

# *The Science of Web Surveys*

By

Roger G. Tourangeau, Frederick G. Conrad, and Mick P. Couper

Oxford University Press (2013)

翻訳版作成 鳩 真紀子, 大隅 昇

(\* 下に表記のページは原書に対応.)

<b>7. ウェブ調査と他のデータ収集方式における測定誤差</b>	<b>129</b>
7.1 調査方式効果を理解するための概念的枠組み	129
7.2 自記式としてのウェブ調査	132
7.3 ウェブ調査と認知的負担	146
7.4 この章のまとめ	149
<b>7. Measurement Error on the Web and in Other Modes of Data Collection</b>	<b>129</b>
7.1 Conceptual Schemes for Understanding Mode Effects	129
7.2 Web Surveys as a Method of Self-Administration	132
7.3 Web Surveys and Cognitive Burden	146
7.4 Summary	149

## 第 7 章 ウェブと他のデータ収集方式における測定誤差

ウェブ調査と従来のデータ収集方式の差のいくつかについては、すでにこれまでの章で触れてきた。第 2 章では、ウェブ調査と他の方式で行われる調査のカバレッジの差について検証した。第 3 章では、ウェブ調査と他のタイプの調査、とくに郵送調査との回答率の差を浮き彫りにした。本章では、ウェブ調査と他のタイプの調査の測定差に焦点を当てた研究を分析する。本章は、測定における調査方式の差について考えるための枠組みをいくつか系統的に提示し、その後でウェブ調査で得られた回答が他のデータ収集方法で得られた回答とどのように違うのかについての知見を再検討する。

### 7.1 調査方式効果を理解するための概念的枠組み

調査方式効果とは、データを収集するために使用した方法によって生じた調査結果の差を意味する。どのような調査方式の差も、測定誤差の一種であると考えられるリサーチャーもいるが、著者らは 2 つの調査方式の差が、調査誤差の主な原因—つまり、標本抽出、カバレッジ、無回答、および測定誤差—の差の総効果を反映するような概念化を好む。著者らの見解では、たとえば、ある郵送調査からの推定値と、並行して行われたウェブ調査からの推定値の差は、これらの手法の間にある 2 つの潜在的な差を反映している。第 1 に、回答率の差、標本抽出の技術、あるいは母集団のカバレッジの差により、さまざまなタイプの人々が 2 つの調査方式で質問に回答することになってしまう。たとえば、ウェブ調査は郵送調査とくらべると、若い世代、あるいは教育水準の高い人々を過大に代表する可能性があり、回答のあった標本の構成におけるこれらの差が、調査方式による回答分布に差を生じる可能性がある。第 2 に、同じ人がウェブ調査と郵送調査に回答するときでさえも、各調査方式で質問をどのように提示するかの違いによって、あるいはこの 2 つの調査方式間の他の測定差によって、ウェブ調査と郵送調査の調査票で異なった回答をする可能性が依然としてある。第 2 章と第 3 章では、標本抽出、カバレッジ、および無回答に起因する差について検討した。本章では、後者—調査方式間の測定差に起因する—のタイプの調査方式による差について集中的に取り組む。

長年にわたって、リサーチャーはさまざまなデータ収集方法におけるそのような測定差を説明するさまざまな概念的枠組みを提案してきた。Tourangeau と Smith (1996) は、最初の 1 つとなる枠組みを提案した。Tourangeau と Smith (Tourangeau, Rips, and Rasinski, 2000 も参照) によると、データ収集の主な手法は以下の 4 つの主要側面で異なっている。

- 質問が自記式によるものか、それとも調査員方式によるものか
- どのように回答者に接触するか(例：直接か、あるいは電話か)

- 調査票はコンピュータ支援によるものか、それとも紙面に提示されるものか
- 質問が回答者に視覚的に伝えられるか、聴覚的に伝えられるか

ウェブ調査では、質問は自記式で行われ、通常はコンピュータを介して視覚的に提示される。回答者との接触方法はEメールであることが多いが、郵送もありうる。TourangeauとSmithは、これらのデータ収集方式の客観的な特性が主要媒介変数—回答者が1人きりになれる感覚や、質問に回答する際に周りに誰かがいるという感覚、回答時の認知的負担<sup>1</sup>、調査の重要性や正当性—に影響をおよぼすと論じている。たとえば、TourangeauとSmithは、写真付きのIDバッジや他の正当性を与える手回り品を提示する調査員が直接接触すると、回答者はその調査がより重要であると思い、結果的に異なる回答をする可能性があることを示唆している。これらの3つの媒介変数は、同様に回答の信用性、欠測データの度合い、あるいは回答者が微妙な内容の情報を開示しようとする意欲といった、さまざまな結果に影響する。

他のリサーチャーたちは、調査方式間の測定差を理解するための代替となる枠組みを提供してきた。Groves, Fowler, Couper, Lepowski, Singer, および Tourangeau (2009)は、データ収集のさまざまな方式について異なる5つの特徴を区別している。

- 1) 調査員の関与度（郵送またはウェブ調査での無関与から、面接調査での最大関与。調査員が最初に回答者に接触し、自動データ収集システムへと移行するデータ収集方式—オーディオ・コンピュータ支援の自記式、あるいは音声自動応答装置—での調査員の関与度を中間レベルとする）
- 2) 回答者との相互行為(interaction)の度合い（行政記録からデータを抜粋する場合はほとんど相互行為がなく、電話聴取を中間レベルとし、回答者の自宅で行われる面接で相互行為の度合いは最大限となる）
- 3) プライバシーの度合い（面接中に調査員や他の人たちが周りにいる場合の低プライバシーから、回答者が面接に自答した場合の高プライバシーまで）
- 4) コミュニケーション・チャンネル（視覚、聴覚、またはこれら2つの組み合わせ）
- 5) テクノロジーの使用（質問紙調査では使用度が低く、コンピュータ支援の自記式調査で調査機関がコンピュータを提供する場合は中程度、回答者が通常、コンピュータ、インターネット接続、ブラウザなどを提供するウェブ調査ではテクノロジーの使用度は高レベルとなる）。

この枠組みはTourangeauとSmith(1996)が提供したモデルをさらに詳しく説明してお

---

<sup>1</sup> 訳注：ここで、cognitive burdenを「認知的負担」とした。これは、調査内容をみた回答者が、その内容や回答方式（面接員を意識する、PCのリテラシーの影響を受けるなど）から、回答行動に心理的になんらかの影響があるということ。本章に頻出する。

り、さまざまなデータ収集方式によって異なる主要な特徴のいくつかは、性質ではなく、むしろ程度の問題であることをはっきりさせている。

De Leeuw (1992; 2005)は、データ収集方式の違いによって変わる3つの要因を区別している一すなわち、媒体に関係した要因、情報伝送に関係した要因、そして調査員効果である。de Leeuw が提唱する最初の要因は、「調査方法で使われる媒体に関連した社会慣習や習わし」と関係している(de Leeuw, 2005, 244 ページ)。たとえば、調査員方式の調査では、調査員がやりとりを開始するので、質問の順番やペースを管理するのは調査員である。質問紙による自記式調査では、ペースと質問の順番を管理するのは回答者である。またウェブ調査では、ウェブ上での他の相互行為にあてはまる慣習や習わし(複数のウィンドウを開きウィンドウを切り替えるなど)が調査設定に引き継がれるため、回答者がどのように調査票と相互行為を行うかに影響をおよぼす可能性がある。de Leeuw の提唱する2つ目の要因つまり、情報伝送一は、提示(視覚に対し聴覚)を行うチャンネルと、それとは異なるチャンネルからさまざまな形で入手することができる潜在的な伝達情報(テキスト、非言語的の手がかり、そしてトーンやタイミングといったパラ言語情報)を網羅している。視覚的な調査では、テキストの特性(ボールド体の使用など)は、口語で実施される調査におけるパラ言語的な手がかり(強調など)と同じように機能する(Redline と Dillman, 2002 を参照)。残る1つの要因となる調査員効果とは、調査員がどのように質問を行ったかの差、または人種や性別といった調査員の個人的特徴の差によって生じる回答の差のことを伝統的に言う。おそらく、バーチャル調査員を調査に組み込まないかぎり、ウェブ調査ではそのような効果が減じられる、または除去される(第6章参照)。

