をどう考えるかの検討も必要とされる。面接調査と同様に、なるべく回答者の顔が見えるような調査が望ましいのだが、インターネットの特性上、これは困難なことでもある。しかし、少なくとも、リソースを作成する段階で、回答者となってくれる人（その意思のある人）の捕捉・登録方法、調査協力への合意形成の方法、インセンティブへの配慮、調査結果の開示方法、等々を丁寧に検討し、一体、インターネット調査とはいかなる特徴をもつ調査法か、どう利用できそうかを、さらに検証することが求められる。同時に、調査をどう考え、また信頼あるものとするためにはどうすべきかといった、インターネットを調査に利用する側の対応や試みも行われるようになってきた⁽¹⁹⁾。

(3) 回答行動と回答者像

今までに行ってきた調査の範囲で、との制約の中で、インターネット調査における回答者には、ある種、特有の回答行動や考え方があるように見える。多くの場合、インターネット調査は「自己参加型(self-selection)」調査であるがゆえに、回答の代表性、一般性が問われると同時に、得られる回答は「回答したかった人の意見」となっているということである。現時点では、インターネット調査における回答者は、言いたいことを、言いたいときに述べている、責任のない発言が多いようにも見える、等の印象がある。

また、メル友関係、インターネット掲示板上での誹謗中傷合戦、(限定された範囲とは思われるが)、メールマガジン(メルマガ)のかなり特異な利用のされ方やその掲示内容、インターネット上でのコミュニケーションや交際(顔を合わせたこともないのに親しくなる)、等々、あれこれ見られる新たな社会現象との関係なども、調査の環境構築を考えるうえで無視できない要素である。

こうした「回答者像」、さらには社会現象として、インターネット上の仮想空間と現実の世界との関係を、一体インターネット調査への回答者はどうとらえてい

るのか、といった「回答行動」の把握に有効な調査方法のあり方を、現在、実施計画を立案中の第4次実証調査でさらに検証を続けることを計画している。

8. 結びにかえて

この特集は、故水野欽司先生を追悼する意味もあると伺った。また、遅筆である筆者が、なんとか執筆をお引き受けした最大の理由はそこにある。

水野先生がご存命であったら、インターネット調査などと聞いただけで仰天し、お叱りを受けることであろう。実際、ある学会で電子メール調査の報告がなされた際に、水野先生が激しく糾弾されたということの間接的にはあるが伺った記憶がある。水野先生の調査に対する主張やお考えからすれば、言わずもがなのことである。

しかし一方では、世の中の事情も大きく様変わりし、かつて可能であった調査方法や調査方式も、理想的な状況下で行うことがきわめて困難となってしまった。確かに、理屈ではわかっている、回収率が50%を下回るような調査結果まで現れるようでは、果たして信頼に足る調査情報が得られるのかと不安になる。しかし、現実には確実にこのような状況に移りつつあると体感している。また、多少楽観的にみても、いずれそうなることは明白である。

しかしながら、こと調査法の研究に関して、関連研究者の多くは、新しいことへの挑戦を怠ってきたと思われてもしかたがない。また、新たなことへの試みに対する協調的な支援やコラボレーションもあまり盛んとは言えなかった。実際、インターネット調査に限らず、研究対象とすべき様々な調査法の研究は、決して盛んとはいえない。ときとして、調査現場の実務家から、もはや統計学の研究からは学ぶ事はない、とまで言われる所以であろう。

こうしたことを考えるたびに、故水野欽司先生が折りにふれて述べられてきた言葉の数々が思い浮かぶ。このいくつかを手元のメモから拾ってみよう。

- (1) 苦労して取得したデータほど可愛いものだ
- (2) データ取得の苦労なくして解析の喜びはない
- (3) 調査で1票でも増えて回収率が上がったときの実感、感動、感激!
- (4) 人の取ったデータで解析をすることほど愚行はなし
- (5) いつでも批評はできる、何を行ったかが重要だ

