

個別面接聴取法における Item Count 法の諸問題と 実用化可能性

土屋 隆裕¹・平井 洋子²・小野 滋³

(受付 2007 年 1 月 16 日; 改訂 2007 年 3 月 8 日)

要 旨

調査項目が社会的に望ましくない内容である場合、それを直接質問しても全ての回答者から正直な回答を得られるとは期待できない。間接質問法の一つである Item Count 法では、直接質問をする代わりに「当てはまる項目数」のみを尋ねることで回答者個人の情報を秘匿し、より正直な回答を促す。それと同時に、二つの等質なサンプル間の項目数の差を利用することで、目的とするキー項目の推定値を理論的には得ることができる。本稿では Item Count 法の特性と実用化可能性を探るため、直接質問法との比較実験調査を行い、「万引き」を含む四つのキー項目の経験率の推定を試みた。調査は個別面接聴取法により実施し、回答者の Item Count 法の受け止め方なども調べた。その結果 Item Count 法では主に文脈効果などが二群の等質性を崩し、推定値を不安定にする可能性があることが示唆された。

キーワード：間接質問法、プライバシー、万引き。

1. Item Count 法とは

1.1 間接質問法と本稿の目的

2005 年の個人情報の保護に関する法律の全面施行以来、プライバシーに対する人々の意識は非常に高い。このプライバシー意識の高さが、社会調査を含む統計調査における近年の回収率低下を招いた要因の一つとも言われている。また調査では普通、目的とする項目（以下、キー項目）を回答者に直接質問する直接質問法 (Direct Questioning Method, 以下 DQ 法) が用いられる。しかし調査内容が社会的に望ましくない行為や薬物使用といった違法行為などの場合、DQ 法では正直な回答を期待できず、推定値が過小となることは想像に難くない。性行動など調査内容・対象によっては、そもそも直接質問することをすら、はばかる場合もある。

以上のような、回答者によっては正確な情報提供をためらう場面では、より正直な回答を引き出すため自記式調査法を用いる、回答用紙を密封してもらうなどの工夫が必要である (Barton, 1958; Sudman and Bradburn, 1982; Lee, 1993)。また間接質問法の利用も選択肢の一つである。間接質問法とは、回答法を工夫し、たとえ回答を見てもキー項目に関する各回答者の状態は分からないよう秘匿しつつ、目的とする推定値を得る方法である。秘匿によって正直な回答を期待できるだけでなく、調査者と回答者の双方にとってストレスのより少ない調査実施が可能となる。

¹ 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7

² 首都大学東京 都市教養学部：〒192-0397 東京都八王子市南大沢 1-1

³ シノベイト株式会社：〒105-0001 東京都港区虎ノ門 3-18-19 虎ノ門マリビル

本稿では間接質問法のうち Item Count 法 (Droitcour et al., 1991) を取り上げ、直接質問法との比較実験調査を通してその特性を検討する。Item Count 法 (以下、IC 法) は、文献によっては Unmatched Count 法 (Dalton et al., 1997) あるいは List 法 (Sniderman and Grob, 1996) とも呼ばれる。IC 法の原理は次節で詳述するが、回答法が単純で理解しやすいという意味で、実施が最も容易な間接質問法の一つである。しかしその適用結果は必ずしも常に成功しているとは言い難い。例えば Biemer and Wright (2004) は 7 万人を対象に IC 法による調査を行い、コカイン経験率の推定を試みている。IC 法の推定値は DQ 法より小さく、精度が向上したとは考えにくい。我が国では IC 法の適用例はほとんど見当たらず、唯一、Tsuchiya et al. (2007) が「万引き」経験率を調べている。DQ 法より 10 ポイント近く大きな推定値を得ることに成功しているものの、モニター形式の Web 調査の結果であり、サンプルが代表性に乏しい (Couper, 2000) ことは否めない。Tsuchiya et al. (2007) では他の調査モードにおける有効性は未確認であり、IC 法の基礎研究は国内外を問わず未だ不足している。

そこで本稿では、個別面接聴取法による DQ 法と IC 法の比較実験調査を実施する。面接法の利点は、IC 法に対する回答者の反応を調査員が直接確認できることである。回答者が IC 法の回答方法を誤解した場合、それを正すことも可能である。一方でデリケートな内容を扱う調査では、面接法など調査員が介在する他記式は不向きとされる (Tourangeau and Smith, 1996; Jobe et al., 1997)。そのため IC 法による研究・調査はこれまで、例えば集合調査や ACASI (Couper et al., 1998) など自記式による実施が多い。他記式による IC 法の実施例は、反黒人感情を調べた Kuklinski et al. (1997a) や Kuklinski et al. (1997b)、反ユダヤ人感情を調べた Kane et al. (2004) など非常に限られる。他記式によるこれらの調査では DQ 法との比較が行われず、IC 法の結果の妥当性は不明のままである。

本稿の比較実験調査の結果をここで先回りして述べれば、IC 法の推定値は DQ 法に比べ不安定であり、その意味では面接法による IC 法の適用は失敗であったと言わざるを得ない。しかしその要因を探ることで、IC 法を成功させるための条件や IC 法の特性・限界をより明確に見極めることができよう。

1.2 Item Count 法

この節では IC 法の原理と IC 法に関する従来の研究を概観する。調査の目的は、大きさ N の母集団 U における、キー項目に当てはまる集団 T (例えばコカインの使用経験がある人、万引きの経験がある人) の割合を知ることである。すなわち母集団の第 i 要素が集団 T に属するか否かを表す指示変数を

$$(1.1) \quad Y_i = \begin{cases} 1, & i \in T \\ 0, & i \in U - T \end{cases}$$

とすると、調査目的は

$$(1.2) \quad \bar{Y} = \frac{1}{N} \sum_U Y_i$$

を知ることである。DQ 法では調査対象にキー項目の値 Y_i を直接尋ねる。

IC 法の手続きは以下のとおりである。母集団 U から大きさ n_A と n_B の二つの等質なサンプル A 群と B 群を独立に抽出する。A 群にはキー項目以外の G 個の非キー項目から成るリスト (以下、短リスト) を提示し、その中で当てはまる項目の数 X のみを回答してもらう。B 群には短リストにキー項目を加えた $G+1$ 個の項目から成るリスト (以下、長リスト) を提示し、同様に当てはまる数 Z のみを回答してもらう。 $Z_i = X_i + Y_i$ であることから、

$$(1.3) \quad \bar{Y} = \frac{1}{N} \sum_U Y_i = \frac{1}{N} \sum_U (X_i + Y_i) - \frac{1}{N} \sum_U X_i = \bar{Z} - \bar{X}$$

であるので、右辺をその推定値で置き換えれば B 群による推定値 \hat{Y}_B が得られる。二つのサンプルは独立であるので、 \hat{Y}_B の分散は $\hat{V}(\hat{Y}_B) = \hat{V}(\hat{Z}) + \hat{V}(\hat{X})$ である。さらに二群の役割を入れ替え、B 群には別の非キー項目から成る短リストの中の当てはまる項目数 V 、A 群にはそれにキー項目を加えた長リストの中の数 W を回答してもらうことで、両サンプルによる推定値 \hat{Y}_{A+B} が得られる。

$$(1.4) \quad \hat{Y}_{A+B} = \frac{1}{n} \left\{ n_B(\hat{Z} - \hat{X}) + n_A(\hat{W} - \hat{V}) \right\} = \frac{1}{n} (n_B \hat{Y}_B + n_A \hat{Y}_A)$$