Couper と Bosnjak (2010)は、ウェブ調査についての議論において、このデータ収集方式の5つの特徴に焦点を当てている。ウェブ調査は自記式で、コンピュータ化されており、双方向的で、分散しており、豊かな視覚性を備えている。Couper と Bosnjak はこれらのほとんどが、従来型のデータ収集方法と比較すると、ウェブ調査の利点であると考えている。たとえば、自記式であることにより、データ収集の費用が減少し、社会的望ましさの偏りを抑制し、誤差源としての調査員変動を取り除く。それでもなお、Couper と Bosnjak が言及しているように、調査員は回答者のやる気をおこさせ、不明瞭な質問をはっきりさせることにおいて重要な役割を果たす可能性がある。ウェブ調査は配信されるということ(つまり、回答者本人のコンピュータを使用し、回答者のインターネット接続を介して回答を行う)は、一長一短である。なぜなら、回答者のシステムが有する特性の多くが、調査がどのように提示されるかに影響する可能性があるからである。Couper と Bosnjak (2010, 541 ページ)は以下のようにまとめている。

「インターネット調査では、… 回答者の装置のロック・アンド・フィールに影響しうる要素がたくさんある。これらには、ブラウザのタイプやバージョン(たとえば、インターネット・エクスプローラなのか、それとも Mozilla Firefox か)、オペレーション・システム(OS)

(たとえば, Windows, Mac, あるいは Linux) , ディスプレイの解像度, ブラウザのセキュリティ設定 (JavaScript が有効か, クッキーは有効か, など), インターネット接続方法 (たとえば, ダイヤルアップなのか, それともブロードバンドなのか), フォントの大きさやその他のブラウザ上の表示設定などが含まれる. インターネットは複数のプラットフォームで機能するよう設計されている一方, これらの差は回答者にとっての調査経験を程度の差はあれ変える可能性があるため, 潜在的に無回答誤差(中断)と測定誤差(データの品質)の両方に影響する.]

従来のコンピュータ支援によるデータ収集方法では, 調査機関のコンピュータや設定を使用しているので, このようなレベルのばらつきが回答者間で生じることはない.

本章では, これまでの章では網羅していないウェブ調査の 2 つの主な特徴に焦点を当てる. つまり, 質問の自記式 (調査員の関与をなくし, 高いレベルのプライバシーを提供する) の使用と, 認知的負担が少ない (回答者が調査のペースを自分で管理でき, 画面上の質問を容易に読み返せる) ことである. もちろん, 読解が困難, またはコンピュータ技能を有さない教育レベルの低い回答者にとっては, ウェブは認知的負担を増すだろう.

## 7.2 自記式としてのウェブ調査

**微妙な質問と報告誤差.** 調査における測定誤差の主な原因の 1 つは, 回答者が面目を保とうとする, または恥ずかしい思いをしないように, わざと回答をねじ曲げることである. 調査では, 妊娠中絶, 違法な薬物の使用, または選挙といった, 微妙な, あるいは潜在的にきまりの悪い内容の質問を回答者にすることが多い. 回答者がそのような話題の質問には正確に回答しないことが多いことにかんしては, 証拠がたくさんある. Fu, Darroch, Henshaw, および Kolb (1998)による研究では, 家族の成長についての全国調査(the National Survey of Family Growth: NSFG) の回答者が, 妊娠中絶について実際の件数の半数程度しか報告していないだろうと推定している. この過小報告の度合いの推定は, NSFG における妊娠中絶の総件数の推定値と, 妊娠中絶の施術者側にたいする全国調査から得た推定とを比較することで導き出した (Tourangeau, Rasinski, Jobe, Smith, および Pratt, 1997 も参照). 同様に, Belli, Traugott, および Beckman (2001)は, 全米選挙調査 (the American National Election Studies: ANES) において非投票者の約 20%が投票を行ったと主張しているだろうと推定している. この推定は, 選挙記録を含んだ調査報告との比較にもとづいている. これらの研究では, 調査回答者は一貫して, 社会的に望ましくない行動について幅広く過小報告し, 社会的に望ましい行動については過大に報告することをはっきりと示している (最近の検討については, Tourangeau と Yan, 2007 を参照).

調査では, 微妙な内容の情報についての報告を改善するためのさまざまな方法が用いられる. 自記式の質問, 偽のパイプライン, 無作為化回答技術などがこれに含まれる. 偽のパイプラインとは, 回答者がウソを探知できると信じる装置や手続きを意味する. たとえば, 喫煙について質問する調査において, 回答者は呼気標本や唾液標本の提供を求められ

る可能性がある(例: Bauman と Dent, 1982)。<sup>2</sup> 無作為化回答技術では、回答者が無作為化装置(スピナーやコイン・トスなど)を使用して、どの質問に回答するかを決める。回答者は、確率が既知の微妙な内容の話題にかんする2つの文のうち、1つを受け取る(A: 私は妊娠中絶をしました。B: 私は妊娠中絶をしたことはありません)。回答者は、スピナーまたはコイン・トスによって選ばれた文について、それに同意するか否かを報告するが、それがどちらの文だったかは明かさない。偽のパイプライン(例: Murray, O'Connell, Schmid, および Perry, 1987)と無作為化回答技術(メタ分析については, Lensvelt-Mulders, Hox, van der Heijden, および Maas, 2005 を参照)の両方ともが、調査報告における社会的望ましさの偏りを低減するのに有効であることを示す証拠はたくさんある。

**自記式によって得られる利益。** 偽のパイプラインと無作為化回答技術の使用には実用上の困難があるため、これらの手法を用いて微妙な内容の情報を収集している全国調査はほとんどないが、多くの全国調査では、画面上に質問を表示し、イヤフォンを介して回答者に録音した質問を再生して聞かせる、オーディオ・コンピュータ支援の自記式(ACASI)の形態で自記式を用いることが多い。たとえば、NSFG と薬物使用と健康に関する全国調査(NSDUH)では質問のいくつかでACASIを使用している。Tourangeau と Yan (2007, pp.863-867)は、質問紙型調査票を用いた自記式とACASIのようなコンピュータ化した自記式の両方で、自記式では微妙な内容の報告が増加するという主張を支持する強力な証拠を報告している。重大な問題となるのは、ウェブ調査が他の自記式手法に見られる利点を残しているかどうかである。

少なくとも14編の論文がこの問題について調べており、質問をウェブで実施する場合と、他の複数の実施法とを比較した。表7.1では、これらの研究における主要特性を提示している。それらの研究のいくつかでは、ウェブ調査と質問紙型調査を比較している(たとえば、Denniston, Brener, Kann, Eaton, McManus, Kyle, Roberts, Flint, および Ross, 2010; Denscombe, 2006)。このうち4つの研究(Chang と Krosnick, 2009; Denniston ら 2010, そして Eaton, Brener, Kann, Denniston, McManus, Kyle, Roberts, Flint, および Ross, 2010; Link と Mokdad, 2005a, b; McCabe, 2004, McCabe, Couper, Cranford および Boyd, 2006)では大規模な標本が関与しており、実際の調査条件下で実施されていることから、著者らはここでこの4つの研究についてより詳しく調べる。

---

<sup>2</sup> 呼気と唾液の標本は両方とも最近喫煙したかどうかを判断するために使用されるため、これは偽の、というよりはむしろ「真の」パイプラインである。

TABLE 7.1. Studies on Web Surveys and Reports about Sensitive Topics

Study	Target Population	Sample Sizes (respondents), by Mode	Key Findings
Bälter et al. (2005)	Adults in a single county in Sweden	Mail: 188 Web: 295	No significant difference by mode in reports of smoking status.
Bason (2000)	Students at a university	Telephone: 161 Interactive Voice Response (IVR): 128 Mail: 204 Web: 115	No significant differences by mode in reported prevalence or frequency of drug use or binge drinking; significantly higher proportions of students report <i>no</i> alcohol use in Web and IVR
Bates and Cox (2008)	Students at a university	Paper: 73 Web: 64	No differences in reported drinking or sexual behavior by mode
Chang and Krosnick (2009)	General population	<b>Pre-Election</b> Telephone: 1,506 Web: Harris Interactive (HI): 2,306 Web: Knowledge Networks (KN): 4,933 <b>Post-Election</b> Telephone: 1,206 Web: Harris Interactive (HI): 1,028 Web: Knowledge Networks (KN): 3,416	Less social desirability bias on racial attitude item in Web surveys
Denniston et al. (2010); same study is also reported by Eaton et al., (2010)	9th and 10th grade students	In-class paper: 1,729 In-class Web (no skips): 1,735 In-class Web (skips): 1,763 Web outside of class (without skips): 559	Web administration in the classroom seen as significantly less private than paper classroom administration
Denscombe (2006)	15-year olds at one school	In-class paper: 220 In-class Web: 69	Only one item in 23 shows a significant difference by mode
Eaton et al. (2010); see Denniston et al. (2010) above	9th and 10th grade students	In-class paper: 1,729 In-class Web (no skips): 1,735 In-class Web (skips): 1,763	Significant differences between Web and paper versions of the survey in reported prevalence for seven of 74 risky behaviors; reported prevalence higher in Web for all seven