(6) 調査が終われば、仕事の8割が終わり

(7) 調査は死んでも同然

筆者が少し意味を取り違えて書きとめた語句もあるかもしれない。また、こうした言葉をどう捉えるかは、人によってさまざまである。しかし筆者には、このどれもが、我々の置かれている今の閉塞的な状況への鋭い指摘であり、また自省を促す言葉に取れるのである。

ともかくにも、これらは、故林知己夫先生とともに提唱してきた、データ科学の精神に通じる事が多いのである。統計学、統計的データ解析、多次元データ解析等々、呼び名はどうかであっても、いずれもが実践の科学、データ指向・重視の方法論であることを標榜していながら、その実態はあまりに現実と乖離した中であつたとしか言いようがない。いまや情報学なる分野の中に埋没しかねない状況にあるが、こうした分野と大きく異なる特質が、現象解明に必要なデータ取得をいかに行うか、またそれに必要な実験計画のあり方、とまさにデータ科学の精神で事に望むべきことこそが、オリジナルな世界を切り拓く術と考える。ここに述べたインターネット調査における実験調査も、そうしたことのさやかな試みの一つと考えていただければ幸甚である。「知は力なり」との言葉があるが、知を知恵とするための術は何かを考えねばならない。

実は、この原稿を書き終える直前に、林知己夫先生が逝去されるという、まことに無念としかいいようのない知らせを受けた。ここに述べたことの多くは、繰り返しとなるが、先生とともに主張してきたデータ科学の精神を、なんとかして実証的に示したいという、筆者のさやかな試みであつた。とくに、インターネット調査については、先生は批判的ではあつたが、しかし非常な関心を持っておられ、科学的に解明することが必要であると、たびたび主張されていた。林先生、水野先生と、調査への深い理解と信念をお持ちの2つの星を失ったことはかえすがえす残念でならない。ここに、改めて、泉下にあるお二人のご冥福をお祈りする。

【補 遺】

本報告を執筆し、この企画を提案された岩坪秀一先生(大学入試センター)や和文誌編集委員長である林文先生と、やり取りを繰り返している中で、第4次実験調査が進み、上に述べた諸事項のうち、とくに「調査の質」に関わる事項の重要な要素であると指摘した「調査不能」「無回答」の情報が具体的に見えるように

なつてきた。折角の機会であるから、速報の意味を兼ねて、その結果の一部を要約して示したい。なお、ここに記述の情報は、日本行動計量学会第30回大会(2002年9月)、特別セッション「インターネット調査の諸特性と今後の展開のあり方」において、吉村幸・大隅昇・清水信夫「インターネット調査の諸特性と今後の展開のあり方—第4次実験調査から見えてきたもの—」として発表された内容の一部である。

ここに示す内容は、第4次実験調査に参加協力いただいた調査機関のうちの2サイト(電通リサーチ、博報堂—東京サーベイ・リサーチ・グループ)で取得のデータの分析から得た情報である。第4次実験調査では、第2次、3次実験調査の基本方針を踏襲し、さらに以下を新たな検証事項として加えた。今回の第4次実験調査の全体像は表A1に示した通りである。ここで、表中に示した「有効回収回答」は、我々が独自の定義のもとに計数したもので、各調査機関がそれぞれ定める有効回答数とは異なる場合がある。またここでは、2社のインターネット調査システム、電通リサーチ社の*i*-system (DENTSU_R-net および Hot Panel) および博報堂—東京サーベイ・リサーチの e-HABIT という、それぞれ独自に構築したシステム下で調査を行った。また、電通リサーチの2回の調査における計画サンプルは、それぞれ独立にリソースから無作為に抽出し作成したものである(いわゆるリソース・サンプリング方式を適用)。一方、e-HABIT では、表A1にあるようにパネル登録者の全数を計画サンプルとした。

1) トラッキング(回答者行動を電子的に追跡し関連データを収集する操作)により、Web調査の特性を分析、とくに技術的な問題や事象を、より詳細に把握するとともに、Web調査の質的側面の評価を、「調査不能」「無回答」の理由を探索することで詳細な分析を試みること。

