A SELECTIVE HISTORY OF CIGARETTE ADVERTISING: PRELIMINARY PERSPECTIVES[1]

Debra Jones Ringold, The American University, Washington, D.C.[2]

ABSTRACT

In one form or another, health is a theme which has been at the center of cigarette advertising since the early part of this century [3]. The use of such health claims has been described as deceitful and representative of egregious abuse [22]. Others have suggested that the presence of such advertising makes it impossible for any smoker, or potential smoker, to ignore or forget the harmful aspects of smoking [5]. This paper briefly presents these two viewpoints and examines the degree to which cigarette advertising has historically incorporated health related information.

INTRODUCTION

Conventional wisdom has long recognized that cigarettes may be unhealthy and/or addictive. The disapproval and distrust of cigarettes is reflected in many of the slang terms traditionally used to describe them: coffin nails, gaspers, dope sticks and poison sausages [8]. As early as 1604, King James is said to have been disgusted at the fact that tobacco users become "obstinately addicted" to the substance [13]. That the American public has historically had some sense of the unhealthy and/or dangerous effects of cigarette smoking is illustrated by the results of two Gallup polls. A 1939 Gallup poll reported that in response to a question about New Year's resolutions, one of the five most frequently mentioned resolutions was to "stop smoking" [11]. In 1949 when Gallup asked, "Do you think cigarette smoking is harmful or not?, 52% of cigarette smokers and 66% of nonsmokers responded yes [12].

According to Tye [22], the early use of health claims in cigarette advertising was designed to offset these common-sense perceptions that smoking is unhealthy. "For 60 years tobacco firms have used advertising to deceive smokers and potential smokers into thinking it is safe to smoke. Cigarette firms have used unfounded health claims to encourage people to smoke despite the risk of harm" [22, p. 30]. Blum [2, p. 113] concurs, "The knowledge and common sense about cigarette smoking were there--but so were the mass media [health claims, "scientific" demonstrations, and physician "endorsements" in advertising] to undermine knowledge and cultivate mass denial."

A very different view of health claims in cigarette advertising was suggested by Brecher et al [3] and developed by Calfee [5, 7]. Succinctly, this view holds that through their own advertising, cigarette companies constantly reminded their customers that cigarette smoking constitutes a real threat to health. Brecher et al [3, p. 145] characterized these campaigns as using the "smoke-our-brand-and-stay-well-it's-those-other-brands-that-hurt-you" approach. Similarly, Calfee [5, 7] has asserted that the purpose of health advertising was to distinguish one brand from its competition; the side effect was to remind consumers constantly of the worrisome symptoms associated with smoking. Calfee [5, 7] further asserts that while effective in promoting one brand over another, such health claims were in some instances highly destructive to the interests of the cigarette industry as a whole.

While the use of health claims in cigarette advertising has been much discussed and highly regulated [7], no one has ever systematically examined the extent to which health claims have been utilized in cigarette advertising. Such an examination is of interest in the context of the two arguments presented above. The purpose of this paper then, is to contribute to our understanding of what types of information, or lack thereof, was/is published by cigarette companies, and was/is available to smokers, or potential smokers, through print advertisments. More specifically, this study may provide some insight into the number and types of claims made by individual brands. Further, this study examines whether cigarette print advertising has provided information about the health consequences of smoking. Since the adverse consequences of smoking often take several decades to reveal themselves, cigarette print advertising throughout the last sixty years is examined.

This paper is organized as follows. Next, the research objectives and methodology are discussed. This is followed by the presentation of the findings. The final section discusses the implications of this study and presents suggestions for additional research.

[1]The author wishes to acknowledge the efforts of Stefan Schechter and Sarah Kvale expended in the collection of the advertisements used in this study. Ms. Kvale and Raheel Masood served as the raters and are due my thanks. The helpful comments of Gary T. Ford are gratefully acknowledged.

[2]Assistant Professor of Marketing, Kogod College of Business

METHODOLOGY

With the foregoing in mind, the objective of this research was to determine: (1) the extent to which cigarette advertising included information about the health consequences of smoking; (2) the type of health information contained in such claims; and (3) how the extent and type of health information varied over time and by brand.

Given these objectives, content analysis was the most appropriate methodological approach. Information provided in advertisements from 1926 to 1985 was analyzed. This time period was selected because magazine advertisements for these products were uncommon prior to 1926 and were not yet available for 1986. Data collection proceeded as follows. Five cigarette brands offered for sale during this time period were selected. The original product type, manufacturer, and introduction date for each of the brands chosen was: Camel (unfiltered, R.J. Reynolds, 1913), Chesterfield (unfiltered, Liggett-Meyers, 1912), Old Gold (unfiltered, Lorillard, 1926), Viceroy (filtered, Brown and Williamson, 1936) and Lucky Strike (unfiltered, American Tobacco, 1916). In addition, the advertisements associated with a later filter brand (Kent, Lorillard, 1952) were examined.

These brands were chosen on the basis of market share and size of firm. Camel, Lucky Strike, and Chesterfield together commanded over 80% of the market in both 1925 and 1935 [5]. By 1950, those three plus Old Gold and Phillip Morris had a combined share of over 80% of the market. Viceroy, therefore, was a less popular brand and was included to contrast the other four. By 1963, the filter brands reduced the combined share of the original big three brands to 21% [5], and thus, Kent was selected to represent these entries. Further, brands offered by both large and small firms were desirable. Camel and Lucky Strike were produced by the two consistently largest firms (R.J. Reynolds and American Tobacco), while the remaining four brands were produced by smaller actors [5].

One print advertisement for each of these brands was sought for each of the sixty years. It was felt that one advertisement per year was sufficient to illustrate how advertising content changed over time. Advertisements were obtained from Time magazine (first published 1923) dated as close as possible to July 1. July 1 was selected in order to avoid seasonal variation and year-end budgetary effects. If an ad for any of the above brands could not be located in Time, an ad was pulled from the corresponding volume of The New Yorker (first published 1925, ceased cigarette advertising in 1964), The Saturday Evening Post (published 1821 to 1969 and 1971 to present), or Life (published 1936 to 1971 and 1979 to present), in that order. Of these publications, only Time is considered to be a "news and opinion magazine", the remaining three are categorized as "general editorial" [23].

Advertisements could not be located, in the publications chosen, for each brand in every year. Further research indicated that the primary reason for a missing advertisement was the total absence of magazine advertising for a given brand in a given year [18]. The next most common reason was, according to Leading National Advertisers [18], the absence of advertising in the publications selected for a given brand in a given year.

Advertisements for Camel were most successfully located with 0.77 of all years represented (n=46, 60 years). The other brands are represented as follows: Kent 0.76 (n=25, 33 years), Lucky Strike 0.73 (n=44, 60 years), Viceroy 0.68 (n=34, 50 years), Chesterfield 0.63 (n=38, 60 years) and Old Gold 0.40 (n=24, 60 years).[3] Of the 211 advertisements collected, 70.6 percent came from Time, 8.5 percent from The New Yorker, 8.5 percent from The Saturday Evening Post, and 12.3 percent from Life.

A coding scheme for the information conveyed in the headline, subhead, copy and visuals was then developed. A review of the literature,[4] and an examination of a variety of ads, suggested the following general categories of information related to: health, cigarette construction, taste, exhortation to purchase/slogans, hedonic satisfaction, tar and nicotine content, price/availability, coupons/contests, and celebrity/athlete endorsements. Each claim or statement (hereafter referred to as claims)[5] was entered only once regardless of the number of times that particular claim was repeated within a single advertisement. For example, if a claim appeared in the headline, the subhead, and the copy, the claim was coded for the headline only. The focus of the study was the number and type of distinct claims or statements, not the amount of repetition within the ads. This is consistent with the approach adopted by Pollay [17].

Two raters (judges), who were unaware of the purpose of the study, were trained in the use of the coding scheme. Each judge examined and coded the entire set of ads and the inter-rater

---

[3] All four of the unfiltered brands later added a filter. When advertisements for the unfiltered variety could not be located, the filtered variety of the same brand was substituted and noted. The years in which these substitutions began were: Camel 1971, Lucky Strike 1965, Chesterfield 1967, and Old Gold 1955.

[4] In particular see Doron [9]; Miles [16]; Taylor [19]; Calfee [5]; Pollay [17]; Tye [21]; Warner [24]; and The American Medical Association [1].

[5] An exhortation to buy a brand is not strictly an advertising claim, it is a directive. Likewise, encouraging contest participation is not an advertising claim, it is an invitation. However, the majority of the information found in these advertisements could reasonably be described as claims.

reliability was found to be 0.81.[6] Differences in code assignments were then resolved and the resulting consensus served as the final data base.