(Continued)

TABLE 7.1. (Continued)

<b>Study</b>	<b>Target Population</b>	<b>Sample Sizes (respondents), by Mode</b>	<b>Key Findings</b>
Knapp and Kirk (2003)	Students at a single university	Paper: 174 IVR: 121 Web: 57	No significant differences on any of 58 sensitive questions
Kreuter, Presser, and Tourangeau (2008)	Alumni at one university	Telephone: 320 IVR: 363 Web: 320	Web respondents significantly more likely to report academic problems than CATI respondents; Web respondents had lowest false negative rates for four academic problems. No differences in reporting by mode for positive academic performance.
Link and Mokdad (2005a); this is the same study reported by Link and Mokdad (2005b) below	Adults in four states	Mail: 836 Telephone: 2,072 Web: 1,143	Web respondents report significantly more days drinking and significantly more binge drinking than telephone respondents (see Table 7.2 for more detailed findings).
Link and Mokdad (2005b)	Adults in four states	Mail: 836 Telephone: 2,072 Web: 1,143	Web respondents significantly differ from telephone respondents in six of eight health conditions (reporting significantly higher rates of diabetes, high blood pressure, obesity, and binge drinking, but lower rates of smoking and STD prevention; see Table 7.2 below).
McCabe et al. (2002); this is the same study reported by McCabe (2004) and McCabe et al. (2006) below	Undergraduates at a single university	Mail: 1,412 Web: 2,194	No differences by mode in reported drinking and smoking habits.

(Continued)



TABLE 7.1. (Continued)

Study	Target Population	Sample Sizes (respondents), by Mode	Key Findings
McCabe (2004); this is the same study reported by McCabe et al. (2006) below	Undergraduates at a single university	Mail: 1,412 Web: 2,194	Two significant differences in reporting by mode in 32 comparisons; both men and women report greater lifetime use of cocaine via the Web.
McCabe et al. (2006)	Undergraduates at a single university	Mail: 1,412 Web: 2,194	No differences by mode in reported prevalence of consequences of drug use.

表 7.1 ウェブ調査と微妙な内容の話題についての報告に関する研究

研究	目標母集団	標本の大きさ (回答者), 調査方式別	主な結果
Balter et al. (2005)	スウェーデン一国の成人	郵送 : 188 ウェブ : 295	報告された喫煙状況に調査方式間で有意差はなかった。
Bason (2000)	ある大学の学生	電話 : 161 IVR : 128 郵送 : 204 ウェブ : 115	報告された薬物使用や過度の飲酒状況に調査方式間で有意差はなかった。ウェブでは IVR とくらべてアルコールの非使用を報告した回答者が有意に多かった。
Bates and Cox (2008)	ある大学の学生	質問紙 : 73 ウェブ : 64	調査方式間で報告された飲酒あるいは性行動に差はなかった。
Chang and Krosnick (2009)	一般母集団	<b>選挙前</b> 電話 : 1506 ウェブ:ハリスインタラクティブ (HI) : 2306 ウェブ:ナレッジネットワークス : (KN) :4933 <b>選挙後</b>	ウェブ調査では、人種問題にたいする態度についての項目で、社会的望ましさの偏りが少なかった。

		電話：1206 ウェブ:ハリスインタラクティブ (HI) :1028 ウェブ:ナレッジネットワークス (KN) :3416	
Denniston et al. (2010); 同じ研究が Eaton et al. (2010)によっても報告されている	9 学年及び 10 学年の生徒	教室内質問紙：1729 教室内ウェブ (スキップなし)：1735 教室内ウェブ (スキップあり)：1763 教室外ウェブ (スキップなし)：559	教室内においてウェブで実施した場合、教室内で質問紙を用いて実施した場合とくらべ、プライバシーが有意に少ないと見なされた。
Denscombe (2006)	ある学校の 15 歳	教室内質問紙：220 教室内ウェブ：69	調査方式間に有意な差があったのは 23 項目中 1 項目だけだった。
Eaton et al.(2010); 上述の Denniston et al. (2010)を参照	9 学年及び 10 学年の生徒	教室内質問紙：1729 教室内ウェブ (スキップなし)：1735 教室内ウェブ (スキップあり)：1763	ウェブ版と質問紙版の調査では、報告された 74 項目の危険行動のうち、7 項目で有意差があった。7 項目すべてにおいて、報告はウェブでより多かった。
Knapp and Kirk (2003)	ある大学の学生	質問紙：174 IVR:121 ウェブ：57	58 項目の微妙な内容の質問のすべてについて有意差はなかった。
Kreuter, Presser, and Tourangeau (2008)	ある大学の同窓生	電話：320 IVR:363 ウェブ：320	ウェブ回答者は CATI の回答者とくらべて学業問題を報告する可能性が有意に高かった。ウェブ回答者では 4 つの学業問題についての偽陰性率が最も低かった。肯定的な学業成績についての報告には調査方式間で有意差はなかった。
Link and Mokdad	4 つの州の成人	郵送：836	ウェブ回答者は電話回

(2005a); これは後述の Link and Mokdad (2005b)と同じ研究		電話：2072 ウェブ：1143	答者とくらべ、過度な飲酒の報告日数が有意に多かった（詳細結果については表 7.2 を参照）.
Link and Mokdad (2005b)	4つの州の成人	郵送：836 電話：2072 ウェブ：1143	8項目中6項目の健康状態についての質問で有意差があった。（ウェブ回答者は電話回答者とくらべ、糖尿病、高血圧、肥満、および過度の飲酒の報告率が有意に高かったが、喫煙と性感染症予防についての報告率は低かった。詳しくは表 7.2 を参照）.
McCabe et al. (2002);後述の, McCabe (2004) および McCabe et al. (2006)と同じ研究.	ある大学の学部生	郵送：1412 ウェブ：2194	報告された飲酒および喫煙習慣について、調査方式間で有意な差はなかった.
McCabe (2004);これは後述の McCabe et al. (2006)と同じ研究	ある大学の学部生	郵送：1412 ウェブ：2194	32件の比較中、2件の有意差が調査方式間にみとめられた。男女とも、ウェブによる報告ではコカインの生涯使用率が高かった.
McCabe et al. (2006)	ある大学の学部生	郵送：1412 ウェブ：2194	報告された薬物使用の結果について調査方式間に有意差はなかった.

**選挙調査の比較.** Chang と Krosnick(2009)は、オハイオ州立大学の調査研究センター (CSR)が実施した2つの電話調査と、同様の調査を2大ウェブ調査パネル—ナレッジ・ネットワーク社 (KN) とハリス・インタラクティブ社 (HI) —の会員を対象にウェブで行った調査から得られた結果とを比較した。初回の調査は2000年の6月と7月に実施され（つまり、その年に行われた大統領選の直前）、2回目の調査は11月（大統領選直後）に行われた。選挙前調査に回答した回答者には、選挙後調査への参加も依頼した。選挙前調査の回答率はCSRの電話調査で43%、KNパネルでは25%だった。HIパネルはボランティアの標本

だったため、回答率は算出できなかった。選挙後調査での再面接率は、CSR の電話標本の回答者で 80%、KN 標本で 82%、HI 標本では 45%だった。調査間の比較ではカバレッジと無回答の差だけではなく、測定差もあきらかに反映していた。これら 3 つ全ての調査から得たデータは、それぞれのデモグラフィック構成の差を調整するために加重した。

著者らの目的では、アフリカ系アメリカ人への支援を増やすべきか、支援を減らすべきか、それとも現状のままとするか、という質問にたいする白人回答者の回答が主な結果であった。Chang と Krosnick は、支援を減らすべきと唱えることは白人にとって社会的に望ましくないことであると論じている。電話回答者では、この回答を選択した人(17.0%)は、ウェブパネル(KN パネルで 35.8%、HI パネルで 42.5%)の会員より有意に少なく、これらの標本間の差はデータに重みづけをし、標本間の背景特性の差を調整するために共分散を使用した場合にも引き続き見受けられた。