2) リソース、パネルの構築方法の類似・差異と調査結果との関連を検証する。とくにセルフ・セレクション型の構築方法と、従来型調査における調査対象者の抽出方法(たとえば、住民基本台帳からの無作為抽出、マスターサンプルの利用等)に基づいて作成されたリソース構築方法との比較に焦点をあてること。

3) 調査票のページ分割: 調査票を複数ページに分割した場合(改ページ方式)、しない場合(巻物方式)で、回答者行動その他にどのような影響が生じるかの検証を行うこと。

4) 設問レイアウト実験: 設問の回答選択肢のデザ

表 A1. 第4次実験調査全体の概要

WEB 調査	電通リサーチ		博報堂-東京サーベイ ・リサーチ	日本リサーチセンター
	第1回：生活意識編			
リソース/パネル名	DENTSU_R-net	Hot Panel	e-HABIT	Cyber Panel
調査方式	リソース内 サンプリング方式	リソース内 サンプリング方式	パネル方式	リソース内 サンプリング方式
調査時期	3/28/2002-4/4/2002	6/13/2002-6/20/2002	3/28/2002-4/8/2002	4/11/2002-4/25/2002
調査票のページ分割	×	△ (半数に実施)	×	○
設問レイアウト実験	○	△ (半数に実施)	○	○
有効回収回答 (%)	939 (61.6)	3,392 (42.4)	931 (88.2)	725 (36.3)
第2回：インターネット編				
調査時期	5/16/2002-5/23/2002	6/20/2002-6/27/2002	5/16/2002-5/27/2002	5/22/2002-6/7/2002
調査票のページ分割	×	×	×	○
有効回収回答 (%)	894 (59.1)	2,587 (32.4)	896 (84.9)	691 (34.6)
オムニバス調査	電通リサーチ：DRPS	日本リサーチセンター：NOS	郵送調査	HABIT2001
	第1回：生活意識編			生活意識+暮らし向き編
調査期間	3/22/2002-3/31/2002	5/7/2002-5/15/2002	調査期間	7/4/2002-7/25/2002
方法	面接法	訪問留置自記式	方法	郵送
対象地域	東京 30 km 圏	日本全国	対象地域	首都圏 40 km・近畿 20 km
対象者	15~59 歳男女	15~79 歳男女	対象者	HABIT2001 の一部 15~69 歳男女
サンプリング法	住民基本台帳からの 二段無作為抽出	住民基本台帳からの 層化多段無作為抽出		
計画サンプル数	1,236	2,200	計画サンプル数	1,000
回収数 (%)	630 (51.0)	1,226 (60.7)	回収数 (%)	946 (94.6)*
不能票数 (%)	606 (49.0)	864 (39.3)	不能票数 (%)	54 (5.4)*
第2回：暮らし向き編			*) 2002年7月25日時点	
調査期間	5/17/2002-5/26/2002	5/7/2002-5/15/2002		
計画サンプル数	1,001	2,200		
回収数 (%)	630 (62.9)	1,389 (63.1)		
不能票数 (%)	371 (37.1)	811 (36.9)		

イン効果を検証するためのレイアウト実験を行う。なおこれは海外の研究者グループとの共同実験である。

5) 比較のための従来型調査として、オムニバス調査及び郵送調査を行うこと。

とくにここでは、事項1)にある「トラッキング」で取得した情報から見えてくることとして、Web調査の質の評価とその標準的なあり方・考え方に着目して得られた結果を報告する。

a. Web調査の質の評価

従来型調査では、まっとうな調査であるなら、少なくとも計画サンプルとその抽出方法、有効回収数、不能票数や不能理由等の情報が報告されてきた。こうした基礎情報を通してその調査の性質や質的内容(調査の質)が客観的に評価され、それを前提として調査結果の分析や内容解釈が行われる。