RESULTS

The results will be presented as follows. First, the most common categories of claims will be described and compared across brands and time periods. This will be followed by an examination of the specific health claims made and an evaluation by brand and by decade. Lastly, the visuals will be described briefly.

Most Common Claims

Taste claims (e.g. mildness, flavor, aftertaste) were the most frequently made in the total set (all brands, all years) of advertisements: they constituted 22.3 percent of all claims made. Next in frequency (19.7%) were statements associated with cigarette construction (e.g. filter, quality of tobacco, package). Health claims were the third most frequently made, with 18.2 percent of all claims being health related (e.g. reduced throat irritation, reduced coughs, protection against adverse health effects). Thus, nearly one-fifth of all claims involved health (health claims exclude the Surgeon-General's Warning and tar and/or nicotine figures). Statements relating to pleasure accounted for 10.7 percent. All remaining information fell into categories constituting 29.1 percent of claims made. In the 211 total advertisements, an average of 5.5877 claims were made (s.d. = 2.2857) and ranged from 1 to 12 claims per advertisement.

Since an advertisement was not available for all brands in all years, comparisons of claims by brand and over time were made on the basis of the percent of total claims constituted by a particular category of claims. The time periods used were as follows: 1926-29, 1930-39, 1940-49, 1950-54, 1955-59, 1960-69, 1970-79, and 1980-85. Analysis of claims by decade is convenient and is consistent with the work of advertising historians such as Fox [10] and Pollay [17]. The decade of the fifties was split in order to more closely examine the effect of the 1955 Federal Trade Commission ban on health claims in cigarette advertising. This allows for a pre- and post-1955 analysis consistent with Calfee [5]. Advertising of tar and nicotine figures by certain brands (Camel, Lucky Strike and Old Gold) was banned by the FTC in 1950-51 [5, 6]; tar and nicotine figures associated with brands not effected by earlier FTC actions were banned by the FTC in 1960; and the inclusion of tar and nicotine figures was subsequently mandated in 1970. Thus, changes in policy regarding the inclusion of tar and nicotine figures coincide well with the decade breakdown used.

As is shown in Table 1, from 1926 to 1954 the three most frequently made claims in cigarette advertising were those related to health, taste, and cigarette construction. Health claims were the most frequently made type of claim in three out of the four periods prior to 1954. A typical health claim made by R.J. Reynolds was "Not one single case of throat irritation due to smoking CAMELS." This was stated in conjunction with the findings "of noted throat specialists after a total of 2,740 weekly examinations of throats of hundreds of men and women who smoked Camels--and only Camels--for 30 days" (Time 1950). Viceroy, a Brown and Williamson product, suggested that..."Filtered cigarette smoke is better for your health" (Time 1951) and Lorillard's Old Gold asked..."Why Risk Sore Throats?" when Old Gold promised "Not A Cough In A Carload" (New Yorker 1929).

TABLE 1. Most Frequently Made Claims (Category as a Percent of Total Claims Per Period) for All Six Brands.

	26-29	30-39	40-49	50-54	55-59	60-69	70-79	80-85
Taste	21.15% (2)	22.18% (3)	23.21% (1)	16.94% (2)	30.68% (1)	32.52% (1)	19.35% (2)	13.21% (3)
Cig Con	15.38% (4)	22.58% (2)	21.94% (2)	14.75% (3)	29.55% (2)	30.89% (2)	11.61% (4)	
Health	23.08% (1)	24.19% (1)	15.19% (4)	26.78% (1)				
Pleasure		13.31% (4)			12.50% (4)	11.38% (4)		
E B/S	19.23% (3)		17.72% (3)	13.11% (4)	14.77% (3)	17.07% (3)		
TN Figs							32.26% (1)	32.08% (1)
Red TN								12.26% (4)
S Gen W							15.48% (3)	14.15% (2)
Other	21.16%	17.74%	21.94%	28.41%	12.51%	8.14%	21.30%	28.30%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%

(Rankings for the period are given in parentheses)

Legend

TN Figs = Tar & Nicotine Figures
Cig Con = Cigarette Construction
S Gen W = Surgeon General's Warning
E B/S = Exhortation to buy/slogan
Red TN = Reduced Tar & Nicotine

From 1955 to 1969, taste claims were most frequently made, followed by cigarette construction, exhortation to buy/slogan, and pleasure. From 1970 to 1985, tar and nicotine figures, taste claims, and the Surgeon-General's warning were most frequently conveyed. During the first five years of the eighties, tar and nicotine figures, the Surgeon General's warning,

[6] The measure of reliability used was the ratio of coding agreements to the total number of coding decisions. It has been suggested that researchers can be quite satisfied with coefficients of reliability above 85 percent [14].

and claims associated with reduced tar and/or nicotine were most frequently presented. Thus, health related information constituted one of the four most frequently made types of claim/statement in every period studied with the exception of 1955 to 1969. See Table 1 for more detail.

As discussed above, prior to 1955, health claims were an important component of cigarette advertising for five out of the six brands studied. See Table 2. Two of these six brands

TABLE 2. Most Frequently Made Claims (Category as a Percent of Total Claims Per Brand) Prior to 1955.

	Camel	Chesterfield	Old Gold	Viceroy	Lucky Strike	Kent
Taste	25.00% (2)	32.21% (1)	14.52% (3)	11.11% (3)	17.61% (3)	
Cig Con	11.27% (4)	14.09% (4)	22.58% (1)	32.10% (1)	29.58% (1)	15.00% (3)
Health	29.41% (1)		22.58% (1)	24.69% (2)	21.13% (2)	35.00% (1)
Pleasure		22.15% (2)	17.74% (2)			10.00% (4)
E B/S	13.24% (3)	16.11% (3)	14.52% (3)		16.90% (4)	10.00% (4)
P / A				9.88% (4)		
Red TN						20.00% (2)
Other	21.07%	15.43%	8.07%	22.22%	14.78%	10.00%
Total	100%	100%	100%	100%	100%	100%

(Rankings Are Given In Parentheses.)

Legend

TN Figs = Tar & Nicotine Figures
Cig Con = Cigarette Construction
P / A = Price/Availability
E B/S = Exhortation to Buy/Slogan
Red TN = Reduced Tar & Nicotine

(Old Gold, Lorillard and Lucky Strike, American Tobacco) incorporated health claims in their advertising in the 1920's. Camel, Chesterfield and Viceroy began to utilize such claims in the 1930's. Calfee [5, 7] has suggested that advertising which included "less unhealthy" claims (suggestions that the advertised brand is less unhealthy than a competitor's) would be a more advantageous strategy for smaller firms and that smaller firms would be likely to engage in such adveritising earlier than larger competitors. This study neither confirms nor refutes this suggestion since health claims were present in the early advertisements sponsored by both American Tobacco and Lorillard, a large and small firm respectively. Because of the small number of ads available in the twenties and thus included in this study, it is impossible to establish which of these companies incorporated these claims first. For Kent, introduced in 1952, such claims represented the central focus of its introductory advertising campaign. Only one brand, Chesterfield, failed to utilize health claims frequently. For all of the remaining brands, health claims comprised either the most or the second most frequently used category.

After 1955, taste and construction claims, together with information associated with tar and nicotine, were most common for five out of the six brands examined. Chesterfield was again exceptional in that tar and nicotine information was not presented by this brand until the inclusion of such information was mandated by the FTC. See Table 3.

TABLE 3. Most Frequently Made Claims (Category as a Percent of Total Claims Per Brand) 1955 and After.

	Camel	Chesterfield	Old Gold	Viceroy	Lucky Strike	Kent
Taste	18.87% (2)	29.63% (1)	37.50% (1)	26.53% (1)	24.66% (1)	21.05% (2)
Cig Con		29.63% (1)	37.50% (1)	19.39% (3)	23.29% (2)	18.05% (3)
Pleasure		24.07% (3)				
E B/S	21.70% (1)	5.56% (4)			16.44% (3)	
TN Figs	16.98% (3)			22.45% (2)	16.44% (3)	22.56% (1)
Red TN	10.38% (4)		25.00% (3)			
S Gen W				11.22% (4)		11.28% (4)
Other	32.06%	11.10%	0.00%	20.41%	19.18%	27.07%
Total	100%	100%	100%	100%	100%	100%

(Rankings are given in parentheses)

Legend

TN Figs = Tar & Nicotine Figures
Cig Con = Cigarette Construction
S Gen W = Surgeon General's Warning
E B/S = Exhortation to buy/slogan
Red TN = Reduced Tar & Nicotine

Health Claims

Of the health claims utilized in the various brand advertisements prior to 1955, those associated with throat irritation, coughs, the irritating properties of cigarettes, and protection from adverse side effects were most common. In the forties and early fifties, it became a common practice to cite "scientific" or "medical" research "substantiating" these types

of claims. In the early fifties the notion of health risk reduction was not an uncommon one. It was the fourth most frequently used health claim in the first years of that decade. See Table 4. Lorillard, makers of Kent, incorporated many of these assertions in a 1953 advertisement: "Here's how you can get the protection you need against nicotine and tars--and the smoking pleasure you want...if you're a sensitive smoker--and published medical reports show that 1 out of every 3 smokers is unusually sensitive to the tars and nicotine in tobacco--you can now enjoy real health protection with every cigarette you smoke...If smoking dulls your sense of taste, gives you a 'raw' throat or 'bunched-up' nerves, chances are you're sensitive to nicotine and tars. So, for your own health, as well as pleasure, you should try KENT" (Time 1953).