IC 法では項目リスト中の「当てはまる項目数」のみを答えてもらうため、回答が「0 個」あるいは「全て」でない限り、長リスト提示群の各回答者がキー項目に当てはまるか否かを調査者が直接知ることはできない。他の間接質問法としては Warner (1965) が提案した Randomized Response 法 (以下、RR 法) がよく知られ、多くの変形手法が存在する (Greenberg et al., 1969; Takahashi and Sakasegawa, 1977; 逆瀬川 他, 1977; O'Brien, 1979; Chaudhuri and Mukerjee, 1988)。さらに Aggregated Response 法 (Warner, 1971; Boruch and Cecil, 1979) や Block Total Response 法 (Raghavarao and Federer, 1979) といった間接質問法も提案されているが、回答法の容易さの点で IC 法は他の間接質問法とは一線を画す。特に RR 法の多くの変形手法では、回答する質問項目を回答者自身が選択する必要があり、その作業は複雑で難解になりがちである。RR 法と IC 法に対する回答者の反応を比較した Hubbard et al. (1989) によれば、RR 法は回答者の「笑いを誘い、不真面目な調査という印象を与えるおそれもある。

反面、IC 法を実用化する上で解決すべき課題は多い。まず、社会的望ましさが影響するキー項目に限ってみても、IC 法の結果が一貫して DQ 法に勝っているわけではない。DQ 法よりも妥当と思われる推定値が得られた例としては、同性愛者に対する暴力や誹謗中傷の経験の有無を調べた Rayburn et al. (2003a) や犯罪被害者となった経験の有無を調べた Rayburn et al. (2003b)、職場における非倫理的行為の経験を調べた Dalton et al. (1994) や Wimbush and Dalton (1997)、性行動などを調べた LaBrie and Earleywine (2000) が挙げられる。ただしいずれもサンプルサイズが数十から数百と小さい上、IC 法の推定値 \hat{Y} の分散を $\hat{V}(\hat{Y}) = n^{-1} \hat{Y}(1 - \hat{Y})$ などで推定していると思われ、統計的な有意差には疑念が残る。さらに IC 法を成功させた要因は必ずしも明らかではなく、偶然であった可能性は捨てきれない。

一方、静注薬物乱用者の割合の推定を試みた Droitcour et al. (1991) では、DQ 法の 3.9% に対し IC 法の推定値は 0.2% にとどまる。先述の Biemer and Wright (2004) では、12 から 17 歳のコカイン経験率が DQ 法の 1.5% に対し IC 法では 0.73%、18 歳以上では DQ 法の 1.9% に対し -0.08% といずれも IC 法の方が小さく、IC 法によって推定値が改善したとは考えにくい。Tsuchiya et al. (2007) は IC 法の過小推定の原因の一つが次のような過少回答傾向という非標本誤差にある可能性を指摘している。一般に IC 法での「当てはまる項目数」の回答は、DQ 法による「当てはまる」項目の合計数より過少となる傾向がある。過少回答の傾向は DQ 法で正直に回答できるリストでも見出され、リストの項目数 L が多いほど大きい。(1.3) 式において短リストの過少回答の大きさの期待値を $\delta_X (> 0)$ 、長リストのそれを $\delta_Z (> \delta_X)$ とすると、

$$(1.5) \quad (E(\hat{Z}) - \delta_Z) - (E(\hat{X}) - \delta_X) = E(\hat{Y}) - (\delta_Z - \delta_X) < E(\hat{Y})$$

となり過小推定となる。Biemer and Brown (2005) は同一の回答者に対して IC 法による回答、当該リストの各項目についての DQ 法による回答、キー項目に対する DQ 法の回答を全て求め、回答傾向を補正した推定値の改善を試みている。

非標本誤差に加え、IC 法の推定値は標本誤差も大きい。IC 法と DQ 法の比較を行った Droitcour et al. (1991) や Biemer and Brown (2005)、Tsuchiya et al. (2007) では、IC 法の方がサンプルが大きい場合でも標準誤差は DQ 法の方が小さい。属性別などの推定のためサンプルの一部だけ

を用いると、標本誤差はさらに拡大することとなる。そこで Tsuchiya (2005a, 2005b) は一定の仮定の下で標準誤差のより小さな属性別の推定法を提案している。

回答者に対し調査の目的・主旨をどのように、どの程度説明すべきかという点も課題の一つである。DQ法より大きなIC法の推定値を得た先述の Dalton et al. (1994) や Wimbush and Dalton (1997) では、回答者に対しIC法の原理を詳細に実演してみせた上で調査を実施している。Ahart and Sackett (2004) は実演の影響を受けてIC法の推定値が大きくなった可能性を指摘し、詳しい説明をせずに調査を実施したところ、必ずしもDQ法より大きな推定値は得られなかったとしている。

非キー項目の内容に関しては主に二通りの考え方があり得る。Dalton et al. (1994) らは、キー項目とは無関係の無難な当たり障りのない内容を用いている。これに対し Biemer and Wright (2004) らは、全く無関係な内容よりも、キー項目と関連するような社会的に望ましくない内容の方が回答者が疑義を抱くことが少ないとしている。

2. 調査デザイン

2.1 比較実験調査計画における視点

比較実験調査の計画にあたっては以下の視点から結果の分析を行うことを考慮した。

(1) キー項目がDQ法でも率直に答えられる内容である場合、DQ法とIC法の推定値は一致するか。逆に率直に答えにくい内容である場合、IC法の方が大きな推定値が得られるか。

(2) 質問文の導入が短くても調査方法や調査目的に関してIC法の回答者が不審を抱かないか。調査では一般に回答法などの説明は簡潔な方がよい。後述のように、短い説明を希望する調査員も多い。IC法の原理の説明はかえって回答者の注意を喚起し、キー項目に対して過剰反応させるおそれもある (Ahart and Sackett, 2004)。一方で短すぎる説明は調査目的等に対する回答者の不審を招き、IC法の失敗原因となる可能性がある。本稿の面接調査ではIC法の原理は説明せず、調査目的も漠然と提示した上で、調査員に回答者の反応を評定してもらう。

(3) IC法の原理を説明しなくともその秘匿特性を正しく理解できているか。回答者がIC法の特性を正しく理解しなければ、IC法は成功しない。回答者の理解度を直接調べることは困難であるが、率直には答えにくい事柄を調べる方法としてIC法とDQ法のどちらがよいかを尋ねることで示唆が得られよう。

(4) IC法の回答の方が、DQ法の「当てはまる」項目の合計数より小さいという回答傾向が認められるか。調査員が回答を確認できる他記式調査でも追認されれば、その原因が回答者の誤解や調査モードの影響である可能性は少なくなる。

2.2 キー項目と非キー項目

キー項目としては表1の四つを用いた (以下、表1の記号あるいは略称を用いる)。これらは Tsuchiya et al. (2007) の結果を参考に、DQ法で率直に答えられるとされた項目 (「献血」と「宝くじ」) とそうでない項目 (「水虫」と「万引き」)、ある程度大きな \hat{Y} が予想される項目 (「献血」と「水虫」) とそうでない項目 (「宝くじ」と「万引き」) という組み合わせで選んだものである。

表2に非キー項目のリストを示す (以下、非キー項目のリストには記号を用いる)。同時に提示する非キー項目の数は $G=3$ と $G=4$ の二通りとした。 $G=2$ ではキー項目を秘匿しにくく、長いリストは非標本誤差を拡大するおそれがあるためである。「献血」と「宝くじ」にはDQ法でも率直に答えられる非キー項目、「水虫」と「万引き」には率直に答えにくい非キー項目を割り当てた。これは、説明が短いため、キー項目が内容的に突出して回答者の不審を招くことがないようにするためである (Biemer and Wright, 2004)。さらに同一の項目リストの中

表 1. キー項目.