一方Web調査では、調査の質の評価に関する情報として、「何を、どこまで、どのように」記述すべきか具体的な概念や共通認識が未だない。表A2(A2-1、

表 A2-1. 電通リサーチ DENTSU_R-net 調査概要

調査回	第1回調査	第2回調査
調査コード	010000-077, 129, 159~162	01000-078
調査期間	3/28/2002-4/4/2002	5/16/2002-5/23/2002
調査テーマ	生活意識編	インターネット編
謝礼	回答者全員に500アルネ (500円相当)	回答者全員に500アルネ (500円相当)
調査方式	リソース内サンプリング方式	リソース内サンプリング方式
リソース	DENTSU_R-net	DENTSU_R-net
登録者数	24,053	24,053
計画標本数	1,542	1,542
依頼発信前除外	17	30
調査依頼発信数	1,525	1,512
有効回収回答 (%)	939 (61.6)	894 (59.1)
無回答 (%)	586 (38.4)	618 (40.9)
未着 (%)	63 (4.1)	34 (2.2)
無接触 (%)	420 (27.5)	414 (27.4)
アクセスのみ (%)	97 (6.4)	163 (10.8)
回答送受信異常 (%)	6 (0.4)	7 (0.5)
アクセス数 (延べ数)	1,270	1,403
アクセスのみ (延べ数) (%)	298 (23.5)	498 (35.5)
回答送信 (延べ数) (%)	972 (76.5)	905 (64.8)
回答送受信異常 (延べ数) (%)	27 (2.1)	11 (0.8)
回収回答 (延べ数) (%)	945 (74.4)	894 (63.7)
重複回収回答 (%)	6 (0.5)	0 (0)
非登録者回答	生起しない	生起しない
ID不明回答	0	0
調査依頼未着 (延べ数)	65	36
登録者数確定日時	3/14/2002	3/14/2002
リマインダー	1回 (E-mail, 4/3)	1回 (E-mail, 5/22)
発信数	720	723
未着数	-	-

*) ここに記す注は、表 A2-1, A2-2 に共通した説明である。

注1) アクセスのみ: 指定 URL へアクセスしたが、回答送信ボタンを押していないもの

注2) 回答送信異常: 回答送信ボタンを押したが、回答が正常に完了していないもの

注3) 重複回収回答数: 延べの回収回答のうち非採用とした重複する回答数

注4) (%) の母数は調査依頼発信数

注5) 未着: 調査依頼メールが戻ってきたもの (住所不明に類似)

注6) 無接触: 調査票にアクセスしない (不在・回答拒否に類似)

注7) アクセスのみ: 調査票にアクセスするが回答は完了せず (回答意思の有無は特定できない)

注8) 回答送受信異常: 回答完了し送信操作を行ったが受信に失敗 (回答の意思あり)

注9) (%) の母数は総アクセス数 (延べ数)

表 A2-2. 電通リサーチ Hot Panel 調査概要

調査回	第 1 回調査	第 2 回調査
調査コード	02000-099~105	02000-107
調査期間	6/13/2002-6/20/2002	6/20/2002-6/27/2002
調査テーマ	生活意識編	インターネット編
謝礼	抽選で 1,200 名に 500 円の図書券	抽選で 1,200 名に 500 円の図書券
調査方式	リソース内サンプリング方式	リソース内サンプリング方式
リソース	Hot Panel 5 & 6	Hot Panel 5 & 6
登録者数	64,629	64,600
計画標本数	7,996	8,000
発信前除外	3	8
調査依頼発信数	7,993	7,992
有効回収回答 (%)	3,392 (42.4)	2,587 (32.4)
無回答 (%)	4,601 (57.6)	5,405 (67.6)
未着数 (%)	560 (7.0)	577 (7.2)
無接触 (%)	3,216 (40.2)	3,494 (43.7)
アクセスのみ (%)	778 (9.7)	1,302 (16.3)
回答送受信異常 (%)	47 (0.6)	32 (0.4)
アクセス数 (延べ数)	5,126	5,226
アクセスのみ (延べ数) (%)*	1,628 (31.8)	2,581 (49.4)
回答送信 (延べ数) (%)*	3,498 (68.2)	2,645 (50.6)
回答送受信異常 (延べ数) (%)*	102 (2.0)	56 (1.1)
回収回答 (延べ数) (%)*	3,396 (66.3)	2,589 (49.5)
重複回収回答 (%)*	4 (0.1)	2 (0.0)
非登録者回答	生起しない	生起しない
ID 不明回答	0	0
調査依頼未着 (延べ数)	781	579
登録者数確定日時	6/7/2002	6/10/2002
リマインダー	1 回 (E-mail, 6/19)	1 回 (E-mail, 6/26)
発信数	5,270	6,072
未着数	-	-