TABLE 4. Four Most Frequently Made Health Claims (Category as a Percent of Total Health Claims Per Period) For All Six Brands.

	26-29	30-39	40-49	50-54
Throat Irritation	16.67% (1)	21.67% (1)	27.78% (1)	22.45% (2)
Protection	16.67% (1)	11.67% (2)	8.33% (4)	12.24% (3)
Coughs	16.67% (1)	10.00% (3)		
Irritation	8.33% (4)		11.11% (4)	
Risk Reduction				10.20% (4)
Citation of Scientific Findings			27.78% (2)	26.53% (1)
Relaxing		10.00% (4)		
Other	41.66%	46.66%	25.00%	28.58%
TOTAL	100%	100%	100%	100%

(Rankings For The Period Are Given In Parentheses.)

An examination of Table 5 suggests that, with few exceptions, there is little to distinguish the various brands in terms of the health claims frequently used. Five out of six brands referred to throat irritation; risk reduction or protection was a frequent theme with five of the six; four out of six referred to "scientific" findings. Camel was unique, however, in that it presented brand endorsements by medical personnel.

The actual placement and/or prominence of health claims within the advertisements was also of interest. Of the total health claims made, 23%

TABLE 5. Most Frequently Made Health Claims (Category as a Percent of Total Health Claims Per Brand) Prior to 1955.

	Camel	Chesterfield	Old Gold	Viceroy	Lucky Strike	Kent
Throat Irritation	21.67% (1)		25.00% (1)	35.00% (2)	23.33% (2)	14.29% (2)
Protection			17.86% (2)		26.67% (1)	28.57% (1)
Coughs					16.67% (3)	
Irritation		25.00% (1)	17.86% (2)		10.00% (4)	14.29% (2)
Risk Reduction		16.67% (3)		15.00% (3)		14.29% (2)
Citation of Scientific Findings	21.67% (1)	25.00% (1)	17.86% (2)	10.00% (4)		
Relaxing	11.67% (3)	8.33% (4)				
Effects on Organs Other Than Throat				40.00% (1)		
Stimulant	8.33% (4)					
Endorsement by Medical Personnel	8.33% (4)					
Other	28.33%	24.99%	21.42%	0.00%	23.33%	28.56%
Total	100%	100%	100%	100%	100%	100%

(Rankings are given in parentheses.)

were presented in the headline, 31% were presented in subheads, and 46% were presented in the copy. Prior to 1955, brands did exhibit differences in the placement and/or prominence of health claims within these ads. Chesterfield and Kent placed the majority of health related claims in the copy, while the four remaining brands (Camel, Old Gold, Viceroy and Lucky Strike) accorded the majority of such claims the prominence associated with headlines and subheads. See Table 6.

TABLE 6. Placement of Health Claims within Advertisements (Location as Percent of Total Health Claims Per Brand) Prior to 1955.

	Camel	Chesterfield	Old Gold	Viceroy	Lucky Strike	Kent
Headline	18.33%	8.33%	35.72%	25.00%	23.33%	28.57%
Subhead	33.33%	16.66%	25.00%	30.00%	53.33%	0.00%
Copy	48.33%	75.00%	39.20%	45.00%	23.33%	71.43%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

After 1955, health claims were very seldom made (2.3% of all claims made) and when they were, over 70% were presented in the copy. The Surgeon-General's warning and tar/nicotine figures were presented, almost without exception, in copy. Claims associated with a reduction in tar and/or nicotine however, were most often presented in the headline or subhead.

Visuals

Visuals here refers to pictorial or photographic representations of subject matter other than cigarette packages or cartons. It was assumed that manufacturers would include representations of their products. In particular, the age of the actor(s), the activity engaged in, implied social or professional success and sexual connotations were of interest.

For the set of advertisements as a whole, actors perceived as being in the 25-35 age group were most often featured. They constituted 57.89% of those shown, while 27.89% were identified as being 36 years of age or older, and 14.21% were categorized as younger than 25. The majority of actors were engaged in activities which imply physical vigor and were portrayed as being socially and professionally successful. The absence of sexual connotations is notable. Only 5.19% of the visually implied claims were of a sexual nature. Eighteen percent of the advertisements utilized either an athlete or celebrity endorsement.

Some changes in the visually implied claims can be noted pre- and post-1955. In the period prior to 1955, actors were almost exclusively perceived to be in the 25-35 age group. Post-1955, the number of actors perceived to be 36 and older was about equal to those in the 25-35 age group. Thus, featured actors appeared to be older post-1955. Another interesting difference is the portrayal of professionally successful individuals. Prior to 1955, actors and settings very often implied professional success. After 1955, this is not the case. Social, not professional, success became the dominant theme. In both periods, actors portrayed as participants in vigorous, physically demanding activities was common. However, almost 90% of the total athlete and celebrity endorsements occurred prior to 1950.

DISCUSSION

While this study is not one of consumer perception or information processing, it does identify the dominant advertising themes associated with several cigarette brands and examines them over time. It has been demonstrated that until 1955, health claims were a consistent and important consideration in the advertisement of the brands analyzed here. Moreover, in the years following 1970, the mandatory disclosure of tar and nicotine yields based on official tests, the Surgeon-General's warning, and claims regarding reduced tar and/or nicotine were prevalent. It was demonstrated that in the period from 1955 to 1969 there was a paucity of commercial information relating to health. During this period taste claims were most frequently made.

Thus, the findings of this study are somewhat at odds with the previously quoted statement by Tye [22, p. 30] which asserts that "for 60 years, cigarette firms have used unfounded health claims to encourage people to smoke..." This study suggests that until the imposition of the Federal Trade Commission (FTC) advertising guides in 1955, the manufacturers of the brands examined here did indeed use health claims in their advertising. After 1955 however, the only statements made in conjunction with health were those mandated by the FTC and the Congress or were related to this required information. In both instances one presumes that the information is of a factual nature. Thus, criticism of the presence and veracity of health claims in cigarette advertising may well be limited to the period before 1955.

While Tye's [22] indictment of all health claims, comparative or otherwise, as false may not be warranted, an examination of the veracity of representative claims made in the period before 1955 is beyond the scope of this paper.[7] However, as early as 1938 Consumer Reports made the following statements in its first test report on cigarettes [3, p. 144]. "There is little difference perceptible between the various brands of any one type of cigarette...none of the popularly advertised brands appears to be more harmful than any of the others..." In 1942, Reader's Digest came, essentially, to the same conclusion, "no single brand is so superior to its competitors as to justify its selection on the ground that it is less harmful" [15, p. 5]. Likewise, the federal government held that among the leading brands there was no meaningful difference in nicotine content or in any other important smoke component and no significant difference in the physiological effect of smoking them [8]. Thus, based on the thinking of the day, comparative health claims appeared to be false and were subsequently prohibited by the FTC.

Consideration of the consumer interpretation of the health related information found in these ads is also beyond the scope of this study. However, Calfee [5, 7] has suggested that brands claiming to be less unhealthy draw attention to the harmful aspects of smoking. Consistent with this view, the present study found that until 1955, ads that incorporated health claims typically

---

[7]Evaluating the veracity of claims made more than 30 years ago could be somewhat difficult in that such evaluations would be based on the current understanding of cigarettes and health rather than the state of understanding at the time the claims were made. On the other hand, a better appreciation of measurement techniques and threshold quantitities of dangerous substances contained in cigarette smoke might indicate important differences where none were previously thought to exist.

emphasized the harmful effects of the product class in an effort to make the featured brand appear less harmful by comparison. It is Calfee's [5, 7] conclusion that in the presence of such advertising claims, it was probably impossible for any smoker, or potential smoker, to ignore or forget that cigarettes lead to throat irritation, smoker's cough, and other symptoms frequently enumerated in the cigarette advertising of the period. It was this constant "screaming at the top of their lungs about nicotine, cigarette hangovers, smoker's cough, mildness and kindred subjects" in advertising that tobaccomen blamed for the "first sustained drop in cigarette consumption in 20 years" [4, p. 66, 20, p. 100]. One tobaccoman was quoted as saying, "The public is being frightened from tobacco by [the] outlandish medical claims [made] by some of the [cigarette] manufacturers" [20, p. 100]. Business Week [4, p. 68] raised the obvious question, "Why has the [cigarette] industry persisted in this 'negative' form of advertising even when, as tobacco growers and others complain, it hurts the trade by making people conscious that cigarettes can be harmful?" Calfee [7] provides one answer: advertising that appealed to smokers' fears could indeed reinforce those fears and thereby suppress marketwide demand, but such advertising could also divert sales to the advertised brands. Apparently, those brands which utilized such health appeals perceived their economic interests as different from those of the entire industry. That is, they were willing to sacrifice total industry sales in order to increase the sales of their individual brands.