記号	略称	キー項目
B	「献血」	献血をしたことがある
L	「宝くじ」	宝くじで一万円以上当たったことがある
A	「水虫」	水虫になったことがある
S	「万引き」	万引きをしたことがある

表 2. 項目リストと各項目の直接質問法による推定値および全回答者による H スコア.

記号	項目	推定値 (%)	H スコア
B	献血をしたことがある	53.11 (1.66)	1.1881 (.0123)
B3	地域のボランティア活動に参加したことがある	47.68 (1.75)	1.2433 (.0189)
	街頭署名に応じたことがある	54.29 (1.78)	1.2503 (.0199)
	よその子どもを注意したことがある	67.27 (1.61)	1.4527 (.0231)
B4	自分と同姓同名の人を知っている	14.56 (1.31)	1.2424 (.0195)
	ハチに刺されたことがある	49.57 (1.66)	1.1781 (.0182)
	フリー・マーケットに出店したことがある	13.61 (0.95)	1.2469 (.0207)
	携帯電話・PHS を便器に落としたことがある	7.47 (0.78)	1.4221 (.0295)
L	宝くじで一万円以上当たったことがある	16.99 (1.06)	1.4801 (.0198)
L3	朝, 二日酔いで苦しんだことがある	36.59 (1.46)	1.3852 (.0236)
	救急車に乗ったことがある	36.03 (1.66)	1.3429 (.0249)
	コンタクト・レンズをしていた (している)	20.06 (1.26)	1.2828 (.0218)
L4	初日の出を見たことがある	68.14 (1.68)	1.1138 (.0155)
	一日に三回以上歯を磨く	23.45 (1.30)	1.3144 (.0232)
	毎日なにかしら乳製品を摂っている	61.21 (1.81)	1.1895 (.0205)
	自分用の万歩計を持っている	22.03 (1.24)	1.1927 (.0206)
A	水虫になったことがある	41.78 (1.57)	1.7817 (.0249)
A3	人のものを借りたまま返さなかったことがある	22.66 (1.37)	2.1644 (.0420)
	電車内にゴミを置いたまま降りたことがある	10.07 (1.00)	1.9686 (.0415)
	店頭の商品を傷つけてしまったことがある	7.32 (0.82)	1.9612 (.0371)
A4	成人前に酒を飲んだことがある	53.66 (1.62)	1.5661 (.0297)
	無言電話をかけたことがある	3.38 (0.56)	2.5306 (.0538)
	電車内の棚に置いてあった新聞・雑誌を手を取ったことがある	25.73 (1.48)	1.6552 (.0350)
	買ったものが気に入らず, 難癖をつけて返品したことがある	9.83 (0.91)	1.9642 (.0396)
S	万引きをしたことがある	—	2.7480 (.0384)
S3	行列に割り込んだことがある	12.20 (1.04)	1.8664 (.0388)
	公園の花を持ち帰ったことがある	4.01 (0.57)	2.1605 (.0431)
	騒音などで近所から苦情を言われたことがある	8.26 (0.77)	1.8119 (.0393)
S4	選挙で棄権したことがある	58.77 (1.77)	1.5561 (.0291)
	他人の傘を持ち帰ったことがある	15.89 (1.24)	2.0345 (.0427)
	店や会社に苦情の投書・電話をしたことがある	26.44 (1.53)	1.7835 (.0353)
	待ち合わせにわざと遅れて行ったことがある	9.21 (0.84)	1.8085 (.0367)

カッコ内は標準誤差

表 3. 質問の配列.

調査票	問 1	問 2	問 3	問 4	問 5	問 6	問 7	問 8
A	B3	S3	L+L3	A+A3	L4	A4	B4+B	S4+S
B	L3	A3	B3+B	S3+S	B4	S4	L+L4	A+A4
リストの項目数 (L)	3	3	4	4	4	4	5	5

B は献血, L は宝くじ, A は水虫, S は万引き, を表す.

では, DQ 法による推定値が大きい非キー項目と小さい非キー項目がなるべく同時に含まれるよう配慮した.

IC 法の調査票は A と B の二種類を用意した. 項目リストの提示順序を表 3 に示す. 順序は, 率直に答えやすいリストとそうでないリストを交互に, 項目が少ないリストから多いリストへと定めた. 例えば調査票 A の問 1 では B3 の三つの非キー項目を表 2 に示す順序で提示した. L+L3 はキー項目「宝くじ」の後に L3 の三つの非キー項目を並べた長リストを表し, B3+B は B3 の後にキー項目「献血」を並べた長リストを表す. キー項目「献血」の推定値は, 調査票 B の問 3 (B3+B) の平均個数から調査票 A の問 1 (B3) のそれを引く, あるいは調査票 A の問 7 (B4+B) の平均個数から調査票 B の問 5 (B4) のそれを引くことで求まる.

2.3 調査票の構成

調査票は A と B の二種類の他に, 比較のため, 調査票 A と B のそれぞれに対応し, 提示したリスト項目それぞれを直接質問する DQ 法の調査票 CA と CB を用意した. 例えば調査票 CA では, 調査票 A と同一の項目リストを同一の順序で提示し, 項目ごとに DQ 法による回答を求めた. ただしキー項目「万引き」については対象者からの強い拒否が予想され, DQ 法で面接調査することは困難と判断し, DQ 法の調査票からは除外した. このため調査票 CA の問 8 は短リスト S4, 調査票 CB の問 4 は短リスト S3 となった.

各調査票の構成は以下のとおりである.

・IC 法と DQ 法

IC 法と DQ 法の質問文作成にあたり, 実査を担当する調査員に対してどのような質問文が適切かヒアリングした (平井・土屋, 2006). 回答者にあまり不審を抱かせず, かつ長くなり過ぎないことを目標にした質問文¹⁾を問 1 の前に置くこととした.

IC 法の各質問では, 当該リストに含まれる項目全てと, 当てはまる項目の数を回答選択肢の形式として回答票で提示した. 例えば三つの項目から成るリストでは, 回答選択肢は「(ア)0 個 (ない)」「(イ)1 つ」「(ウ)2 つ」「(エ)3 つ」の四つとなる. DQ 法では, 対応する IC 法の調査票と同一の項目リストと, 「(ア)当てはまる」「(イ)当てはまらない」という回答選択肢を提示し, 項目ごとに回答選択肢を選んでもらった. IC 法, DQ 法のいずれにおいても「その他」と「わからない」を隠しコードとした.