**青少年危険行動調査 (Youth Risk Behavior Survey: YRBS) の実験。** 2 番目に規模の大きな調査方式の比較研究は、Denniston と同僚ら (2010, Eaton ら 2010 も参照) によって行われた。彼らは YRBS の調査票への回答を詳しく調べた。彼らの研究では、1) 学生に教室内で質問紙型調査票に回答してもらう (これは YRBS で通常行われている手続き)、2) 教室内で実施するウェブ調査 (質問紙型調査票により似せるため、スキップパターンがないバージョンとしてプログラムされている)、3) スキップパターンを含んだウェブ調査を教室内で実施する、という 4 つのデータ収集条件を比較している。参加者は、15 の州の 85 校から成る 9 年生および 10 年生の便宜的標本である。各校で 4 つの学級を選択し、1 学級をそれぞれの実験条件に割り当てた。ほとんどの学校で、ウェブ調査は校内のコンピュータ・ラボで実施した。

調査票には 2007 年の YRBS からの 77 項目が含まれ、「不慮のけがおよび暴力、喫煙、アルコールおよび他の薬物使用、性行動、体重管理行動、および身体活動」について問う、微妙な内容の質問が 70 項目含まれた (Eaton ら, 2010, 141 ページ)。12 項目の追加質問では、調査自体についての質問を行った。5000 人以上の生徒が 3 つの教室内条件に参加し、追加で 500 人以上の生徒が自記入方式の調査票に回答した。この研究者らは、2 つの教室内ウェブ条件にはほとんど差がないことを発見し、これら 2 つのグループの結果を 1 つにまとめた。残る 1 つのウェブのグループは、グループ内での回答率が低かったことから脱落したようである (その条件下での回答率は 28%、それにたいし残り 3 つの条件下では 90% 以上)。学校内のウェブ回答者は、質問紙型調査票に回答した回答者とくらべると、74 項目中 7 項目の危険行動についての報告が有意に高かった (飲酒した運転者に同乗した、学校に武器を持ち込んだ、デートの相手に殴られた、学校の敷地内でマリファナを使用した、セックスの前に飲酒した、あるいは薬物を使用した、禁煙の努力をしていない)。さらに、Denniston と同僚ら (2010) は、教室内のウェブ調査は、質問紙型調査票とくらべてプライバシーが有意に少ないと思われていたが、それでもなお、ウェブ版の調査票の回答者は質問紙型調査票の回答者よりも多くの危険行動を報告した。

行動危険因子監視システム（Behavioral Risk Factor Surveillance System (BRFSS)) の実験. Link と Mokdad (2005a,b)による論文の2編では、もう1つの大規模調査を報告しており、ここでは BRFSS で使用した電話版、郵送版、そしてウェブ版の調査票の比較が行われた。BRFSS は従来から電話で実施されているが、電話調査の回答率は減少し続けており (Curtin, Presser, および Singer, 2005), リサーチャーはデータ収集の代替方法を模索していた。Link と Mokdad は、2003 年の秋に2つの実験を4つの州で行った。1つ目の実験では、標本構成員を招いてオンラインで調査に回答してもらい、2つ目の実験では標本構成員に調査票を郵送した。両実験とも、回答は同月に同じ州で実施された現行の BRFSS のコンピュータ支援による電話聴取 (CATI) の回答と比較を行った。電話標本は RDD で選出し、住所と適合する電話番号のみを標本に残した。郵送およびインターネットのグループで返信のない個体にたいしては、電話による追跡(フォローアップ)聴取を行った (だが、これらの電話による追跡 (フォローアップ) 聴取は、ここでの説明には含まない)。合計すると 6000 人以上の回答者が調査に回答した。

Link と Mokdad (2005a,b)は、さまざまな健康状態や飲酒行動についての報告を検証し、州やさまざまなデモグラフィック変数について、3つの調査方式グループ内で報告された発生率を調整した。表 7.2 では、その2編の論文から得た主な結果を示している。電話聴取と比較すると、ウェブによるデータ収集では飲酒およびいくつかの健康状態の報告率が高かった。大部分において、これらの報告における差は各モデルのデモグラフィックをコントロールした後も継続して見られた。

TABLE 7.2. Prevalence Estimates, by Condition in BRFSS Experiments

	CATI	Web	Mail	Web vs. CATI
<b>HEALTH CONDITIONS</b>				<b>ADJUSTED ODDS RATIO</b>
Asthma	11.7	12.0	11.9	1.06
Diabetes	9.5	11.9	10.2	1.30*
High blood pressure	31.1	38.1	33.2	1.30*
BMI greater than 30	21.6	26.5	25.6	1.31*
Current smoker	22.8	16.9	17.3	0.77*
Binge drinking	14.4	12.3	21.6	1.87*
STD prevention	8.2	4.3	3.3	0.51*
Tested for HIV	38.8	30.8	32.1	0.85
<b>DRINKING BEHAVIORS</b>				<b>SIGNIFICANCE LEVEL</b>
Mean number of days in the last 30 had 1+ drinks	4.5	4.7	5.2	$p < .01$
Mean number of drinks per day on days with 1+ drinks	2.1	2.1	2.2	ns
Mean number of days had 5+ drinks	1.0	1.2	1.9	$p < .001$
Percent had a drink in last 30 days	55	52	60	ns

Note: Significance levels in the bottom panel of the table are not adjusted for demographic differences across mode groups. \* indicates that the adjusted odds ratio differs significantly from 1.0.

表 7.2 BRFSS 実験における条件別の発生率推定

	CATI	ウェブ	郵送	ウェブ対 CATI
<b>健康状態</b>				<b>調整オッズ比</b>
喘息				
糖尿病				
高血圧				
BMI30 以上				
現在喫煙者である				
過度の飲酒過度の飲酒				
性感染症予防				
HIV の検査を受けた				
<b>飲酒行動</b>				<b>有意水準</b>
過去 30 日で 1 杯以上飲酒した平均日数				
1 杯以上飲酒した日における 1 日あたりの平均飲酒回数				
5 杯以上飲酒した平均日数				
過去 30 日間の飲酒者の割合 (%)				

注：表の下のパネルにおける有意水準は、調査方式グループ間でデモグラフィック差を調整していない。\* は、調整オッズ率が 1.0 より有意に差があるものを示している。

**大学生の実験。** McCabe と同僚ら (McCabe, 2004, McCabe, Boyd, Couper, および Crawford, および d'Arcy, 2002; McCabe, Couper, Cranford, および Boyd, 2006) は、ある大規模大学の学部生の標本を使用して実験を行い、郵送とウェブでのデータ収集を比較した。ウェブ条件に割り当てられた学生には、オンラインで調査に回答するための招待状を送付した。McCabe と同僚らは、2001 年度の学生生活調査 (the 2001 Student Life Survey) の質問を改作して使用した。質問には、薬物やアルコールの使用についての項目が含まれた。学生たちは、8 種類の違法薬物の生涯使用状況および昨年度における使用状況を報告した。McCabe (2004) は、2 つの標本間のデモグラフィック差をコントロールし、16 件の比較のうち、有意差があったのは 1 件のみであったことに気づいた。男女とも、ウェブ調査票では郵送版の調査票よりもコカインの生涯使用率が有意に高かった。追跡調査 (McCabe ら 2006) では、調査票における他の 10 項目の微妙な内容の質問への回答に差は見られなかった。

**メリーランド大学同窓生の調査。** Kreuter, Presser, および Tourangeau (2008) が実施したもうひとつの追加研究には、詳細に説明するだけの価値がある。なぜなら、リサーチャーが調査回答のいくつかの項目と、大学の記録とを照合することができたからである。

Kreuter と同僚らは、調査員にメリーランド大学の同窓生に電話で接触させた。いくつかのスクリーニング質問に答えた後、回答者は学部在学中の経験にかんする質問を含んだ調査票に回答するよう依頼された。およそ 3 分の 1 の回答者が、無作為に電話での本調査に割り当てられ、次の 3 分の 1 は本調査に回答する際に IVR システムに切り替わり、残りの回答者はオンラインで質問に回答するよう指示を受けた。調査票には、優等賞を授与された、あるいはある課程で落第点をとったなど、回答者の学部在学時における学業面での明暗についての質問が含まれた。Kreuter と同僚らは、これらの質問にたいする調査回答と、その同窓生の成績証明書とを比較した。