Calfee's [5, 7] work suggests that an empirical investigation of how consumers process "less harmful" claims is warranted. Do such claims, as he suggests, inform users that consumption of the product is harmful and therefore provide information pertinent to the assessment of risk? Did the use of health claims in advertising constitute an implicit, yet effective warning to smokers? If this is in fact the case, our future evaluations of cigarette advertising content and the policies regulating it should focus less on the technical veracity of individual claims and more on the impact of health claims on total product class consumption.

REFERENCES

1. The American Medical Association Board of Trustees. "Media Advertising for Tobacco Products," Journal of the American Medical Association, 255 (8), 1986, p. 1033.

2. Blum, Alan. "When 'More Doctors Smoked Camels': Cigarette Advertising in the Journal," in The Cigarette Underworld: A Front Line Report on the War Against Your Lungs, Alan Blum, ed., Secaucus, New Jersey: Lyle Stuart, Inc., 1985, pp. 107-111, 113.

3. Brecher, Ruth, Edward Brecher, Arthur Herzog, Walter Goodman, and Gerald Walker. The Consumers Union Report on Smoking and the Public Interest, Mount Vernon, N.Y., 1963.

4. Business Week. "Cigarette Scare: What'll the Trade Do?" Business Week, December 5, 1953, pp. 58-68.

5. Calfee, John E. "Cigarette Advertising, Health Information and Regulation Before 1970," working paper, Bureau of Economics, Federal Trade Commission, 1985, pp. 1-79.

6. _____ Bureau of Economics, Federal Trade Commission, personal communication, 1986.

7. _____ "The Ghost of Cigarette Advertising Past," Regulation, Nov/Dec, 1987, pp. 35-45.

8. Consumer Reports. "Cigarettes, What CU's Tests Showed...the Industry, Its Advertising...and: How Harmful Are They?" Consumer Reports, 18 (February 1953), pp. 59-74.

9. Doron, Gideon. The Smoking Paradox, Cambridge, Massachusetts: Abt Books, 1979.

10. Fox, Stephen. The Mirror Makers: A History of American Advertising and Its Creators, New York: William Morrow and Company, Inc., 1984.

11. Gallup, George H. "New Year Resolutions," The Gallup Poll: Public Opinion 1935 - 1971, 1, 1939, p. 133.

12. _____ "Cigarettes", The Gallup Poll: Public Opinion 1935 - 1971, 2, 1949, p. 874.

13. Garner, Donald W. "Cigarette Dependency and Civil Liability: A Modest Proposal," Southern California Law Review, 53 (1371), 1980, pp. 1423-1465.

14. Kassarjian, Harold H. "Content Analysis in Consumer Research, "Journal of Consumer Research, 4 (June 1977), pp. 8-18.

15. Littell, Robert. "Cigarette Ad Fact or Fiction," Reader's Digest, 41 (July 1942), pp. 5-8.

16. Miles, Robert H. Coffin Nails and Corporate Strategies, Englewood Cliffs: Prentice-Hall, Inc., 1982.

17. Pollay, Richard W. "The Subsiding Sizzle: A Descriptive History of Print Advertising, 1900-1980," Journal of Marketing, 49 (Summer 1985), pp. 24-37.

18. Publishers' Information Bureau/Leading National Advertisers. Magazine Advertising Expenditures, New York: Publishers' Information Bureau, all years extant.

19. Taylor, Peter. Smoke Ring: The Politics of Tobacco, London: The Bodley Head, 1984.

20. Time. "Cigarette Hangover," Time, November 9, 1953, p. 100.

21. Tye, Joe B. "Cigarette Marketing: Ethical Conservatism or Corporate Violence?" New York State Journal of Medicine, July, 1985, pp. 324-327.

22. _____ "Cigarette Ads Reveal a History of Deceit," The Wall Street Journal, August 5, 1986, p. 30.

23. Ulrich's International Periodical Directory. 24th Edition, New York: R.R. Bowker Company, 1985.

24. Warner, Kenneth E. "Tobacco Industry Response to Public Health," Health Education Quarterly, 12 (2), 1985, pp. 115-127.

THE QUALITY OF ECONOMIC MICRODATA: THE GOOD, THE BAD AND THE UGLY

Greg J. Duncan and Williard Rodgers, University of Michigan[1]

ABSTRACT

Two validation studies of economic microdata and related work are used to speculate on the likely quality of such data and to develop guidelines for researchers using such data. Data quality is found to be highly variable, but appears to depend systematically on the salience of the topic and recall task. Lessons about the likely effects of measurement error presented in econometrics texts are not very useful.

During the 1950s and 1960s the empirical work of economists was dominated by analyses of aggregate time series data. Concerns over measurement error in such data were understandably relegated to a back burner, since it could be argued that aggregation eliminated most of the offsetting micro-level errors and the time series focus on change made persistent measurement biases relatively unimportant. When compared with the statistical problems of identification, multicolinearity and serial correlation inherent in most time-series analyses, measurement error quite properly was considered relatively unimportant.

Micro-level analyses, usually performed on data gathered from sample surveys, were carried out infrequently in the late 1960s but had gained a position of dominance by the middle 1980s.[2] The agenda of statistical problems has changed accordingly, with far less concern over problems such as multicollinearity and far more for issues like sample selection bias. Problems of measurement error in micro-level data, long a concern in other social sciences, have been blissfully ignored in most micro-level economic studies. No longer able to argue that aggregation eliminates most measurement error, micro-based economic analysts have retreated to the position that, yes, measurement errors probably exist in micro-data, but until we know more about their nature there is little we can do about them.

Getting a handle on the measurement properties of variables typically used in micro-level behavioral data is, at best, difficult and expensive, and, at worse, virtually impossible. In some cases, as with health, crime and voting behavior, it is possible to use independent sources of information to measure the accuracy of some of the survey reports. When validation against records is impossible, a second-best strategy has been to use multiple reports by different respondents or by the same respondent at intervals of several weeks or months (e.g. Bielby and Hauser, 1977) to measure the reports' reliability. Low test-retest reliability clearly points to problems; high reliability is consistent with either the absence of problems or the presence of persistent over- or underreporting.

Luckily for analysts of economic and demographic behavior, such behavior is often, at least in theory, subject to independent validation. But, with the prominent exception of work done in the 1950s and 1960s by Ferber, Maynes and others on wealth data, by Neter and Waksberg on expenditure data, and a few scattered attempts on other topics, little effort has been devoted to investigating the measurement properties of economic micro-level data.

Fortunately, this situation seems to be changing. Two substantial validation studies, of which one (the PSID Validation Study) focused on the quality of employment information and the other (the Study of Michigan Generations) on a range of household demographic and economic information, have recently been completed and offer many lessons about the quality of micro-level data. Insights gleaned from these two studies are reviewed in this paper, along with results from a handful of other methodological projects that provide information on the quality of economic microdata.

Our focus on these two studies forces us, for reasons of space, to be highly selective in reviewing the large body of methodological work that has preceded them. These two studies deserve special attention, however, because they are quite recent and the quality of the independent validation information to which they had access is unusually high. This greatly enhances their usefulness for quantifying measurement error, although it also limits the range of measures for which measurement properties could be established. We have organized the results according to their implications for micro-level data analyses, ranging from point-in-time description to more model-based analyses.

Can respondents report reliably about qualitative aspects of their current situation?

Tables 1 and 2 provide evidence on errors in reports of a wide range of measures of current (i.e., time of interview) fringe benefits. The diversity in the amount of error is striking. Only one percent of the respondents to the PSID employment validation survey gave erroneous reports of medical insurance coverage, but ten percent misreported life insurance coverage (Table 1). Knowledge of the details of health insurance coverage was investigated in a different validation study designed expressly for that purpose and reported by Walden et al. (1984). Pressing for such details produces higher error rates, ranging from 12 percent for in-patient surgery to 71 percent for ambulatory psychiatric care (Table 2).[3]

---

[1]Program Director and Associate Research Scientist

[2]In reviewing labor-related articles from eight leading economic journals, Stafford (forthcoming) found that the respective fractions using time series and micro-level data changed from 42 and 11 percent in the 1965–69 period to 16 and 46 percent in the 1980–83 period.