・回答者反応

問 8 が終わった段階で, 調査員には回答者の反応を以下の四つの中から一つ選んでもらった.

- (1) 回答方法を直ちに (問 1 の質問文を一度読み上げるだけで) 理解し, 調査の目的や内容も不審に思っていないようだった
- (2) 回答方法は直ちに理解したが, 調査の目的や内容を不審に思っているようだった
- (3) 回答方法は直ちに理解できなかったが, 調査の目的や内容を不審に思っていないようだった
- (4) 回答方法を直ちに理解できず, 調査の目的や内容も不審に思っているようだった

表 4. 調査実施の概要.

	A 群	B 群	CA 群	CB 群
質問の種別	Item Count 法	Item Count 法	直接質問法	直接質問法
調査時期	10 月 7～10 日	11 月 10～13 日	9 月 9～12 日	
計画サンプル	2,000	2,000	1,044	956
回収サンプル (率)	1,328 (66.4%)	1,313 (65.7%)	710 (68.0%)	621 (65.0%)

・率直に答えられるか

問 9 と問 10 では、一般にキー項目や非キー項目が DQ 法でどの程度率直に答えられるかを評定してもらった²⁾。回答者の負担軽減のため、28 個の非キー項目を 14 個ずつ二分し一方を調査票 A と CA に、もう一方を調査票 B と CB に割り当てた。四つのキー項目は全ての調査票に含めた。さらに一つの調査票に割り当てた $14 + 4 = 18$ 個の項目を二分し、一回の質問では 9 個の項目を同時に提示した。回答選択肢は「(ア)率直に答えられる」「(イ)まあ率直に答えられる」「(ウ)あまり率直には答えられない」「(エ)率直には答えられない」の四つであり、「その他」と「わからない」を隠しコードとした。

・質問法に対する評価

最後に問 11 で、IC 法の調査票では、率直には答えにくい事柄を調べる方法として IC 法と DQ 法のどちらが答えやすいか、あるいはどちらも同じかを尋ねた。DQ 法の調査票では、率直には答えにくい事柄を調べる方法として DQ 法はどの程度答えやすいかを 5 段階で評定してもらった。

2.4 調査実施の概要

調査は、社団法人中央調査社が個別面接聴取法で毎月実施しているオムニバス調査 (母集団は全国の 20 歳以上男女個人) を利用して 2005 年の秋に行った。一般に IC 法の推定値の標本誤差は DQ 法より大きくなるため、可能なコストの範囲でなるべく大きなサンプルを得ることが望ましい。調査員が複数の調査票を取り違えないよう、10 月は調査票 A を用いた調査、11 月は調査票 B を用いた調査とした。DQ 法の調査票 CA と CB は 9 月の調査でサンプルを地点ごとに二分し、各々割り当てた。

サンプルの抽出は地域および市郡規模を層、国勢調査区を第一次抽出単位 (PSU) とした層化二段確率比例抽出である。層の数は 30、抽出された PSU の数は 158 である³⁾。第二段の個人の抽出は住民基本台帳あるいは選挙人名簿による。表 4 に回収率を含めた調査実施の概要を示す。

3. 調査結果

3.1 キー項目の推定値

以下の分析では、IC 法あるいは DQ 法の八つの質問項目の全てに「その他」と「わからない」がなかった回答者のみを用いる。分析に含めなかった回答者数は A 群が 7 名 (回収サンプルの 0.5%)、B 群が 12 名 (0.9%)、CA 群が 34 名 (4.8%)、CB 群が 26 名 (4.2%) であり明らかに DQ 法の方が多い。この点に関しては最後の節で考察を加える。

まず分析の視点 (1) に従い、調査票および質問法ごとの各キー項目の推定値を比較した結果が表 5 である。カッコ内は標準誤差である。2.3 節で述べたとおり「万引き」は DQ 法では質問していない。IC 法の A 群や B 群とは、当該群に長リストを割り当てたときの (1.3) 式による推定値である。IC 法の合計は (1.4) 式による。

表 5. 各キー項目の推定値 (%).

	献血	宝くじ	水虫	万引き
直接質問法				
CA 群	55.18 (2.06)	16.12 (1.54)	40.24 (1.91)	—
CB 群	50.76 (2.11)	17.98 (1.52)	43.53 (2.37)	—
合計	53.11 (1.66)	16.99 (1.06)	41.78 (1.57)	—
Item Count 法				
A 群	47.50 (4.50)	4.07 (4.46)	18.69 (3.55)	-1.78 (4.55)
B 群	65.23 (6.05)	15.74 (5.45)	39.28 (4.60)	19.41 (2.81)
合計	56.30 (2.87)	9.86 (2.98)	28.91 (2.09)	8.73 (2.11)

カッコ内は標準誤差

表 6. 調査員による回答者反応の判断.

	(1) 方法 OK 目的 OK	(2) 方法 OK 目的 NG	(3) 方法 NG 目的 OK	(4) 方法 NG 目的 NG	回答者数
直接質問法					
CA 群	75.6 %	14.9 %	7.7 %	1.8 %	676 名
CB 群	72.9 %	16.6 %	8.9 %	1.5 %	595 名
合計	74.4 %	15.7 %	8.3 %	1.7 %	1,271 名
Item Count 法					
A 群	76.4 %	14.2 %	7.2 %	2.2 %	1,321 名
B 群	73.6 %	16.4 %	7.8 %	2.2 %	1,301 名
合計	75.0 %	15.3 %	7.5 %	2.2 %	2,622 名

合計の推定値で DQ 法と IC 法を比較すると、「献血」は 3.19 ポイント差で IC 法の方が大きく、「宝くじ」は 7.13 ポイント、「水虫」は 12.87 ポイント差で IC 法の方が小さい。同一質問法内で比較すると、DQ 法は最大でも「献血」の 4.42 ポイント差であるが、IC 法は「献血」が 17.73 ポイント、「宝くじ」が 11.67 ポイント、「水虫」が 20.59 ポイント、「万引き」が 21.19 ポイントの差があり、A・B 群間の差はかなり大きい。さらに「万引き」の A 群では負の推定値が得られている。IC 法による推定はかなり不安定であり、成功したとは言い難い。

3.2 回答者反応

以下では残り三つの分析視点に沿いながら、IC 法による推定値が安定しなかった原因を探る。まず表 6 は、分析の視点 (2) から調査員に判断してもらった回答者の反応を集計したものである。

どの群もほぼ 4 人のうち 3 人が「(1) 方法 OK 目的 OK」であり、DQ 法と IC 法の違いは認められない。IC 法群では IC 法特有の回答法の意図などを特に説明しなかったが、調査員が判断する限り、回答者が特に不審を抱いたということはなさそうである。すなわち表 6 からは、推定値の不安定さの原因は見出せない。詳細は割愛するが、DQ 法を含まない群においても、回答者の年齢があがるにつれ「(1) 方法 OK 目的 OK」が減り、70 歳以上では 57.3% (CA 群) から 65.1% (A 群) となっている。

3.3 質問法に対する評価

次に分析の視点 (3) に基づき、率直に回答しにくい項目では IC 法と DQ 法のどちらが答えやすいか尋ねた結果を表 7 に示す。実際に IC 法で回答した A 群と B 群に対してのみ、この質問

表 7. 質問法に対する評価.