A2-2)、表 A3 は、第 4 次実験調査で得た情報に基づき、独自に整理を試みた、調査の質的側面を示すと思われる事項の要約の“一部”である。Web 調査の報告に際しては、少なくともこの程度の情報を、共通した調査概要情報として開示し、調査の結果の品質を明らかにする必要がある。

a-1. 「無回答」の内訳から見る Web 調査の質

現行の多くの Web 調査では「調査不能」や「無回答」についての情報が記述されることはまずない。しかし、従来型の調査と同様、無回答となった理由や内訳を検

証することは調査の質と信頼性を高める上で必要不可欠である。

たとえば「未着」(未達)は発信された調査依頼が発信先に届かずに戻ってきた数である。リソースやパネルの維持・管理状態がここに反映されることが可能性として考えられる。

「無接触」は調査依頼に対して何の反応もなかった数である。従来調査での不在や回答拒否に類似する。表 A2、表 A3 を比べると、リソースまたはパネルごとにその値は異なり、しかし各リソース内では 2 回の調査

表 A3. 博報堂-東京サーベイ・リサーチ e-HABIT 調査概要

調査回	第1回調査	第2回調査
調査コード	200201	200202
調査期間	3/28/2002-4/8/2002	5/16/2002-5/27/2002
調査テーマ	生活意識編	インターネット編
謝礼	回答者全員に5ポイント (500円相当)	回答者全員に5ポイント (500円相当)
調査方式	パネル方式	パネル方式
パネル	e-HABIT	e-HABIT
登録者数	1,055	1,055
計画標本数	1,055	1,055
依頼発信前除外	0	0
調査依頼発信数	1,055	1,055
有効回収回答 (%)	931 (88.2)	896 (84.9)
無回答 (%)	124 (11.8)	159 (15.1)
未着 (%)	0	3 (0.3)
無接触 (%)	113 (10.7)	125 (11.8)
アクセスのみ (%)	11 (1.0)	31 (2.9)
回答送受信異常 (%)	—	—
アクセス数 (延べ数)	1,136	1,153
アクセスのみ (延べ数) (%)	149 (13.1)	219 (19.0)
回答送信 (延べ数) (%)	987 (86.9)	934 (81.0)
回答送受信異常 (延べ数) (%)	—	—
回収回答 (延べ数) (%)	987 (86.9)	934 (81.0)
重複回収回答 (%)	56 (4.9)	38 (3.3)
非登録者回答	生起しない	生起しない
ID不明回答	7	6
調査依頼未達 (延べ数)	—	—
登録者数確定日時	3/25/2002	3/25/2002
リマインダー	1回 (E-mail, 4/4)	1回 (E-mail, 5/23)
発信数	356	394
未着数	2	4

注1) 未着: 調査依頼メールが戻ってくる(住所不明に相当)。なお、第1回調査では初め2ケースの未達が生じたが電話でアドレスを再確認の後、再発送、2ケースとも回収回答に至っている。また第2回調査では初め6ケースの未達が生じたが、電話確認の後、うち3ケースが回答回収に至っている。