表 7.3 では、この研究の主な結果を提示している。上のパネルは、調査から得た 9 つの微妙な内容の質問の報告率であり、これらについては記録データが入手可能である。下のパネルでは、望ましくない項目（課程で D や F をとるなど）についての偽陰性率と、望ましい項目（たとえば、GPA で高得点を得た）についての偽陽性率を示している<sup>3</sup>。ウェブ回答者は CATI の回答者とくらべ、少なくとも 1 つ以上の望ましくない項目を報告する可能性が有意に高かった。またウェブ回答者は、CATI とくらべ社会的な望ましさを誤って報告する可能性が有意に低かった。これらの調査方式の差は、望ましい学業成績の特徴より、望ましくない学業成績の特徴で大きかったが、望ましい学業成績の特徴は望ましくない学業成績の特徴とくらべ内容が微妙ではないともいえる。ウェブと電話の報告には、4 つの望ましくない質問のうち 2 つと、4 つ全ての合計において有意な差があった。

TABLE 7.3. Percentage Reporting and Misreporting Rates for Desirable and Undesirable Characteristics, by Condition in Maryland Alumni Survey

	CATI	IVR	Web
<b>PERCENT REPORTING</b>			
<b>UNDESIRABLE CHARACTERISTIC</b>			
GPA lower than 2.5	1.8	3.7	6.2
At least one D or F	42.2	44.3	50.7
Dropped a class	46.7	45.6	50.6
Received warning or placed on academic probation	10.2	13.4	13.8
<b>DESIRABLE CHARACTERISTIC</b>			
GPA higher than 3.5	23.8	23.8	24.2
Received honors	16.3	19.9	15.5
Ever donated to alumni fund	42.1	40.5	41.3
Donated in last year	44.2	41.9	40.5
Member of Alumni Association	24.8	21.5	23.6
<b>FALSE NEGATIVE RATE</b>			
GPA lower than 2.5	83.3	69.2	61.5
At least one D or F	33.0	28.3	19.9
Dropped a class	34.3	34.2	31.6
Received warning or placed on academic probation	33.3	33.3	25.0
<b>FALSE POSITIVE RATE</b>			
GPA higher than 3.5	7.4	1.9	6.0
Received honors	5.2	5.7	6.4
Ever donated to alumni fund	24.3	19.2	20.3
Donated in last year	25.6	25.9	23.3
Member of Alumni Association	10.7	10.1	8.1

Note: Data are from Kreuter, Presser, and Tourangeau (2008), Tables 6 and 9.

<sup>3</sup> 偽陰性率は、望ましくない特徴（たとえば、落第した人）があってもそれを否定する個体の割合である。同様に、偽陽性率とは望ましい特徴（GPA が 3.5 以上）をそなえていないにもかかわらず、あやまってその特徴を報告する個体の割合である。

表 7.3 メリーランド大学同窓会調査における条件別の望ましい特徴と望ましくない特徴についての報告率と誤報告率

	CATI	IVR	ウェブ
<b>望ましくない特徴の報告率</b>			
GPA が 2.5 以下 D または F を 1 回以上とった 落第した 学業不振のため警告を受けた, あるいは仮進級扱いになつた			
<b>望ましい特徴</b>			
GPA が 3.5 以上 受賞した 同窓会基金に寄付したことがある 昨年募金した 同窓会の会員である			
<b>偽陰性率</b>			
GPA が 2.5 以下 D または F を 1 回以上とった 落第した 学業不振のため警告を受けた, あるいは仮進級扱いになつた			
<b>偽陽性率</b>			
GPA が 3.5 以上 受賞した 同窓会基金に寄付したことがある 昨年募金した 同窓会の会員である			

注 : Kreuter, Presser, および Tourangeau (2008), 表 6, 表 9 からデータを得た.

**メタ分析の結果.** 潜在的にきまりの悪い情報を収集する手段としてのウェブにかんするこれらの研究にもとづいて, 私たちが到達できるような一般的な結果はあるのだろうか? 著者らは, 表 7.1 に要約した 10 種類の調査方式研究についてメタ分析を行い, その結果に全体的な傾向があるかどうかを究明した. これらの研究は著者らが以下の 3 つの基準を満たしているとみなした研究である.

- 第 1 に, ウェブによるデータ収集と (データ収集の方式を無作為に割り当てる) 他の



データ収集方式あるいは真の実験と密接に近似する疑似実験（たとえば，Chang と Krosnick, 2009）とを比較し，真の実験を報告した研究であること．回答者が回答する方式を自ら選択している研究は除外した．

- 第 2 に，あきらかに社会的望ましさの偏りが生じる可能性がある調査質問への回答を詳しく調べた研究であること．社会的に望ましい回答の心理学的測定といった非調査項目を詳しく調べている研究は除外した．
- 第 3 に，効果の大きさの標準測定に変換することが可能な(平均値や割合と言った)量的推定量を報告している研究であること．

**TABLE 7.4. Mean Effect Sizes and Standard Errors for Studies, by Study and Mode Comparison**

<b>Study</b>	<b>Sample Sizes</b>	<b>Mean Effect Size</b>	<b>Standard Error</b>
<b>WEB VS. MAIL/PAPER</b>			
Bälter et al. (2005)	Mail: 188 Web: 295	0.054	0.309
Bason (2000)	Mail: 204 Web: 115	-0.168	0.129
Bates and Cox (2008)	Mail: 73 Web: 64	-0.014	0.180
Eaton et al. (2010)	Mail: 1,729 Web: 3,498	0.070	0.012
Denscombe (2006)	Mail: 267 Web: 69	-0.256	0.120
Knapp and Kirk (2003)	Mail: 174 Web: 57	-0.077	0.119
Link and Mokdad (2005a, 2005b)	Mail: 836; 804–820 Web: 1,143; 948–1,139	0.068	0.039
McCabe (2002, 2004; McCabe et al., 2006)	Mail: 1,412 Web: 2,194	0.006	0.019
<b>WEB VS. TELEPHONE</b>			
Bason (2000)	Phone: 161 Web: 115	-0.503	0.132
Chang and Krosnick (2009)	Phone Pre-Election: 1,456 Web Pre-Election: HI: 2,313 KN: 4,914 Phone Post-Election: 1,206 Web Post-Election: HI: 1,040 KN: 3,408	0.172	0.035
Knapp and Kirk (2003)	Phone: 121 Web: 57	0.193	0.126
Kreuter et al. (2008)	Phone: 320 Web: 363	0.157	0.060
Link and Mokdad (2005a, 2005b)	Phone: 2,072; 2,066–2,070 Web: 1,143; 948–1,139	0.026	0.031
<b>WEB VS. IVR</b>			
Bason (2000)	IVR: 128 Web: 115	0.108	0.143
Kreuter et al. (2008)	IVR: 320 Web: 363	0.081	0.060

Note: Sample sizes fluctuate due to item nonresponse. The mean effect sizes are weighted averages of the log odds ratios comparing reports from Web surveys to reports obtained under other modes of data collection.

表 7.4 研究と調査方式の比較別効果の大きさ平均と研究の標準誤差

研究	標本の大きさ	効果の大きさ平均	標準誤差
ウェブ対郵送/質問紙			
Balter et al.(2005)	郵送： ウェブ：		
Bason (2000)	郵送： ウェブ：		
Bates and Cox (2008)	郵送： ウェブ：		
Eaton et al. (2010)	郵送： ウェブ：		
Denscombe (2006)	郵送： ウェブ：		
Knapp and Kirk (2003)	郵送： ウェブ：		
Link and Mokdad (2005a, 2005b)	郵送： ウェブ：		
McCabe (2002, 2004; McCabe et al., 2006)	郵送： ウェブ：		
ウェブ対電話			
Bason (2000)	電話： ウェブ：		
Chang and Krosnick (2009)	電話選挙前： ウェブ選挙前： HI： KN： 電話/選挙後： ウェブ/選挙後 HI： KN：		
Knapp and Kirk (2003)	電話： ウェブ：		
Kreuter et al. (2008)	電話： ウェブ：		
Link and Mokdad (2005a, 2005b)	電話： ウェブ：		

ウェブ対 IVR			
Bason (2000)	IVR: ウェブ :		
Kreuter et al. (2008)	IVR: ウェブ :		

注：標本の大きさは項目無回答によって増減する。効果の大きさ平均は、ウェブ調査から得られた報告と他のデータ収集方式によって得られた報告とを比較したログオッズ率の加重平均である。