[3]All respondents in the PSID validation study sample were eligible for medical, dental, life insurance and pension benefits. Reporting errors were possible only if respondents didn't think they were eligible. Eligibility was much more variable in the Walden et al. study and it should be viewed as a better estimate of validity for common benefits such as dental coverage.

Only 3 percent of the employment study respondents failed to give accurate reports of pension coverage, but again, when pressed for details such as vesting and early retirement provisions, error rates jump substantially. A safe generalization would appear to be that reports of the general aspects of salient current circumstances are well-reported. Pressing for detail dramatically increases the likely error, as does asking about less salient conditions.[4]

Table 3, based on the Study of Michigan Generations, shows similar patterns. Birth dates, while obviously salient, may be misreported if respondents want to present themselves as younger (or, perhaps, in some cases older) than they really are. The study found reassuringly little error in such reports. Reports of several automobile-related measures were somewhat more problematic, although probably reliable enough for most analytic purposes.

How about quantitative aspects of their current situation?

Here it is useful to distinguish between a measure's bias and its error variance. If respondent reports agree on average with validation sources, even if that agreement is the fortuitous result of offsetting individual errors, then there is no bias. Discrepancies between individual interview reports and "truth" as revealed by validation sources produce error variance that may or may not be accompanied, in the aggregate, by bias. Bias tends to be more serious for simple descriptive analyses; error variance and covariance are more important for regression-based modeling.

Consistent with the notion that the more detail sought, the worse the survey report, Table 4 shows substantially more error — some bias and substantial error variance — in reports on various quantitative aspects of the respondent's current situation in the Study of Michigan Generations data than was the case for the qualitative evidence shown in Table 3.

Respondents overestimated their house values, on average, by a slight amount (3 percent, after an adjustment for an apparent bias in the assessors' records) and their property taxes by a much larger amount (24 percent). Although the most salient neighborhood characteristic — racial composition — is reported with very little bias, less salient characteristics have more, with respondents underestimating the proportions of their neighbors with low income, overestimating the proportion with high incomes and substantially overestimating the fraction of neighbors who are elderly.

Although bias may be small in some cases, the error variance is usually much larger, as indicated by the average absolute difference between interview and record reports. These average discrepancies were close to 20 percent or more in all measures but racial composition.

Do respondents give reasonably unbiased reports of past flows of income and employment-related measures?

Table 5 summarizes bias and error variance in reports of past calendar years' amounts of earnings, work hours, unemployment and average hourly earnings gathered in the PSID validation study. Generally there appears to be little bias in measures, even if recalled in the June, 1983 interview for the 1981 calendar year, but substantial error variance. Not surprisingly, earnings showed the least bias, perhaps due to their salience from W-2 forms, tax returns, etc. Work hours were significantly underreported, on average, perhaps because of difficulty in recalling episodes of overtime work.

Bias in survey reports of some stock and flow information gathered from a representative sample can also be ascertained by checks against aggregate information, as with the Avery et al. (1987) and Curtin et al. (1987) preliminary comparisons of wealth data from the 1983 Survey of Consumer Finances with flow-of-funds aggregates; the comparisons by Duncan, Hill and Ponza (1984) of PSID transfer income reports with aggregates expenditure totals; and the U.S. Census Bureau's (1983) comparison of transfer and employment income reports from the Current Population Survey and the new Survey of Income and Program Participation with similar aggregate expenditure totals. With the assumption that nonrespondents are similar to respondents, these studies typically find evidence of underreporting that is relatively minor in some cases (transfer measure reports in SIPP and the PSID, mortgage debt, corporate stock and mutual funds in the Survey of Consumer Finances) and major in other cases (earnings in both SIPP and CPS, installment debts in the Surveys of Consumer Finances).

Error variance, as indicated either by the average absolute deviation between interview and record report (Table 5), or by the ratios of error variance to true variance (Table 6), is highly variable.[5] For example, the error-to-true variance ratio (.011) shown in the last row of the first column of Table 6 indicates that tenure with the company is reported with very little error. The extent of true variation in tenure was quite large (the standard deviation was 15.9 years and the range was from 1 to 42 years), and the resulting ratio of error-to-true variance was very small.

Reports of annual earnings and unemployment for the prior calendar year also have relatively small error-to-true variance ratios (.154 and .129, respectively). Curiously, the natural logarithm of annual earnings has a considerably higher ratio than the earnings measure expressed in dollars. This was due in large part to four cases for which very low earnings were reported in the interview and for which measurement error was quite high. Removal of these cases dropped the error-to-true variance considerably.

Validation of reports of work hours was possible only for hourly workers. There was relatively high error variance in the reports of annual work hours. This is not surprising given the complex task facing the respondent in reporting the components of this amount, especially an hours-per-week figure that is averaged over the entire year. Nor is there anything equivalent to the W-2 statement of earnings that summarizes annual work hours.

Errors in interview reports of average hourly earnings (defined as the ratio of interview reports of annual earnings to annual hours) were enormous. The ratio of error-to-true variance was 2.8 for 1982 and 1.8 for 1981. Several factors, spelled out in Duncan and Hill (1985), argue against the likely precision of these estimates for the entire working population but not against the proposition that reports of average hourly earnings have exceedingly small signal-to-noise ratios.

---

[4]See Cannell et al. (1981) for an excellent summary of evidence, mostly from the health area, on how salience, time and social desirability affect survey responese.

[5]As shown below, the ratio of error to true variance is a useful measure that, under certain conditions, represents the extent of bias in an error-ridden measure's coefficient if it were included as an independent variable in a multiple regression.

The ratios of error-to-true variance were generally much higher for calendar year 1981 amounts than for 1982 amounts. The jump was most dramatic for unemployment (from .129 to .518) and for work hours (from .336 to .919). For hourly earnings, however, 1982 reports had higher error-to-true variance ratios than 1981 reports. This appeared to be due to the greater tendency in 1981 for the errors in reports of earnings and hours to offset one another.

Can one reduce measurement error by removing apparent outliers?

Analysts often examine their data for extreme cases before using them to describe or model the behavior they represent, believing that such a procedure will purge their data of the most erroneous cases and improve the consistency of their estimates. Duncan and Hill (1985) identified outliers in the univariate distributions of the interview data using a rule of five standard deviations on either side of the mean, with exceptions made when a judgment of what an analyst might do under the circumstance dictated something else.

With the exception of hourly earnings, removal of these extreme cases produced the surprising result that the error-to-true variance ratios generally increased. A closer look at the data confirmed that the interview reports of outliers identified in this way did not have large amounts of error. Their removal reduced the true variance a great deal and reduced the error variance by a smaller relative amount. On balance then, the error-to-true variance increased in most cases. Of course, these cases may have been outliers in a conceptual sense (e.g. when earnings fell to zero if a self-employed businessman had a bad year), and their removal may improve the fit of a model designed to explain only "conventional" behavior. Whether the analyst should choose to handle such conceptual outliers by deleting them or by reformulating the model to cover the behavior they represent is an open question. Validation studies suggest the importance of pondering such questions in the development and estimation of models.

What about recall of time use?

Social scientists from several disciplines have turned to time as a useful unit for organizing analyses of economic and social activities. The work hours and unemployment reports reviewed in the previous section are, of course, examples of time use. But what if one wants a more complete accounting of time allocation? Reliable reporting of patterns of time use has been the subject of intensive methodological work in the past decade, which, although never based on direct validation techniques, does provide a useful body of evidence from which some lessons can be drawn.

Several different approaches have been taken to elicit reports of how respondents spend their time. The most direct is simply to ask questions such as "How much time did you spend doing housework (work, childcare, television viewing, etc.) last week (month, etc.)". More cumbersome methods involve leave-behind diaries and 24-hour recall diaries in which respondents begin at midnight of a given day and report the sequence of these activities over the next 24 hours. Robinson (1986) compared responses to the "How much time. . ." questions and the 24-hour recall method with the "truth" as revealed by reports when a randomly programmed beeper signalled and respondents were instructed to mark down their activities at that point. He found clear evidence that the "How much time. ." questions were affected by the social desirability of the activity. Housework and childcare were substantially overreported; the less direct recall methods did not shown a similar bias.

Apart from the issue of social desirability, recall of time use would appear to place heavy burdens on respondents' memories, especially for less salient activities. Studies of the quality of the 24-hour recall method justify this fear, showing that data quality declines if the period recalled is several days in the past (Juster, 1986). Not surprisingly, the more salient weekend days are recalled more accurately than weekdays. Also worthy of note is that techniques for stimulating recall appear to improve the quality of reporting.

In sum, data on time use activities that are distant, irregular or have elements of social desirability or undersirability are likely to be suspect, as are respondents who have not been properly motivated to do the cognitive work necessary to provide the required information.