注2) 総アクセス数(延べ数)、アクセスのみ(延べ数)、回答送信数(延べ数)、回収回答数(延べ数)にはID不明者による回答は含まれない。

間で値が類似している。このことから「無接触」にはリソースやパネルの構築方法やその維持・管理の状態が影響する可能性が示唆される。

「アクセスのみ」は調査票ページにアクセスするな

ど、調査依頼に対し何らかの行動を起こしたものの、回答送信まで至らなかった回答者の数(接触数)である。表A2、表A3から、リソース間での違いは確かにあるようで、しかも共通して第2回調査でこの値がほぼ倍

増している。またここには記載しないが、第2回調査での平均回答所要時間は第1回よりも長く、これを考慮すると、「アクセスのみ」には調査の内容や量が影響することが考えられる。

「回答送受信異常」は、回答者が回答送信の操作を行ったとの記録があるにも関わらず回答データが取得されなかった数である。郵送調査における郵便事故に類似する。必ずしも調査実施機関の問題だけでなく、インターネットの仕組みや回答者のコンピュータ・リテラシーに起因する要素もあるだろうが、Web調査の技術面の評価の一つの手がかりでとなる。

a-2. 「アクセス数」から見た Web 調査の質

上で述べた無回答の内訳はいずれも回答者数（調査客体）を基準として見たものである。ここではアクセス件数（発生事象）を基準とした考察を行う。

表 A2, 表 A3 に示した「アクセス数（延べ数）」は、回答者個人を区別せずに延べて調査票への接触がどれだけあったかを表している。どのリソースあるいはパネルにおいても、第2回調査のアクセス数が第1回に比べて多い。また、第2回調査では「アクセスのみ」が増え、「回答送信数」、「回収回答数」は減っていることが共通しているのも興味深い。つまりアクセスはするがそれが回答に結びつかなかったというケースが、第1回調査に比べ第2回で増えていることを意味する。おそらくは両調査の内容と量の違いに起因する現象であると推察される。たとえばアクセス数に対する回収回答数の割合（%）を回答の「完遂率」と定義すると、こうした現象を計量的に記述することも可能となる（今回の場合の完遂率は、第1回調査：DENTSU_R-net = 74.4%, Hot Panel = 66.3%, e-HABIT = 86.9%; 第2回調査：DENTSU_R-net = 64.5%, Hot Panel = 49.5%, e-HABIT = 81.0%, となった）。

以上にみるように、インターネット調査、とくに Web 調査を適切に使いこなすには、その調査の質をいかに計量的な指標として記述し評価するかが不可欠である。とくに、回答者行動の追跡方法（トラッキング）と、その具体的な技法を実装したインターネット調査システムを組織的に構築することが重要であることを、これらの結果が示している。こうしたシステムから得られるデータを詳細分析することにより、リソースやパネルの維持管理状態、調査票設計、回答取得システム等の適切な評価方式や調査自体の質の評価に有効な各種指標を得られることを示唆している。同時に、こうした結果は、インターネット調査の特性を、従来

型調査との対比の中で客観的かつ実証的に議論することから得られるものであり、調査機関・研究者間のコラボレーションが肝要であるとの我々の主張を裏付けるものである。もちろん、本稿に述べたことが、インターネット調査が抱える、無数にある本質的な問題の解決に直ちに結びつくとは限らない。しかしながら、こうした実験調査や調査現場との共同研究を継続的に続けることこそが、問題解決への早道であることは間違いない。

謝 辞

第1次調査から第3次調査まで、一連の実験調査の主旨に賛同いただき、コラボレーションにご協力いただきました関係各位、各社に、この場をお借りして厚く謝意を表す。

また、第4次調査における貴重な情報の開示、とくに国内で初めてのインターネット調査に関する調査不能理由情報の開示に際して、これをご快諾いただいた電通リサーチ、博報堂、東京サーベイ・リサーチの3社ならびに関係者諸氏に厚く御礼申し上げます。