総合すると、これらの研究では 223 件の調査方式比較を報告しており、そのうちの 160 件にはウェブと質問紙型調査票が関与した。表 7.4 では、それぞれの研究の効果の大きさの平均（つまり、各研究で提示される推定値の平均ログオッズ率）と、それに関連するこれらの研究レベルでの平均となる標準誤差とを示している。肯定的な効果の大きさは、他方式条件下の回答者とくらべ、ウェブ条件の回答者が微妙な特徴あるいは行動を報告した割合が高かったことを示す。たとえば、Balter と同僚らによる研究から得られた効果の大きさ平均 0.054 は、平均すると、その研究におけるウェブ回答者が郵送調査の回答者よりも微妙な情報を提供する可能性が高いことを示している。研究の平均値は加重平均であり、それぞれの効果の大きさ推定は標準誤差の 2 乗の逆数を加重したものである(Lipsey と Wilson, 2001)。

著者らは、これら 10 件の研究から得たデータを分析し、効果の大きさ推定を研究別のクラスター（層）として扱った。メタ分析の結果は、2 つの主な結論を支持した。第 1 に、調査員方式の電話調査とくらべ、ウェブによるデータ収集はより多くの微妙な内容の情報を引き出すようである (Chang と Krosnick, 2009; Kreuter, Presser, および Tourangeau, 2008; Link と Mokdad, 2005a, 2005b; Bason, 2000 もあきらかな例外として参照)。2 つの方式を 6 件の研究で比較したところ、全体的な効果の大きさは 0.088 だったが、その大きさはゼロ( $t=1.69$ ,  $df=7$ )とは有意に異ならなかった。これは一部には Bason が報告した大逆転が原因となっている。Bason の研究を脱落させれば、電話とウェブを比較した効果の大きさ平均は、.105 まで上昇する（標準誤差は.052）。このグループに属する研究の 1 件以上 (Kreuter et al., 2008) で、報告の増加が精度の上昇を意味することを示している（上述の表 7.3 の 2 つのパネルの下側を参照）。これらの結果は、潜在的にきまりの悪い情報を引き出す際には自記式が役立つという先の分析と一致している (Tourangeau と Yan, 2007)。

第 2 に、質問紙型の自記式にたいしオンラインでの自記式による報告は、ほんのわずかに優位であるにすぎないようである。ウェブと質問紙の比較における平均効果の大きさは全体で 0.030、標準誤差は 0.023 である。それに先立って Tourangeau と Yan(2007) が行ったメタ分析においても、微妙な内容の情報の報告増加について、コンピュータ化された調査票（必ずしもオンライン調査票とは限らない）が質問紙型調査票より優れていた利点は

有意ではなかった（コンピュータによる実施では、質問の種類によっては社会的に望ましい回答が増減する可能性があることに気づいた、Richman, Kiesler, Weisband, および Drasgow, 1999 も参照してほしい。ここで論じた研究で用いられたような、微妙な行動についての質問項目では、コンピュータ化によって社会的に望ましい回答が増加するようであった）。

**データ収集設定の影響力。** 大学やその他の学校母集団は、ウェブ調査にとって魅力的なターゲットである。なぜなら、高校生や大学生ではインターネットへのアクセスが多いことから、これらの母集団ではウェブ調査の回答率が比較的高くなる可能性が高いからである。実際、McCabe が学生標本を用いて行った実験では、ウェブ調査の回答率が郵送調査の回答率よりも有意に高く (McCabe, 2004; McCabe et al., 2006), 通常とは逆の傾向となった (Lozar Manfreda, Bosnjak, Berzelak, Haas, および Vehovar, 2008; Shih と Fan, 2008; 第 3 章参照)。表 7.1 に要約した他のいくつかの研究でも、微妙な内容の情報を収集する上でウェブの効果をテストするために学生標本を使用している。

学生母集団について自然に生じてくる 1 つの問題は、そのデータを学校内で収集するか（たとえば、教室やコンピュータ・ラボなど）、それとも自宅や学校外のどこかで収集するかである。リサーチャーらは、若者たちから違法薬物の使用状況についてのデータを収集するには、自宅よりも学校の方がよいと論じてきた (Fendrich と Johnson, 2001; Fowler と Stringfellow, 2001)。いくつかの研究では、データ収集の設定の効果について実験的に詳しく調べている。Brener, Easton, Kann, Grunbaum, Gross, Kyle, および Ross (2006) は、危険行動についての質問を、コンピュータまたは質問紙による自記式で、教室内および学校外—通常は回答者の自宅—で実施することで比較を行った。彼らが調べた 55 項目のうち 30 項目で、設定に有意な効果がみとめられた。すべての場合において、危険度の高い行動は、自宅よりむしろ学校で回答したときに多く報告された（このことは、学校内でのデータ収集は自宅でのデータ収集よりも微妙な内容の行動の報告が多いという Fendrich と Johnson, 2001, および Fowler と Stringfellow, 2001 の推測を確認することにつながった）。そのうち 7 つの項目では、コンピュータで実施した場合、質問紙の場合よりも報告度が有意に高くなり、55 項目中 5 項目で、設定と方式が互いに影響をおよぼした。ほとんどの場合において、学生が調査票に自宅で回答した場合よりも、学校内で回答した場合にコンピュータと質問紙の差は大きくなるようであった。Beebe と同僚ら (Beebe, Harrison, McCrae, Anderson, および Fulkerson, 1998) は、学校で行われたある調査において、コンピュータと質問紙の間に全体としてはほとんど差が見られなかったが、学生が互いに近接して座席につき、恐らくはプライバシーが少ない場合では、質問紙方式よりもコンピュータ方式で実施した場合に報告が減少した。

Brener と共著者、および Beebe と同僚らは両者ともコンピュータ方式と質問紙方式に目を向けたが、ウェブを経由しないコンピュータ方式であった。Bates と Cox (2008) は、ウェブ方式と質問紙方式で実施する微妙な内容の質問の聴取を 3 つの異なる設定で詳しく

調べた。3つの設定とは、教室におけるグループ実施、専用オフィスにおける個別実施、そして回答者自身が選んだ場所での個別実施である。5項目中2項目で、設定に有意な効果が見られた（他の設定よりも、自分で選んだ場所で個別に質問に回答した場合は、より多くの回答者が微妙な行動について報告した）が、この変数はデータ収集の方法とは関係していなかった。

Denniston と同僚ら(Denniston et al., 2010)は、回答者が教室で行われるウェブ方式のデータ収集は、同じ設定における質問紙方式のデータ収集よりも機密性が低いと思っているが、それでもなおウェブ方式ではより多くの微妙な情報を報告した、と報告している。

そこで全体としては、高校生から微妙な内容の情報を収集する場合には、自宅より学校の方が場として適している可能性があるが、設定としては自記式の方法とは関係がないようである。学生母集団では、質問をウェブで実施するものを含め、質問紙方式よりもコンピュータによる自記式で実施する方が微妙な情報の報告度を高めるようである。

**インターフェイスに人間らしさをもたせる。**第6章では、ウェブ調査においてインターフェイスに人間らしさをもたせることの賛否について論じた。たとえば、「バーチャル」の調査員に質問を実施させるといったことである。ここではそのような人間らしさをウェブ調査に加えることで、具体的には社会的望ましさの偏りや調査員効果の低減といった、自記式によるデータ収集がもつ利点のいくつかを弱める可能性があるかどうかについての知見を手短に見直したい。

Naas と同僚ら、および Kiesler と同僚らによる研究は、最小限の人間らしさの特性（インターフェイスで使用される音声など）さえもが、実際の動作主体が生じさせるのと同じ反応をユーザーから誘発する可能性を提起した（男女をめぐる固定観念など。例として、Nass, Moon, および Green, 1997; Sproull, Subramini, Kiesler, Walker, および Waters, 1996 も参照）。Tourangeau, Couper, および Steiger (2003)は、人間らしさの手がかりを加えることが、回答者が健康行動、違法薬物の使用、および他の微妙な話題についての情報を開示しようとする積極的意志にどう影響をおよぼすかを見極めるために、ウェブ・インターフェイスの特性を系統的に変えて行った一連のウェブ実験を報告した。これらの研究では、男性または女性のリサーチャーの写真をウェブ調査に含めること、回答者の回答にもとづいて個々に対応した双方向的なフィードバックを提供すること、そして個人的な言葉を使用すること（たとえば、回答者を名前と呼ぶなど）の効果を詳しく調べた。彼らの研究では、調査にそのような人間らしさの特性が含まれると、回答者が微妙な行動についての質問にたいする回答を変えるということを示した証拠はほとんど見つからなかった。彼らが見つけた証拠は、リサーチャーの写真が男女の性別による役割にかんする一連の質問にたいする回答に影響をおよぼしたことである。ウェブ調査で女性リサーチャーの写真を表示したときには、男性リサーチャーの写真を表示したときとくらべ、回答者はより女性を支持する回答をもたらした。その効果は小さいものであったが統計的には有意であり、同様に実際の男女調査員で行った場合の結果と同じ傾向であった（Kane と Macaulay,

1993) .