How good are retrospective event-history data?

The above analysis indicates that respondents are able to provide error-prone but generally unbiased reports of annual amounts. Panel data, however, are increasingly being used to analyze event histories where precise timing of transitions is essential. To obtain this level of precision in timing it is necessary to ask respondents to date occurrences retrospectively since the prior interview.

It has long been known that respondents often tend to "telescope" more distant events by reporting them as occurring more recently than they actually do. With a panel design, one can "bound" the recall of such events using information collected in prior interviews (Neter and Waksberg, 1965). More problematic is accumulating evidence that respondents simply forget events, even presumably quite salient events such as an overnight stay in a hospital. For example, Cannell and Fowler (1963) found that more than one-quarter of hospitalization episodes that appeared in hospital records were not reported in interviews with the ex-patients if they occurred one year or more prior to the interview. Neter and Waksberg (1965) found about half as many repair expenditures of $20 or less reported if they occurred three months as opposed to one month prior to the interview. Larger expenditures didn't appear to produce substantial memory decay, indicating once again the importance of salience.

The PSID validity study also provides valuable evidence on the quality of episodic recall of unemployment. Table 5 showed that respondents provided relatively accurate reports of overall annual unemployment amounts. A second way of casting information about the unemployment experiences of the respondents, however, is in terms of spells. Company records gave the precise dates on which all spells of unemployment among its workers began and ended. The information obtained in the interviews was not as precise, dating unemployment only to the month in which it occurred. Nevertheless, it was almost always possible to determine whether the respondent reported accurately each spell of unemployment that appeared in the company records.

Table 7 (taken from Mathiowetz, 1986) shows the performance of respondents in reporting unemployment spells of various lengths. It is obvious that respondents have great difficulty in recalling spells of unemployment, especially short ones. Two-thirds (66 percent) of all spells of unemployment that appeared in the company records were not reported in the interview. Even long spells (lasting more than 5 weeks) were seriously underreported; the fraction not reported was about one-half. And only one-fourth of very short spells, lasting up to one week, were reported in the interview. These findings contrast sharply with the relatively greater reliability of the more aggregate measures of annual unemployment hours, discussed above, and

argue for extreme caution in analyzing spell-level information of retrospective reports of unemployment. Other kinds of event-history measures, such as those focused on demographic events like births or marriages, may be more reliable because of the greater salience of these events to respondents.

What statistics books tell you about how error variance affects model estimation

In the simple bivariate regression case, dependent (Y^*) and independent (X^*) variables are typically assumed to be related in the following way:

$$Y_i^* = \beta_0 + \beta_1 X_i^* + \epsilon_i \quad (1)$$

where Y_i^* and X_i^* are measured without error, ϵ_i is distributed with mean zero and variance σ_ϵ^2, and with $\mathrm{cov}(X_i^*, \epsilon_i) = 0$.

Measurement error in the dependent variable in this model that is uncorrelated with Y_i^*, X_i^*, and ϵ_i will increase the standard errors but will not bias the estimates of β_0 or β_1. As we shall soon see, although the absence of such correlations is almost universally assumed, validation studies frequently show substantial correlations between measurement errors in measures frequently used as dependent variables (e.g. earnings) and both the true values of those measures and the true values of typical right-hand side measures.

Measurement error in the X variable almost always produces bias in the estimate of β_1. If the measured independent variable X is related to its true value (X^*) according to:

$$X_i = X_i^* + u_i \quad (2)$$

and if u_i is uncorrelated with Y_i^*, X_i^*, and ϵ_i, then the probability limit of the estimate of β_1 is:

$$\mathrm{plim}\,\beta_1 = \beta_1 - \frac{\beta_1 \sigma_u^2}{\sigma_{X^*}^2 + \sigma_u^2} = \frac{\beta_1}{1 + \frac{\sigma_u^2}{\sigma_{X^*}^2}} \quad (3)$$

Measurement error in the Y variable does not effect this plim formula as long as we adopt the far from innocent assumptions that measurement error is independent of X*, u, and ϵ.

Given those assumptions, the crucial feature of the measurement error of X is the ratio of the error variance to the true variance of X, $\frac{\sigma_u^2}{\sigma_{X^*}^2}$. For example, if this ratio is .5, then the true value of β_1 will be 50 percent higher than its value estimated with the erroneous measure of X. The figures reported in Table 6 are precisely these error-to-true variance ratios. Rodgers and Herzog (1986) develop the more general case where u_i and measurement error in Y_i are allowed to be correlated with Y_i^*, X_i^* and ϵ_i. Virtually nothing can be said about the extent of bias in β without knowledge of these covariances.

What does measurement error really do to regression relationships?

As developed above, conventional wisdom holds that measurement error in the dependent variable biases standard errors but not regression coefficients, while measurement error in independent variable biases regression coefficients to zero. Duncan and Hill (1985) explored some of the implications of the measurement error observed in the validation data with models that related annual earnings to tenure, both of which could be validated, and education, for which validation was not possible. They found that the estimated payoff to tenure was roughly 30 percent lower if interview reports of annual earnings and tenure were used to estimate the relationship than if the company records were used. The source of the bias, however, was not in the amount of measurement error in the tenure measure, because, as shown in Table 6, tenure appears to be extremely well-reported in the interview. Rather, it can be traced to a negative correlation between the measurement error in the report of earnings and the true level of tenure. In the language of the statistical models, there was a strong negative covariance between measurement error in Y_i and the level of X_i. Older workers tended to underreport earnings while younger workers overreported them somewhat.

Such correlations are typically ignored in most measurement error models, and, contrary to conventional wisdom about the effects of measurement error, could bias estimated coefficients either toward or away from zero. In this case, the negative correlation caused a downward bias. But had younger workers been the underreporters, the tenure coefficient would have been biased upward. Conventional wisdom, which holds that measurement errors bias regression coefficients downward, would simply have been wrong in this case.

Rodgers and Herzog (1986) take a more general look at measurement error covariances using data from the Study of Michigan Generations. For several of the measures they consider they find substantial correlations between both the signed and absolute amounts of measurement error and the "true" values of those measures. Correlations between measurement errors and demographic variables were mostly small, although substantial biases on regression coefficients were occasionally introduced by these correlations. Finally, they found that measurement errors tended not to be correlated with with one another, particularly when the variables were not in the same topic area.

What is a poor microdata researcher to do?

It is clear that measurement error is an important phenomenon in economic microdata and that, under certain circumstances, it can seriously affect estimates of simple statistics such as means as well as regression coefficients. It is also clear that these effects can be highly variable. Some measures, especially the most salient, least threatening and easiest to recall, tend to be reported quite well. Others, with less salience or occurring further in the past, are more problematic.

It is crucial for researchers to know as much as possible about the measurement properties of their microdata. An obvious but not always very helpful bit of advice is to restrict one's analysis, when possible, to valid measures. A more realistic position in which researchers find themselves is one of trading off the benefits of good measurement with the costs of omitted variable bias. Omitting a crucial but poorly measured variable may bias coefficients on the included variables more than including it.

One solution to this problem is to build information about measurement properties into the regressions. If one knows that an independent variable has a certain amount of measurement error that is uncorrelated with its own true level, the true level of other variables in the model and other measurement errors, then adjusting that variable's coefficient for the measurement is quite straightforward, using the formulae developed above. Knowledge of error covariances, obtained from validation studies such as those summarized in this paper, can also be incorporated into the regression estimates to adjust for the effects of the measurement errors. It is unlikely that information about all such covariances will be available to the researcher, so the optimal strategy may be one of building in the known covariances and examining the sensitivity of the estimates to assumptions about the unknown ones. Programs such as LISREL are designed to estimate flexible structural models with measurement error variance and covariance.

Sensing yet another opportunity to plead for further research, we conclude with just such a plea. Measurement error problems do not disappear when they are assumed away in econometric models. The binding constraint in advances in dealing with these problems is factual knowledge about the nature of measurement error in the kinds of microdata that we now use routinely. Validation studies are expensive, tedious, and absolutely crucial.

TABLE 1. Error Rates in Reports of Union Status and Several Fringe Benefits

Question	Percent Invalid
Is your current job covered by a union contract?	1
Do you belong to that union?	1
Do you have medical, surgical or hospital insurance that covers any illness or injury that might happen to you when you are not at work?	1
Do you receive sick days with full pay?	9
Are dental benefits provided to you on your main job?	5
Do you have life insurance that would cover a death occurring for reasons not connected with your job?	10
Do you get paid vacation days?	1
Now I need to get some information about any pension or retirement plan you may be eligible for at your place of work. Not including Social Security or Railroad Retirement, are you covered by a pension or retirement plan on your present job?	3
Have you worked under the main or basic plan long enough to earn (the right of vesting?)	11
If you wished to retire earlier (than time needed to receive full benefits) could you receive part but not full benefits from this plan?	28

SOURCE: Duncan and Hill (1985), based on the data from the PSID Validation Study.