属性	Item Count 法	どちらも同じ	直接質問法	他・D.K.	回答者数
性別					
男性	47.1 %	30.8 %	15.8 %	6.3 %	1,242 名
女性	53.8 %	23.7 %	14.9 %	7.6 %	1,380 名
年齢層					
20 歳代	54.3 %	27.8 %	14.4 %	3.4 %	291 名
30 歳代	55.8 %	24.2 %	14.6 %	5.4 %	446 名
40 歳代	58.0 %	23.9 %	14.1 %	4.1 %	440 名
50 歳代	50.6 %	27.4 %	16.4 %	5.5 %	543 名
60 歳代	45.8 %	29.8 %	14.5 %	9.9 %	496 名
70 歳以上	40.4 %	29.3 %	17.5 %	12.8 %	406 名
学歴					
中学	43.2 %	26.9 %	15.2 %	14.7 %	435 名
高校	49.0 %	29.4 %	15.6 %	6.0 %	1,321 名
大学	57.2 %	23.5 %	14.9 %	4.4 %	860 名
調査票					
A 群	47.5 %	26.4 %	16.2 %	9.8 %	1,321 名
B 群	53.8 %	27.7 %	14.4 %	4.1 %	1,301 名

は行っている。ほぼ 5 割前後の回答者が「Item Count 法の方が答えやすい」としており、これらの回答者は IC 法の秘匿特性を正しく理解したと考えてよいであろう。属性別に見ると男性より女性が、高齢層より若年層が、低学歴より高学歴が「IC 法」を選ぶようである。

「IC 法」という回答は B 群の 53.8% に比べ、A 群では 47.5% と 6.3 ポイント少ない。群間差の修正 Wald 検定 (Fellegi, 1980) の結果は明らかに有意であり ($F_{(3,258)} = 5.51; p = .001$), A 群と B 群が等質という IC 法的前提条件が成立していない可能性がある。

3.4 H スコア

この節では各項目の H スコア (hesitation score) を分析に用いる。H スコアとは、個々の項目の DQ 法での答えやすさを尋ねた質問 (問 9 と問 10) において、「率直に答えられる」という回答を 1, 「率直には答えられない」を 4 としたときの項目ごとの平均である。仮に当該項目を DQ 法で尋ねたときにどの程度「率直に答えることをためらうか」を点数化したものであり、Tsuchiya et al. (2007) が分析に用いている。表 2 には当該項目に回答した全ての回答者を用いて求めた H スコアを示した。「万引き」の H スコア 2.7480 は表中で最も高く、率直には答えにくい項目であることが確認できる。

図 1 は群ごとに求めた H スコアをプロットした結果である。ただし調査票の設計上、キー項目は全四群で尋ねているが非キー項目群は二分して尋ねている。同一の非キー項目の H スコアの比較は A 群と CA 群の間、B 群と CB 群の間では可能だが、A 群と B 群の間では不可能である。そこで A 群と B 群の間の比較もある程度可能となるよう、図 1 では Tsuchiya et al. (2007) が Web 調査により求めた H スコアとの差を便宜的に用いて各項目をプロットした⁴⁾。

まず明らかに IC 法群 (A 群と B 群) は DQ 法群 (CA 群と CB 群) より H スコアが大きい。DQ 法のサンプルが、質問に対し「率直に答えられる」人々に当初から偏っていたとは考えにくい。詳細は割愛するが、Tsuchiya et al. (2007) の Web 調査データを再分析しても、IC 法群の方が DQ 法群より H スコアが高い傾向が認められる。DQ 法や IC 法への回答行動が、その後の「率直に答えられるか」への回答に影響したと考えてよい。

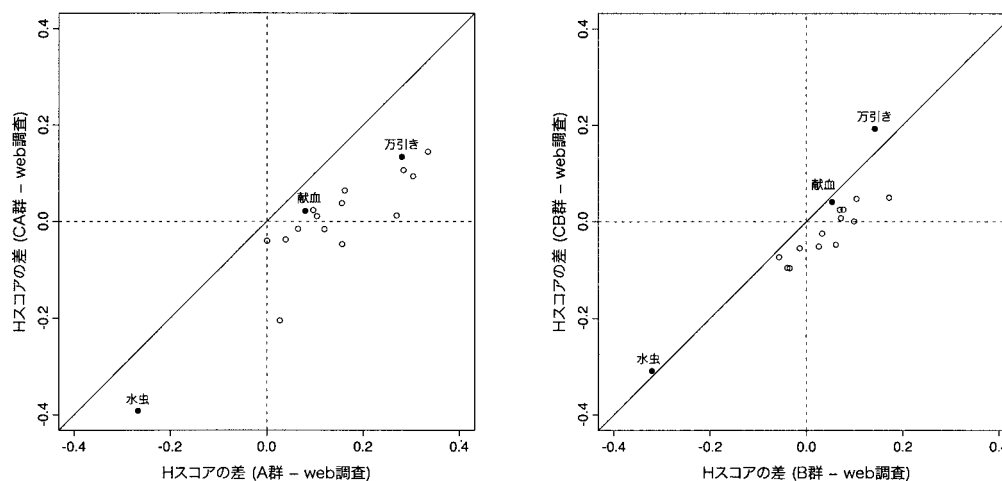


図1. 各群のHスコア.

さらにB群ではWeb調査とのHスコアの差が高々 ± 0.1 程度にとどまるのに対し、A群ではプラスの方へ大きく広がる。キー項目を比較すると、「献血」はA群の1.21に対しB群は1.19 ($F_{(1,260)} = 0.46; p = .500$), 「宝くじ」はA群の1.54に対しB群の1.46 ($F_{(1,260)} = 2.42; p = .121$), 「水虫」はA群の1.83に対しB群の1.78 ($F_{(1,260)} = 0.66; p = .419$), 「万引き」はA群の2.83に対しB群の2.69 ($F_{(1,260)} = 2.02; p = .156$)と、いずれも有意ではないもののA群がB群より大きい。Hスコアの値はその直前の質問群に影響され得る事実や、表7の結果を考え合わせると、調査票Bに比べ調査票AはIC法の質問に対する回答者の警戒感を強めた可能性がある。

3.5 個数の比較

最後に分析の視点(4)に従い、横軸を L/H 、縦軸を $(\bar{C} - \bar{D})/(\bar{C} + \bar{D})$ として、S3+SとS4+Sを除く14の項目リストをプロットしたものが図2の左である。ただし L はリストの項目数、 H はリストに含まれる項目のHスコア(表2に示される)の平均、 \bar{C} はIC法による回答の平均、 \bar{D} はDQ法で調べた当てはまる個数の平均である。