より最近の研究では、バーチャル調査員をコンピュータ支援の調査に含めることが、Tourangeau, Couper,および Steiger が発見したよりも大きな効果を回答にもたらすことを示唆している。Krysan と Couper (2003)は、実際の調査員と、デジタルビデオで録画した同じ調査員が質問を読み上げたものとを比較した。また彼らの実験では、調査員の人種を系統的に変えた。実験は比較的小規模なもので(回答者数は合計 160 人)、研究結果のほとんどは統計的に有意ではなかった。それでもなお、黒人の回答者では、調査員の人種の有意な効果が見られた(たとえば、黒人回答者は、白人調査員より黒人調査員にたいして白人にたいする否定的な態度の報告が多かった)。これらの効果は、実際の調査員とバーチャル調査員では同様の傾向を示しており、これは Schuman と Converse (1971)の、実際の調査員との面接における黒人回答者についての有名な研究結果を再現する傾向があった。Fuchs (2009)は、男女のバーチャル調査員が質問を読み上げるビデオ録画と、テキストのみという 2 つのバーチャル調査員の条件を比較した。(Krysan と Couper の研究では、オンライン調査ではなくラップトップで行うコンピュータ支援の調査を用いた。) Fuchs の研究で行われた 4 つの比較のうち 3 つで、女性回答者は微妙な内容の性的な情報(性感染症にかかったことがあるかどうか、といったこと)を男性のバーチャル調査員より、女性のバーチャル調査員に打ち明ける可能性が有意に高かった。男性回答者ではこのパターンはそれほどあきらかではなく、4 つの比較で調査員の性別に有意差があったのは 1 項目だけだった。最後に、Conrad, Schober, および同僚ら(Conrad, Schober, and Nielsen, 2011, and Lind, Schober, Conrad, and Reichert, 2011)は、人種問題への意識にかんする質問(Conrad et al., 2011)と、質問を行う際に顔の動きを伴う(訳メモ: facial movement/顔の動き, 顔面運動→表情の変化と言ってよいか)バーチャル調査員が微妙な行動に関して質問する際の、社会的に望ましい回答の増加にかんして、バーチャル調査員の人種の効果を示した(Lind et al., 2011)。バーチャル調査員を用いるウェブ調査は、面接調査と同様の結果が生じた項目もあれば、面接調査とオーディオ CASI による面接調査の中間的な結果が生じた項目もあった(Lind et al., 2011)。

総合すると、バーチャル調査員が実物そっくりになってきていることで、最近の結果ほど、バーチャル調査員の効果は実物の調査員とより似たものになり、調査員の人種または性別が意識についての質問への回答におよぼす効果や、微妙な行動にかんする質問への社会的に望ましい回答の増加を示唆している。

**要約.** ウェブ調査による実施は、初期の自記式形態がもつ長所を共有しているようである。著者らのメタ分析では、ウェブ方式は、どちらかといえば微妙な情報を入手するには、質問紙方式よりも優れている。Kreuter, Presser, および Tourangeau (2008)による実験は、ウェブ方式のデータ収集が、調査員が実施する電話調査とくらべると、報告の精度を改善することを示している(少なくとも社会的に望ましくない質問にかんしては)。ウェブによる情報収集は学生母集団ではとくに有効であるかもしれない。ウェブ調査においてインタ

ーフェイスに人間らしさをもたせる実験—とくにバーチャル調査員の使用を通じて—は、微妙な行動についての報告の減少、人種や性別にかんする意識についての質問における調査員の人種、性別の効果といった、バーチャル調査員が実物の調査員と同じ欠点を示す可能性があることを示唆している。それでもなお、第 6 章で注目したようにバーチャル調査員を用いることには、回答者の関与度が高まるといった見返りとなる利点があるだろう。微妙な質問や人種、性別、あるいは他のあきらかな調査員の特徴に関連した質問では、標準的なインターフェイスが人間らしさの手がかりが組み込まれたインターフェイスよりもふさわしい可能性が高い。

### 7.3 ウェブ調査と認知的負担

質問の提示を主に視覚にたよる他の自記式手法のように、ウェブ調査は調査員方式のデータ収集、とくに聴覚的提示に依存する方式（電話聴取など）とくらべて認知的負担が低くなる可能性がある。読み書き能力の低い回答者ではもちろん、この関係性は真逆となる。つまり、電話調査とくらべ、ウェブ調査では負担が高まる可能性があるのだ。ウェブ調査では、回答者は自分の好きなときに、自分のペースで質問に回答することができる。さらに、質問を読み返すことも容易にできる。視線追跡調査は、回答者がウェブ調査で質問を読み返すためによく後戻りしていることを示している。これとは対照的に、電話調査では、回答者が質問を繰り返し読み上げるよう調査員に頼むことはめったにない。ウェブ調査が電話調査とくらべて認知的負担を低減する可能性があり、したがってより良いデータをもたらすということを示した 2 つの線の証拠がある。

**知識についての質問。** 回答者の知識を評価するためのウェブ調査と電話聴取で比較を行った研究が 2 つある。1 つ目の研究では、Fricker, Galesic, Tourangeau, および Yan (2005) が米国国立科学財団 (the National Science Foundation) が定期的に米国民の科学的リテラシーを測定するために使用している、一連の基本的科学知識を問う質問を用いて実験を行った (Miller, 1998)。Fricker と同僚らは、RDD で選出した成人の全国標本にたいし、これらの質問を実施した。回答者は数問のスクリーニング質問に回答した後、インターネットにアクセスできる標本が、電話あるいはオンラインでの本調査（科学的知識を問う質問を含む）への回答に無作為に割り当てられた。電話条件に割り当てられた個体は、ウェブに割り当てられた個体とくらべ調査に回答する可能性が非常に高かった（ほぼ 98% が電話を切らずに知識についての質問に回答したのに対し、ウェブ条件では 52% だった）にもかかわらず、2 つの回答者グループではデモグラフィック特性や教育的背景に有意な差はなかった。オンライン回答者は、電話で回答した回答者よりも知識得点が高く、平均正答率は 70% だったのに対し、電話回答者の正答率は 64% だった。方式による差は、真偽問題よりも自由回答質問でより大きく、このことはより認知的に負担が大きくなる質問ではウェブの支援がより大きかったことを示唆している。また、Fricker と共著者は、ウェブ回答者は電話回答者よりも質問の回答に時間をかけており、その差のほとんどが、自由回答質問に



オンラインで回答する際に費やした時間が長かったことを反映していた。

Strabac と Aalberg (2011)は、Fricker と同僚らと同様の研究結果を報告している。彼らは、米国とノルウェーにおいて、ウェブパネルの会員と電話調査の回答者にたいし、政治的知識を問う 6 つの質問を行った。ギャラップ社が両国で電話調査を実施し、研究のウェブ部門では、二国間で異なるウェブパネルを使用した。標本構成員を無作為に電話またはウェブによるデータ収集に割り当てるのではなく、既存のウェブパネルを使用したことから、この研究は真の実験とは言えない。回答者は 3 人の政治的指導者(例: Robert Mugabe) と 3 つの国際組織(例: OPEC) を特定するよう求められた。12 件の非加重比較のうち 5 件では正答率の割合に有意差があり、その 5 件のすべてにおいてウェブの方が好まれた。回答者の年齢、性別、および教育の分布を調整するためデータに加重をした場合、結果はいっそうあきらかとなった。12 件の比較中 11 件で、正しい答えを出した回答者の割合はウェブ回答者でより高く、これらの差の 6 件は統計的に有意であった。

これら両方の研究は、ウェブ回答者が調査に回答する際に回答をオンラインで調べる可能性を除外することを試みた。Strabac と Aalberg (2011)は、各質問で回答を提出する時間を 30 秒しかウェブ回答者にあたえなかった。Fricker と同僚らは、ウェブ回答者の利点が、回答をオンラインで容易に調べることができる質問より、オンラインで調べることが困難な質問(コントロール・グループを設定した研究は、コントロール・グループを設定しない研究よりも優れているのはなぜかを問う質問など)で大きかったことを報告している。