TABLE 2. Error Rates in Reports of Own Coverage by Public and Private Insurance for Selected Health Services

Variable	Percent Invalid
Semi-private room in hospital	14
Physician inpatient surgery	12
Other inpatient physician	19
Maternity	45
Eye examination for glasses	27
Routine dental care	22
Orthodontia	31
Ambulatory X-rays and diagnostic tests	30
Ambulatory physician	46
Prescription drugs for ambulatory patients	46
Ambulatory psych. or other mental health care	71
Inpatient mental health	68
Semi-private nursing home or similar facility	67

SOURCE: Walden et al. (1984).

TABLE 3. Error Rates in Reports of Automobile and Birth Data

Variable	Percent Invalid
Auto make	9
Auto year	13
Have a driver's license?	8
Birth date	2

SOURCE: Rodgers and Herzog (forthcoming), based on data from the Study of Michigan Generations.

TABLE 4. Summary of Reporting Error for Reports of House Value, Property Taxes, and Neighborhood Characteristics

Variable	Mean of Simple Difference (Interview-Record)[1]	Mean of Absolute Difference |Interview-Record|
House value (expressed as a percentage difference)	2.9	18.4
Property Tax (expressed as a percentage difference)	24.2*	39.0
Percent of neighbors who		
–are age 60 or older	22.7*	25.2
–are Black	3.5*	6.7
–have incomes over $10,000	–6.7*	18.6
–have incomes over $30,000	8.7*	19.8

[1]Hypothesis test is a two-tailed test for a zero mean.

*p<.05

SOURCE: Rodgers and Herzog (forthcoming), based on data from the Study of Michigan Generations.

TABLE 5. Summary of Reporting Error for Reports of 1982 and 1981 Earnings, Work Hours, Unemployment Hours, and Average Hourly Earnings

Variable	Mean of Simple Difference (Interview-Record)[1]	Mean of Absolute Difference |Interview-Record|
Annual Earnings[2]		
1982	–$55	$2123
1981	–$294	$2567
Annual Work Hours[3]		
1982	90*	157*
1981	115*	211*
Unemployment Hours[2]		
1982	–11	52
1981	–16	45
Hourly Earnings[3]		
1982	–0.63	2.68*
1981	–0.66	2.13*

[1]Hypothesis test is a two-tailed test for a zero mean.

[2]Means for earnings and unemployment based on total sample (n=357).

[3]Means for work hours and hourly earnings are presented for hourly workers only (n=184).

*p<.05

SOURCE: Duncan and Hill (1985), based on data from the PSID Validation Study.

TABLE 6. Ratios of Error Variance to True Variance for Reports of Earnings, Unemployment, Work Hours, and Tenure

	Unaltered		Outliers Removed[1]	
Variable	1982	1981	1982	1981
Annual Earnings	.154	.301	.183 (5)	.325 (2)
ln (Annual Earnings)	.316	.412	.199 (5)	.553 (2)
Annual Unemployment Hours	.129	.518	.214 (6)	.964 (2)
Annual Work Hours (hourly workers only)	.366	.919	.462 (4)	.866 (13)
Average Hourly Earnings (hourly workers only)	2.801	1.835	2.439 (4)	1.495 (4)
Tenure	.011	–	–	–

[1]Number of outliers removed is given in parentheses.

SOURCE: Duncan and Hill (1985), based on data from the PSID Validation Study.

TABLE 7. Percent of Unemployment Spells Unreported by Length of Recall Period and Duration of Unemployment Spell[1]

Duration of Unemployment	Length of Recall Period				
	<8 Months	9–12 Months	13–18 Months	19+ Months	Total
1 week	53% (17)	66% (58)	76% (97)	81% (75)	74% (247)
2 weeks	58% (12)	64% (45)	66% (21)	71% (42)	67% (120)
3–4 weeks	46% (13)	50% (10)	60% (5)	60% (5)	52% (33)
5 or more weeks	40% (5)	41% (56)	63% (8)	61% (18)	47% (87)
Total	51% (47)	56% (169)	73% (131)	75% (140)	66% (487)

[1]Numbers in parentheses are the denominators for the cell percentages. Since some respondents have multiple unemployment spells, the observations are not independent.

SOURCE: Mathiowetz (1986), based on data from the PSID Validation Study.

REFERENCES

Avery, Robert B., Elliehausen, Gregory E. and Kennickell, Arthur B. "Measuring Wealth with Survey Data: An Evaluation of the 1983 Survey of Consumer Finances," mimeo, 1987.

Bielby, W. T. and Hauser, R.M. "Response Error in Earnings Functions for Nonblack Males." Sociological Methods and Research 6, (2), 1977, 241–279.

Cannell, Charles and Fowler, Floyd. A Study of the Reporting of Visits to Doctors in the National Health Survey. Research Report. Ann Arbor: Institute for Social Research, 1963.

Cannell, Charles, Miller, Peter and Oksenborg, Lois. "Research on Interviewing Techniques." In Samuel Leinhardt (Ed.,) Sociological Methodology 1981. San Francisco: Jossey-Bass, 1981.

Curtin, Richard T., Juster, F. Thomas and Morgan, James N. "Survey Estimates of Wealth: An Assessment of Quality". Prepared for Conference on the Measurement of Saving Investment and Wealth, Baltimore, 1987.

Duncan, Greg and Hill, Daniel "An Investigation of the Extent and Consequences of Measurement Error in Labor Economic Survey Data", Journal of Labor Economics, 3, No. 4, October 1985, pp. 80–101.

Duncan, Greg J, and Mathiowetz, Nancy. A Validation Study of Economic Survey Data, ISR Research Report, Ann Arbor, MI: Institute for Social Research, 1985.

Duncan, Greg J.; Hill, Daniel; and Ponza, Michael. May, 1984. "How Representative is the PSID? A Response to Some Questions Raised in the Unicon Report." Mimeographed. Ann Arbor: University of Michigan, Survey Research Center.

Juster, F. Thomas. "Response Errors in the Measurement of Time Use." Journal of the American Statistical Association 81, 394, June 1986, 390–402.

Mathiowetz, Nancy A. "The Problem of Omissions and Telecopying Errors: New Evidence from a Study of Unemployment" in American Statistic Association Proceedings, Social Statistics Section, 1986.

Neter, John and Waksberg, Joseph. Response Errors in Collection of Expenditure Data by Household Interviews. An Experimental Study Technical Paper No. 11, Washington, D.C.: Bureau of Census, 1965.

Robinson, John P. "The Validity and Reliability of Diaries versus Alternative Time Use Measures." In F.T. Juster and F.P. Stafford (Eds.) Time, Goods and Well-being, Ann Arbor: Institute for Social Research, 1986.

Rodgers, Willard and Herzog, A. Regula. "Covariance of Measurement Errors in Survey Responses," Mimeo, Survey Research Center, 1986.

Rodgers, Willard and Herzog, A. Regula. "Interviewing the Elderly: The Accuracy of Factual Information." Journal of Gerontology, forthcoming.

Stafford, Frank P. "Forestalling the Demise of Empirical Economics: The Role of Microdata in Labor Economics Research." In Handbook of Labor Economics, O. Ashenfelter and R. Layard (Eds.), North Holland: forthcoming.

U.S. Bureau of the Census, Current Population Reports, Series P-60, No. 137, Money Income of Households, Families and Persons in the United States, 1981, U.S. Government Printing Office, Washington, D.C., 1983, Table A-2, p. 216.

U.S. Bureau of the Census, Current Population Reports, Series P-70–83–4, Economic Characteristics of Households in the United States: Fourth Quarter 1983, Table D-3, p. 47.

Walden, Daniel C., Horgan, Constance M. and Cafferata, Gail L. "Consumer Knowledge of Health Insurance Coverage." in Charles Cannell and Robert Groves (Eds.) Health Survey Research Methods, Proceedings of the Fourth Conference on Health Survey Research Methods, DHHS Publication No. (PHS) 84–3346, 1984.

THE EFFECTS OF FAMILY SIZE ON RESPONDENT BURDEN: THE CASE OF PROXY REPORTING FOR CHILDREN'S HEALTH STATUS

Ken R. Smith[1] and David A. Huth,[2] University of Utah

ABSTRACT

Many surveys rely on self reports as well as proxy reports. This study investigates the burden imposed on respondents who are asked to provide health information about other individuals and the potential adverse effect this has on the quality of the proxied responses. The general finding is that more proxy interviews tend to be associated with greater underreporting of health conditions for the proxied individuals. This report offers alternative survey designs for proxy interviews.