参 考 文 献

- [1] Black, G.S. (2000). *The Harris Interactive Election Forecast*. <http://www.harrisinteractive.com/news/USA>.
- [2] Clemente, P. (1998). *State of the Net-The New Frontier*, McGraw-Hill.
- [3] Commerce Net/Nielsen (1996). *Commerce Net/Nielsen Internet Demographic Survey*, The CommerceNet, March/April 1996, USA.
- [4] Couper, M.P. (2001). *The Promise and Perils of Web Surveys, The Challenge of the Internet*, (private information from the author), Edited by A. Westlake et al, Compilation (c) 2001 Association for Survey Computing.
- [5] Couper, M.P. (2001). *Web Surveys: A Review of Issues and Approaches*, *Public Opinion Quarterly*, 64 (4), 464-494.
- [6] Couper, M.P., Baker, R.P. and others (1998). *Computer Assisted Surveys Information Collection*, John Wiley & Sons.
- [7] de Leeuw, D. and Nicholls, W. (1996). *Technological Innovations in Data Collection: Acceptance, Data Quality and Costs*, *Sociological Research Online*, vol. 1, no. 4, USA.

- [8] Dillman, D.A. (2000). *Mail and Internet Surveys-The Tailored Design Method*-, John Wiley & Sons.
- [9] Jones, S. ed. (1999). *Doing Internet Research-Critical Issues and Methods for Examining the Net*-, SAGE Publications.
- [10] Saris, W.E. (1991). *Computer-Assisted Interviewing*, A Sage University Paper, SAGE Publications.
- [11] Vehovar, V., Couper, M., Manfreda, K.L., Virtual Selves and Web Survey Nonresponse, Paper presented at the *12th International Workshop on Household Survey Nonresponse*, Oslo, 12-14 September 2001.
- [12] インターネットビジネス研究会 (2001). インターネットビジネス白書 2002, ソフトバンクパブリッシング.
- [13] インターネット協会監修 (2001). インターネット白書 2001, インプレス.
- [14] インターネット協会監修 (2002). インターネット白書 2002, インプレス.
- [15] 総務省 (2001). 平成 13 年版情報通信白書.
- [16] 細井 勉 (2002). マーケティング・リサーチ領域におけるインターネット調査概論—現状と展望—, エストレーラ, 95号, 2月号, 2-10.
- [17] 横原 東 (2001). マーケティングにおけるインターネット調査の実状と課題, 統計数理研究所公開講演会抄録, 統計数理, 49, 215-222.
- [18] 吉村 宰 (2001). インターネット調査にみられる回答者像, その特性, 統計数理研究所公開講演会抄録, 統計数理, 49, 223-229.
- [19] (株)日本マーケティング・リサーチ協会, インターネット・リサーチ委員会編 (1999). 日本におけるインターネットリサーチの現状と課題—ガイドラインについての考え方—.
- [20] 大隅 昇 (2000). 「調査環境の変化に対応した新たな調査法の研究」報告書, 文部省科学研究費, 特定領域研究「統計情報活用のフロンティアの拡大」(略称: ミクロ統計データ), 研究計画 A02 班(公募研究)「マイクロデータ利用の社会的制度の問題点」(課題番号: 09206117).
- [21] 大隅 昇 (2001). 調査環境の変化と新しい調査法の抱える問題, 統計数理研究所公開講演会抄録, 統計数理, 49, 201-213.
- [22] 大隅 昇 (2002). インターネット調査の適用可能性と限界—実験調査から見えること—, 輿論科学協会創立 56 周年記念講演, 市場調査, 250号.
- [23] 大隅 昇, Ludovic Lebart (2000). 調査における自由回答データの解析—InfoMinerによる探索的テキスト型データ解析—, 統計数理, 48, 339-376.
- [24] 大隅 昇 (2000). 定性情報のマイニング—自由回答データの解析—, エストレーラ, 74号, 5月号, 14-26.
- [25] 大隅 昇 (2002). インターネット調査, 「社会調査ハンドブック」, 朝倉書店.
- [26] 大隅 昇, Ludovic Lebart (2002). テキスト型データの多次元データ解析. —Web 調査自由回答データの解析事例, 「多変量解析実例ハンドブック」, 朝倉書店.

(2002年2月12日受付)