項目リストA3を除く全ての項目リストで $\bar{C} < \bar{D}$ であり、IC法の方が平均個数は少ないという回答傾向が認められる。例として表8に項目リストB3の回答分布を示す。明らかにDQ法に比べIC法は個数が少ない方へ偏る ($F_{(3,257)} = 13.62; p < .001$)。図2では全体として右下がりの傾向が認められ、リストの項目数 L が多いほど \bar{C} と \bar{D} の差は広がる。またHスコアの平均 H が大きいほど \bar{C} と \bar{D} の差は縮小する。これは、DQ法では率直な回答ができず、 \bar{C} と同程度に \bar{D} も小さくなるためと考えられる。同時に \bar{C} を引き下げる要因はIC法特有の過少回答傾向が主であり、IC法では率直な回答が得られている、つまりIC法の秘匿特性は基本的には理解されているということが示唆される。

図2では、特に項目リストB3、S3、L+L3、A+A3の四つが下方ヘシフトしている。これらは調査票AとCAの最初の四問である。そこで図2の右には、横軸に質問(項目リスト)の提示順序 x 、縦軸に x 番目の項目リストも含めそれまでに提示された全ての項目に当てはまらない(当てはまる項目が0個)とする回答者の割合を示す。B群とCB群では、1問目と2問目が終わった時点で「どれも当てはまらない」とする回答者の割合に差はない。3問目以降、IC法のB群の方がわずかに「0個」と答え続ける回答者が多くなる。

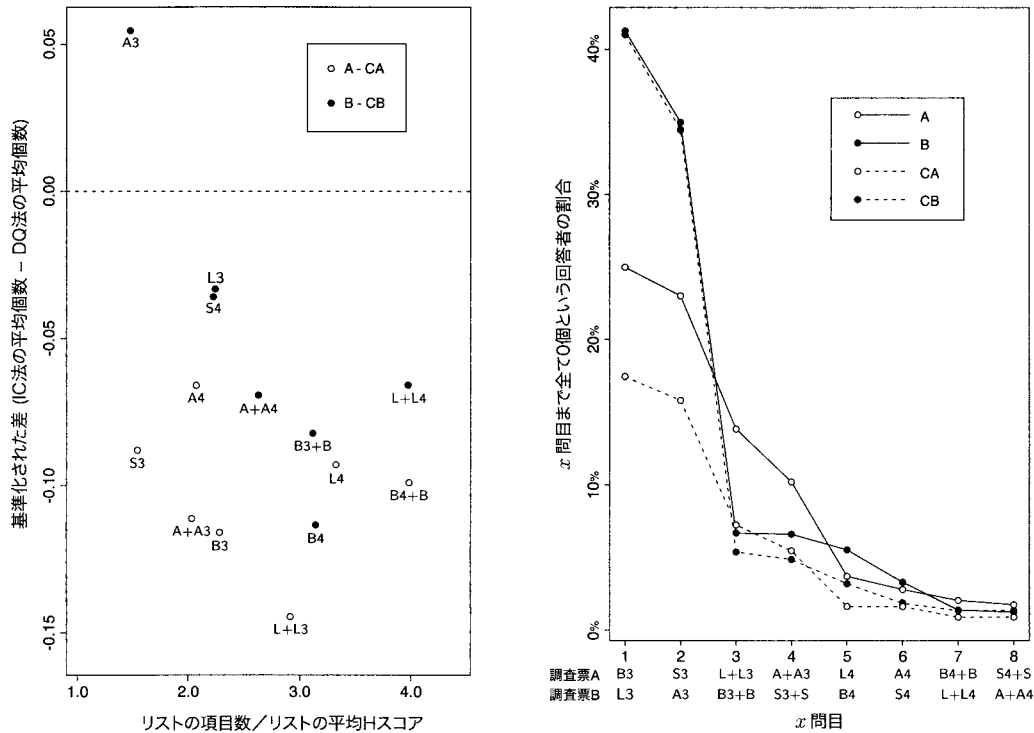


図 2. 当てはまる項目の個数の比較.

表 8. 項目リスト B3 に対する回答分布.

調査票	0 個	1 個	2 個	3 個	回答者数
A 群 (IC 法)	25.0 %	35.9 %	25.0 %	14.2 %	1,321 名
CA 群 (DQ 法)	17.5 %	26.2 %	33.3 %	23.1 %	676 名

一方 A 群と CA 群では、1 問目で既に 7 ポイント以上の差がある (表 8 も参照)。H スコアの低さが示すとおり、1 問目のリスト B3 の各項目は率直に答えにくい内容ではない。7 ポイントの差は明らかに A 群の回答に何らかの“異常”があることを示す。A 群と CA 群の差は質問が進んでも解消せず、4 問目が終わった時点で 5 ポイント近い。4 問目までに提示される 12 個の非キー項目は四つ全ての調査票で同一である。キー項目に違いはあるが、四問全てを通して「0 個」と答える回答者の割合は、本来、四つの調査票でほぼ一致するはずである。A 群を除く三群は 5~6% で確かに一致しているが、A 群は 10% を超えて他の三群と異なる。A 群が 10% を割るのは 5 問目である。5 問目のリスト L4 には、「初日の出を見たことがある」や「毎日なにかしら乳製品を摂っている」といった、当てはまる人が多く、かつ H スコアの小さな項目が含まれる。

以上の分析結果をまとめると、IC 法による推定値が安定しなかった主な原因として以下のことが考えられる。IC 法の A 群では、当てはまる個数を“異常”に過少回答させる何らかの要因がはじめに強く働いた。その影響は 4 問目まで続き、5 問目には薄らいだ。一方、B 群では“異

常”な過少回答は生じなかった。その結果、最初の四問を使う推定値では、A群が短リストの場合には過大（「献血」の65.23%）となり、長リストの場合には過小（「宝くじ」の4.07%、「水虫」の18.69%）となった。後半の四問を使う推定値は、全体にやや過小であるものの、DQ法に近い推定値（「献血」の47.50%、「宝くじ」の15.74%、「水虫」の39.28%）が得られた。

4. 考察

本稿の実験調査におけるIC法の適用自体は、一言でいえば成功しなかったと結論せざるを得ない。しかしその原因の究明を通じ、今後のIC法の研究・実用化へ向け多くの示唆が得られた。以下にそれらをまとめた。

IC法による推定値の不安定さは、今回は主にA群とB群の過少回答傾向の大きさの違いに起因した。二群の等質性がIC法的前提である。その等質性が崩れた主因は、サンプルの構成方法というより調査の実施方法にある。例えば調査員にとってIC法の実施はA群が初めてであり、B群は二回目といった調査員の慣れの影響である。また調査票における文脈効果も無視できない。オムニバス調査では不可避の問題であるが、調査票Aは内閣支持など定例的な時事項目の直後に位置したのに対し、他の調査票は調査の冒頭に位置した。前問の有無により回答行動が異なった可能性は十分ある。図1や図2に鑑みれば、項目リストの提示順序、特に最初数個の項目リストもIC法では回答に大きく影響しよう。

二群の等質性を担保するには以下の三つの対策が考えられる。第一は冒頭数問の項目リストを二群で共通とし、両群の回答者をIC法に同等に慣れさせることである。当然、質問数が増し、回答者を消耗させるおそれはある。しかしIC法に不慣れな回答者に、主目的となる項目リストを直ちに数えさせる方がリスクはより大きい。共通リストの内容には未だ議論の余地があろう。警戒感を避けるため率直に答えられる内容がよいという考え方もあれば、やや答えにくい内容とすることでその後のリストへの抵抗感をやわらげるという考え方もある。Droitcour et al. (1991)は少なくとも一個は当てはまるリストをウォーミングアップに用いているが、図2では調査票Bの方が調査票Aより「0個」回答が多く、一個も当てはまらないリストが不適当とも限らない。