**尺度回答。**ウェブを介することで認知的負担が軽減されること、あるいは質問をより思慮深く処理できることから生じるもう 1 つの潜在的な重要性は、回答の信頼性あるいは妥当性の増大であろう。同一の質問をウェブ方式と質問紙方式で実施、比較した研究がいくつかある。たとえば、Ritter, Lorig, Laurent, および Matthews (2004) は、これらの実験の 1 つを説明している。彼らの研究には、インターネットまたは郵送で 16 項目の健康に関する質問(そのいくつかは 1 つの質問のみを含む)に回答した 397 名の回答者が関与した。また、30 名のウェブ回答者から成る小規模の副標本も 2 回目の質問に回答した。質問紙方式(とくに郵送方式)は、おそらくはウェブ方式と同様に、自分のペースで回答したり、質問文を読み直したりする機会を提供している。全体的には、Ritter と同僚らは 16 の尺度の手法および尺度の信頼性(クローンバッハの  $\alpha$  による推定)にかんして有意差を発見しなかった。これらの結果は、ウェブによるデータ収集と、同じ複数尺度を質問紙で実施した場合とを比較した、より大きな研究母体を代表しているようである。

話しが変わってくるのは、データ収集を、ウェブで行う場合と電話で行う場合とで比較する場合である。Chang と Krosnick (2009)は、無作為の測定誤差が電話調査よりも 2 つのウェブパネルで低かったことを示した彼らの研究から、いくつかの比較を報告している。3 つの標本全体で、項目についての信頼性や妥当性の推定が最も低かったのは、電話標本であった。たとえば、2000 年の大統領選における票の選択は、質問を電話で行った場合より

オンラインで実施した時の方がより一連の予測因子と密接に関連していた。

電話回答者は郵送やウェブの回答者とくらべ最も極端な回答選択肢を選択する可能性が高いと論じているリサーチャーもいる(例: Christian, Dillman, および Smyth, 2008; Dillman と Tarnai, 1991). 最も極端な回答を選択するという傾向を減らすことで、ウェブによる実施は、電話聴取とくらべて尺度回答の妥当性を高めることができる可能性がある。Ye, Fulton, および Tourangeau (2011)によるメタ分析では、調査方式によって回答分布に系統的な差があるが、尺度の肯定的な端ではその差がより顕著であるということに気づいた。電話回答者は、郵送やウェブの回答者よりも肯定的な最極端の回答を選択する可能性が高いことに彼らは気づいたのである。Ye と同僚らは、主要変数は提示の視覚的チャンネルではなく、むしろ調査員の存在であると論じている。最も肯定的な選択肢が選ばれたのは、電話で聴取を受けた回答者よりも面接を受けた回答者で多く、また IVR による面接では調査員方式の電話調査の場合よりも選ばれることが少なかった。Ye と共著者らが再検討を行った調査のほとんどが顧客満足度の評価を収集していた。

ウェブ調査では、尺度の信頼性あるいは妥当性は回答を入力する方法(たとえば、視覚的アナログ尺度にたいし、ラジオボタンなど)または画面上の回答選択肢の視覚的配置(たとえば、垂直的配置または水平的配置)に影響される。第 6 章ではすでにスライダー・バーや他の視覚的アナログ尺度の使用について論じており、それらを推奨する証拠がほとんどないことに気づいた。第 5 章では、意識尺度の回答選択肢の間隔取りや配置が提起した問題を詳しく調べた。そこで再検討された研究から浮かび上がった、少なくとも 1 つのはっきりとした結果は、尺度選択肢の見た目はその根底をなす次元の構造と並行するべきであるということである。たとえば、尺度の視覚的中間点は、その次元の概念的な中間点と一致するべきで、尺度点が等間隔を表すよう意図されているときには、選択肢は等間隔に配置されるべきである(Tourangeau, Couper, および Conrad, 2004)。

**ウェブと面接によるデータ収集。**ウェブを含んだ調査方式比較のほとんどが、インターネットによるデータ収集と郵送あるいは電話調査との比較を行っている(たとえば、表 7.4 を参照してほしい)。Heerwegh と Loosveldt (2008)による研究は、1 つの例外である。彼らはウェブ調査と面接によるデータ収集を比較する無作為化実験を行った。Heerwegh と Loosveldt は、報告した結果はまちまちである。彼らは、ウェブ調査では面接のときとくらべて非差別化が少なかったが、ウェブ回答者は面接を受けた人たちよりも「わからない」や中間的な回答選択肢を選択することが多いことに気づいた。これらは両方とも、調査では最低限の要求を満たせばそれによしとする潜在的な兆候である。ウェブ調査もまた、平均すると面接とくらべてより早い時間で回答が行われていた。

いくつかの追加実験では、発生可能性評価(contingent valuation:CV)データを収集するウェブ調査と面接調査とで比較を行っている。CV 研究では通常、回答者にある環境財についての詳細情報を提示し、そして回答者がそれを保護するためにいくら払うつもりがあるかを聞き出す。Lindhjem と Navrud (2011a)は、面接とウェブの CV 調査を比較した 6 つの

研究を見直したが、2つのデータ収集方式間に一貫した差を見出さなかった (Lindhjem と Navrud, 2011b; Marta-Pedroso, Freitas,および Domingos, 2007; および Nielsen,2011 も参照のこと)。一般的にこれらの研究を通じて、支払い意欲の平均はデータ収集方式で異ならず、ウェブ調査でデータの品質が低下したことを示すあきらかな証拠も得られなかった。

#### 7.4 この章のまとめ

ウェブ調査には、他のデータ収集方法とくらべ測定誤差を減らすことのできるいくつかの重要な特性がある。

- 1) 質問は調査員よりむしろコンピュータによって実施する。これは回答における社会的に望ましきの偏りを減らし、調査員効果を取り除く可能性のある特性である。
- 2) 質問を提示する主なチャンネルは視覚的であり、写真、ビデオクリップなどを組み込むことができる。
- 3) 調査が双方向性になるので、回答者にさまざまな種類のヘルプを出し、また適切な質問に回答者を導くことができる。
- 4) 回答者が調査質問のペースを管理し、質問を容易に読み返すことができる。こうした特性により、回答の際の認知的負担を減らすことができる。

ウェブ調査のこれらの特性はそれぞれ、回答者がいかに容易に、そしていかに正確に質問に回答できるかに影響する。第5章と第6章では、ウェブ調査の視覚的提示および双方向性の機能がもつインパクトを詳細に論じた。本章では、ウェブ調査の1番目と4番目の特徴一つまり、調査員の除外と調査員が認知的負担におよぼす効果—に焦点を当てた。もちろん、これらの4つの特徴に加え、ウェブ調査は自動化されており、自動データ収集がもつ通常の利点はすべて提供している (たとえば、事前に読み込まれた情報や前の質問への回答にもとづいて質問を調整する能力、自動スキップや自動ルーティング、回答が事前に決められた範囲内におさまっているかを確認する編集チェック、質問や回答選択肢の並び順の無作為化など)。これらの特性は、ウェブ調査が高度に複雑な調査票を実施することができることを意味している。

回答者から微妙な情報を収集するためとしては、ウェブは、初期の自記式手法の長所を共有しているようである。表7.4に要約したメタ分析は、きまりの悪い情報を引き出す際に、ウェブ調査が少なくとも郵送あるいは他の質問紙による自記式と同程度であることを示している。またそこでの結果は、薬物使用、性的行動および他の潜在的に微妙な話題について情報を収集する際には、ウェブ調査が調査員方式の電話聴取より優れていることも示唆している。ウェブによるデータ収集のこうした利点は、とくに若い回答者で大きくなる可能性があり、質問に回答する設定が落ち着けるものではなかった場合や、バーチャル調査員がインターフェイスに使用されている、あるいは他の人間らしきの手がかりが組み込ま

れている場合にはこうした利点が損なわれる可能性があった。

回答者がウェブ調査のタイミングやペースを管理できること、そして回答者が質問を容易に読み返すことができることから、彼らは知識についての質問により正確な回答を提供し、回答者の意識や他の心理的特性を評価する質問群により信頼度の高い妥当な回答をもたらしている。これらの利点は、ウェブ調査が誤解を招く方法で回答選択肢を配置した場合、相殺される可能性がある。そして読解が困難である、あるいはコンピュータ技能が低い母集団にかんしては、ウェブ調査のいかなる認知的利点も消え去る、あるいは逆転する可能性が高い。