対策の第二は対応するリストの提示順序を二群間で揃えることである。例えば調査票Aのx番目の項目リストがB3ならば、調査票Bのx番目の項目リストはB3+Bとする。本稿では冒頭に項目の少ないリストを置くことでIC法に慣れさせ、リストの項目数を揃えることで二群間の等質性の確保を狙った。しかし冒頭に共通のウォーミングアップ質問を置けば、その後はリストの項目数よりも内容を揃える方が、文脈効果の点からみて等質性の保持には有効と考えられる。

第三は以上の二つの対策で等質性が確保できない場合に備え、統計的な調整用の補助変数として共通の項目リストを追加することである。対応するリストの順序を揃えても、長リストではキー項目が加わるため二群間で異なる文脈効果が生じるおそれがある。IC法形式の共通の変数があれば、例えば傾向スコア (Rosenbaum and Rubin, 1983; Rosenbaum, 2002) などを用いた調整が可能となろう (土屋 他, 2006)。文脈効果調整用の補助変数は、当然、文脈効果が生じた後に置く必要がある。つまり主目的の項目リストより後に質問するため、ウォーミングアップ質問では不十分である。ただしこの第三の対策は万一のときの保険と位置づけ、プリテストなどを通じ、無調整でも等質性を確保できるよう努めるべきである。仮に非等質性が疑われる場合でも、その原因を追求せず機械的に調整することは好ましくない。

DQ法との比較では、IC法の方が個数が少ない回答傾向が Biemer and Write (2004) や Tsuchiya et al. (2007) に続いて再認められた。自記式・他記式を問わず、この回答傾向は一般的な現象と

考えてよい。もちろん DQ 法が過大回答であるという仮説も否定はできないが、IC 法による過少回答の方がより疑わしい。IC 法の推定精度を高めるには過少回答の原因を今後明らかにすることが不可欠である。手がかりの一つは「その他」と「わからない」である。DQ 法に比べ IC 法はその割合が小さかった。DQ 法では「その他」「わからない」となる回答者が、IC 法ではいずれかの回答選択肢、特に「0 個」を選び、平均個数を引き下げる要因の一つになったと想像できる。そうであれば「その他」「わからない」を回答選択肢としてあえて提示し、非標本誤差を縮小できると考えられる。

ただし表 8 で分かるとおおり、「0 個」を除いても IC 法の回答は個数が少ない方へ偏る。原因の解明と対策には、IC 法の心理的な回答プロセスにまで踏み込んだ研究が必要である。そのため統計的な調査手法と同時に、例えば、回答中に考えていることをそのまま話してもらう thinkalouds (Sudman et al., 1996) などの技法の利用も有効である。本稿の調査票 A を用いて thinkalouds インタビューを実施した土屋・平井 (2006) によれば、秘匿性を持つ IC 法は正直な回答への動機づけを高めると同時に、正確な回答への動機づけを低下させるおそれがあるという。この後者の特性が十分な記憶検索を阻害し、過少回答を引き起こす要因となった可能性が考えられる。

IC 法の導入文を聞いた、あるいは読んだときの回答者の思考も thinkalouds インタビューによって調べることができよう。本稿の調査員判断によれば、IC 法の説明が短くとも特に問題はなさそうであった。しかし表 6 では A 群と B 群の間に明確な差が見られず、調査員判断の妥当性や有効性には疑問が残る。IC 法に関する調査員の説明が短いため、回答者は項目リストの内容から調査目的等を憶測することになり、これが大きな文脈効果の誘因となったのかもしれない。文脈効果が生じること自体は避けようもないが、少なくとも不適切な文脈効果を生じさせないような非キー項目の内容や組み合わせについても今後研究が必要であろう。

質問法に対する回答者の評価結果を見る限り、特に説明はなくとも IC 法の秘匿性は多くの回答者に理解されたようである。図 1 に見られる、DQ 法群と IC 法群の間の H スコアの明らかな差も、この主張を裏付けるものと考えられる。ただし表 7 のとおおり、IC 法への評価は回答者の年齢や学歴によって少なからず異なる。IC 法は、標本誤差を犠牲にしても非標本誤差を小さくすることで誤差全体の縮小を目指す方法である。IC 法の秘匿性を回答者が理解できず、非標本誤差を低減できないのであれば、IC 法を採用する積極的な理由はない。キー項目の推定精度を上げるには、回答者の属性に応じて IC 法と DQ 法を使い分けるハイブリッド型の質問法などの開発も視野に入れた研究が必要となる。

注.

1) IC 法における質問文は以下のとおおりである。

これから、日頃のちょっとした経験や体験についてうかがいます。良いことも良くないこともいろいろ出てきますが、この調査では、全体として、皆さんがそういった事をどのくらい経験しているのか、調べています。まず、この 3 つのことがらのうち、あなたに当てはまることからは全部でいくつありますか。3 つとも当てはまったら「3 つ」、2 つだけだったら「2 つ」というように、当てはまる数でお答えください。

また、DQ 法における質問文は以下のとおおりである。

これから、日頃のちょっとした経験や体験についてうかがいます。良いことも、良くないことも、いろいろ出てきますが、この調査では、そういった事をどのくらいの人が経験しているのか、統計的に調べることだけが目的です。それぞれあなたに

当てはまるかどうか、なるべく率直にお答え下さい。

2) 「率直に答えられるか」を問う質問文は以下のとおりである。

では次に、ここにある9つのことがらは、それぞれ、このようなアンケート調査で、当てはまるかどうかを直接、たずねたときに、誰にでも率直に答えられることがらだと思いますか、それとも人によっては、率直には答えられないことがらだと思いますか。

3) 一つのPSUのみ回収した層が二つ(10月調査では三つ)あり、これらの層は、標準誤差の算出に当たって、それぞれ隣の地域の同じ市郡規模の層に合併した。

4) Web調査では、キー項目「宝くじ」と非キー項目「成人前に酒を飲んだことがある」のHスコアは調べられていないため、図には表示していない。

参 考 文 献

- Ahart, A. M. and Sackett, P. R. (2004). A new method of examining relationships between individual difference measures and sensitive behavior criteria: Evaluating the unmatched count technique, *Organizational Research Methods*, **7**, 101–114.
- Barton, A. H. (1958). Asking the embarrassing questions, *The Public Opinion Quarterly*, **22**, 67–68.
- Biemer, P. and Brown, G. (2005). Model-based estimation of drug use prevalence using item count data, *Journal of Official Statistics*, **21**, 287–308.
- Biemer, P. P. and Wright, D. (2004). Estimating cocaine use using the item count methodology: Preliminary results from the national survey on drug use and health, Paper presented at the annual meeting of the American Association for Public Opinion Research, Phoenix, Arizona.
- Boruch, R. F. and Cecil, J. S. (1979). *Methods for Assuring the Confidentiality of Social Research*, University of Pennsylvania Press, Philadelphia.
- Chaudhuri, A. and Mukerjee, R. M. (1988). *Randomized Response: Theory and Techniques*, Marcel Dekker, New York.
- Couper, M. P. (2000). Web surveys: A review of issues and approaches, *Public Opinion Quarterly*, **64**, 464–494.
- Couper, M. P., Baker, R. P., Bethlehem, J., Clark, C. Z. F., Martin, J., Nicholls II, W. L. and O'Reilly, J. M. (1998). *Computer Assisted Survey Information Collection*, John Wiley & Sons, New York.
- Dalton, D. R., Wimbush, J. C. and Daily, C. M. (1994). Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior, *Personnel Psychology*, **47**, 817–828.
- Dalton, D. R., Daily, C. M. and Wimbush, J. C. (1997). Collecting “sensitive” data in business ethics research: A case for the unmatched count technique (UCT), *Journal of Business Ethics*, **16**, 1049–1057.
- Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsley, T. L., Visscher, W. and Ezzati, T. M. (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application, *Measurement Errors in Surveys* (eds. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz and Sudman Seymour), 185–210, John Wiley & Sons, New York.
- Fellegi, I. P. (1980). Approximate tests of independence and goodness of fit based on stratified multistage samples, *Journal of the American Statistical Association*, **75**, 261–268.
- Greenberg, B. G., Abul-Ela, A-L. A., Simmons, W. R. and Horvitz, D. G. (1969). The unrelated ques-

- tion randomized response model: Theoretical framework, *Journal of the American Statistical Association*, **64**, 520–539.
- 平井洋子, 土屋隆裕(2006). Item Count 形式の質問紙は調査員にどのように受け止められたか, Research Memo., No. 1003, 統計数理研究所, 東京.
- Hubbard, M. L., Casper, R. A. and Lessler, J. T. (1989). Respondent reactions to item count lists and randomized response, *Proceedings of the Survey Research Section of the American Statistical Association*, 544–548, American Statistical Association, Washington, D. C.
- Jobe, J. B., Pratt, W. F., Tourangeau, R., Baldwin, A. K. and Rasinski, K. A. (1997). Effects of interview mode on sensitive questions in a fertility survey, *Survey Measurement and Process Quality* (eds. Lars Lyberg, Paul Biemer, Martin Collins, Edith de Leeuw, Cathryn Dippo, Norbert Schwarz and Dennis Trewin), 311–329, John Wiley & Sons, New York.
- Kane, J. G., Craig, S. C. and Wald, K. D. (2004). Religion and presidential politics in Florida: A list experiment, *Social Science Quarterly*, **85**, 281–293.
- Kuklinski, J. H., Cobb, M. D. and Gilens, M. (1997a). Racial attitudes and the “New South”, *The Journal of Politics*, **59**, 323–349.
- Kuklinski, J. H., Sniderman, P. M., Knight, K., Piazza, T., Tetlock, P. E., Lawrence, G. R. and Mellers, B. (1997b). Racial prejudice and attitudes toward affirmative action, *American Journal of Political Science*, **41**, 402–419.
- LaBrie, J. W. and Earleywine, M. (2000). Sexual risk behaviors and alcohol: Higher base rates revealed using the unmatched-count technique, *The Journal of Sex Research*, **37**, 321–326.
- Lee, R. M. (1993). *Doing Research on Sensitive Topics*, Sage, London.
- O’Brien, D. M. (1979). *The Comprehension Factor in Randomized Response*, University Microfilms International, Michigan.
- Raghavarao, D. and Federer, W. T. (1979). Block total response as an alternative to the randomized response method in surveys, *Journal of the Royal Statistical Society, Series B (Methodology)*, **41**, 40–45.
- Rayburn, N. R., Earleywine, M. and Davison, G. C. (2003a). An investigation of base rates of anti-gay hate crimes using the unmatched-count technique, *Journal of Aggression, Maltreatment & Trauma*, **6**, 137–152.
- Rayburn, N. R., Earleywine, M. and Davison, G. C. (2003b). Base rates of hate crime victimization among college students, *Journal of Interpersonal Violence*, **18**, 1209–1221.
- Rosenbaum, P. R. (2002). *Observational Studies*, 2nd ed., Springer-Verlag, New York.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, **70**, 41–55.
- 逆瀬川浩孝, 高橋宏一, 鈴木達三 (1977). ランダム化器具を使用しないランダム回答法の実験的調査結果について, 統計数理研究所彙報, **24**, 79–94.
- Sniderman, P. M. and Grob, D. B. (1996). Innovations in experimental design in attitude surveys, *Annual Review of Sociology*, **22**, 377–399.
- Sudman, S. and Bradburn, N. M. (1982). *Asking Questions: A Practical Guide to Questionnaire Design*, Jossey-Bass, San Francisco, California.
- Sudman, S., Bradburn, N. M. and Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*, Jossey-Bass, San Francisco, California.
- Takahashi, K. and Sakasegawa, H. (1977). A randomized response technique without making use of any randomizing device, *Annals of the Institute of Statistical Mathematics*, **29**, 1–8.
- Tourangeau, R. and Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context, *The Public Opinion Quarterly*, **60**, 275–304.
- Tsuchiya, T. (2005a). Domain estimators for the item count technique, *Survey Methodology*, **31**,

41–51.

- Tsuchiya, T. (2005b). On the assumption required for the domain estimators for the item count technique, Research Memo., No. 944, 統計数理研究所, 東京.
- 土屋隆裕, 平井洋子 (2006). Thinkaloud インタビューによる Item Count 法の回答傾向の分析, Research Memo., No. 1013, 統計数理研究所, 東京.
- 土屋隆裕, 平井洋子, 小野 滋 (2006). 個別面接聴取法による Item Count 法と直接質問法の比較実験調査, Research Memo., No. 1004, 統計数理研究所, 東京.
- Tsuchiya, T., Hirai, Y. and Ono, S. (2007). A study on the properties of the item count technique, *Public Opinion Quarterly* (in press).
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, **60**, 63–69.
- Warner, S. L. (1971). The linear randomized response model, *Journal of the American Statistical Association*, **66**, 884–888.
- Wimbush, J. C. and Dalton, D. R. (1997). Base rate for employee theft: Convergence of multiple methods, *Journal of Applied Psychology*, **82**, 756–763.

An Empirical Study of Item Count Technique Based on Face-to-face Interviews: Some Suggestions for Its Application

Takahiro Tsuchiya¹, Yoko Hirai² and Shigeru Ono³

¹The Institute of Statistical Mathematics

²Faculty of Urban Liberal Arts, Tokyo Metropolitan University

³Synovate K. K.

When a survey questionnaire asks about sensitive topics, it can not be assumed that all respondents will give truthful answers. The item count technique, one of the most practical indirect questioning methods, requests merely the number of items applied to keep the respondents' secret. An estimate of the target key-item is derived from the differences between mean item count responses of two homogeneous samples. This paper attempts to estimate the percentages of four key-items including shoplifting by both item count and direct questioning techniques. The survey was conducted via face-to-face personal interviewing. The result suggests that the item count estimates can be unstable because of the context effect which disturbs the homogeneity of two samples.