INTRODUCTION

Daily conversations are largely made up of questions and answers. The answer that is given depends on many factors--for example, the person doing the asking, what is being asked and how the respondent feels when the question is being asked. Similar variables are at play when surveys are administered. A major objective among survey scientists is to ask questions in a reasonable way without generating too much of an inconvenience for the respondent, thereby allowing him/her to provide as accurate information as possible.

This paper focuses on a survey design that is being used with increasing frequency: proxy interviews. With proxy interviews, an individual is asked to provide information for himself and for other individuals as well, typically fellow household members. The first objective of this study is to first determine whether there is any evidence that the reporting behavior of the respondent is adversely affected as the number of proxy respondents increases. The study hypothesis is as follows: the greater the number of proxy respondents, the more difficult it is for a single respondent to provide information about each individual.

The second objective is to examine respondent characteristics in relationship to the respondent's ability to provide accurate answers. Characteristics that are examined here are the education and health of the respondent and whether the respondent is the mother or father. The paper concludes with some possible design options that may minimize possible respondent burden problems.

---

[1]Director, Survey Research Center

[2]Senior Programmer Analyst, Division of Social Science Research

PROXY INTERVIEWS AND RESPONDENT BURDEN

The proxy respondent design may be thought of as a survey technique that affects respondent burden. Respondent burden has been interpreted to mean different things [10]. For example, the Office of Management and Budget (OMB), the agency empowered to monitor the requests for information by the federal government, defines respondent burden in terms of time needed by an individual to complete a questionnaire or request for information. Others have referred to burden in terms of the subjective reaction of the respondent to the interview [9]. More recently, Bradburn [2] elaborated the concept of burden so that it referred to the perception of the respondent along four dimensions: (1) length of the interview, (2) effort required of the respondent, (3) frequency of being interviewed, and (4) the stress experienced by the respondent due to the disturbing nature of the questions. These factors were examined experimentally by Sharp and Frankel [8] who found that among suburban households in Philadelphia, interview length was the most important factor in increasing perceived respondent burden. If an individual is asked to provide increasing amounts of proxy information, it is likely that the respondent burden will increase and reporting to be of lower quality.

PROXY INTERVIEWS, RESPONDENT BURDEN, AND REPORTING BEHAVIOR

If proxy interviews are burdensome, why are they used in numerous survey settings? Cost and sampling error considerations have been the primary justifications for using a proxy interview design. The expense of asking a single adult to provide information about other household members is substantially less than interviewing each household member separately. Furthermore, each interview represents several other proxy interviews so that the cost per interview declines while the effective sample size increases and sampling error declines.

Many survey scientists are rightfully concerned about using proxy interviews and whether it is asking too much of the respondent. After all, respondents often have difficulty answering questions dealing with themselves, let alone others. Recent evidence suggests that while a proxy response may be more useful than a self-report for some types of questions (e.g., stigmatizing

health events, [4], proxy response may lead to underreporting in other instances (e.g., physician contact rates, [5]).

It is surprising to find that the studies seeking to investigate differential responses between proxy and self reports have generally focused on comparisons among adult respondents. For the large majority of surveys, the number of proxy adult respondents is only one or two. It is therefore possible that the potential burden imposed on the respondent is not being adequately studied because total household size is not being considered. That is, an adult proxy response may have come after the respondent had already proxied for other household members.

It is intuitive to think the greater the household size, the greater the burden on the respondent reporting for other household members. From Bradburn's perspective, increasing household size will generally lengthen the interview and add to the amount of effort required to respond and increase perceived burden for the respondent. In some cases, burden may increase when a parent reporting about his or her children becomes distressed or psychologically uncomfortable (Bradburn's fourth measure of respondent burden) if the questions pertain to unpleasant behaviors or health conditions of the children.

A clearer picture of the potential adverse effects of respondent burden may emerge by studying response patterns across a wide range of household sizes. Therefore, the presence of children and information collected about children will be examined in the context of household size because the number of children will generally be larger than the number of adults. The relationship between household size and proxy reporting will be studied in two distinct ways:

(1) Proxy reporting for children as a function of household size.

(2) Proxy reporting for children as a function of household size and characteristics of the reporting adult.

In this study, burden will be measured indirectly by changes in reporting behavior rather than by direct measures of burden which might be based on the respondent's reactions to the interview [8].

MATERIALS AND METHODS

A sample of 6000 Utah households was selected by randomly sampling telephone numbers based on a two-stage cluster sample developed by Waksberg [12] and attributed to Mitofsky. The survey was conducted between February and June of 1986 and represented over 20,000 household members, with a response rate of 81%. This study is based on households that had at least one child under 18 in the home and the respondent was either the mother or father. The sample used in this study included 8105 children from households with a mean of 5.8 household members and 3.8 children per household.

When a sampled household was contacted, one person was identified who would serve as the source of information for all other household members. Any adult 18 years or older who responded that they knew the most about the health of all other household members was selected as the informant for the entire household. Because the respondent is speaking for the entire household rather than just herself, it is appropriate for any knowledgeable and responsible adult (as opposed to a randomly selected adult) to speak about the health of all household members.

The structure of the questionnaire is based on a household-then-person design. This means that the respondent was asked whether anyone in the household had a certain health condition or used a type of health care service in a given period of time. If the answer was yes, the respondent was then asked to identify the particular individual. This follows the design used by the National Health Interview Survey [7].

This design therefore asks respondents to think across potentially many individuals over a period of time. The time, effort, and possibly the psychological discomfort involved in providing answers to these questions can potentially impair their ability to sort through the health events for the past year. This is a potential problem for all types of proxy reporting methods. Several recent studies have shown that proxy interviews for even a single individual, often the spouse, may yield very useful information for some questions while underreporting on other items [1, 5].

For this study, respondent burden is likely to increase as household size increases. The more children in the household, the greater the cognitive demands for recalling whether any one individual experienced a given health event. In addition, the number of children (depending upon their age composition) also may be a crude measure for household activity. It is possible that the larger the number of children, the greater are the distractions to the respondent in terms of immediate and local distractions as well as psychological preoccupation that might reduce the respondent's level of concentration at the time of the interview.

Another measure, the length of the interview, may also be an indicator of the burden imposed on the respondent. Such an effect may be due to the inconvenience imposed on an individual because of the sheer sacrifice of time made by the respondent. With each passing minute, the respondent may become less committed to concentrating and answering each question. Alternatively, longer interviews may represent a greater willingness to cooperate indicating that the respondent is extending the interview because it is enjoyable. One of the major advantages of testing for respondent burden effects attributable to the size of the household in Utah is the relatively

large average household size in Utah. If more proxy reporting does lead to greater under-reporting, it is likely to be observed here because of the greater prevalence of large families. The average Utah household size is 3.2 and nationally it is 2.8.

DESCRIPTIVE RESULTS

The survey asked a knowledgeable adult about 24 different chronic conditions, visits for medical advice, overnight hospitalizations, and assessments of general health status, for all household members (see Appendix A). The relationship between these health variables and household size is shown in Table 1. This table shows that the average number of conditions reported for a given child (the sum of all the conditions numbered 1 to 24 in Appendix A) shows a general decline between households with 4 to 5 members to those with 8 or more. Removing a relatively common condition, influenza, from this total (called "chronic conditions") shows that the negative relationship between household size and the average number of chronic conditions is also observed.

Table 3 reports a similar set of relationships for health care utilization patterns. Overall, the larger the household size, the less likely a child is reported to have seen a doctor for medical advice or have been hospitalized overnight.

Tables 1 through 4 also break down these relationships by age to control for any confounding between household size and the age of the child and their relationship to health reports. Table 1 shows that across age groups, children from larger households appear to have fewer chronic conditions than similarly aged children from smaller households. The exception to this relationship appears to be adolescent children where household size does not have a consistent effect (an exception that is evident in Tables 2 and 3 as well).

Table 1. Average Number of Reported Health Conditions for Children by Age and Household Size

	Household Size				
	2-3*	4-5	6-7	8+	Total
Children Less than 1	0.145	0.349	0.321	0.279	0.293
N	76	172	78	43	369
Children Aged 1 to 4	0.534	0.419	0.441	0.373	0.430
N	178	921	547	241	1887
Children Aged 5 to 12	0.323	0.327	0.305	0.227	0.300
N	201	1506	1513	730	3950
Children Aged 13 to 17	0.229	0.292	0.241	0.322	0.276
N	223	760	551	363	1897
All Children	0.327	0.345	0.320	0.279	0.324
N	678	3359	2689	1377	8103

ANOVA for all 4 household sizes $F_{3,8099} = 3.02$; p=0.03

ANOVA for 3 largest household sizes $F_{2,7422} = 4.50$; p=0.01

*Nearly all are only children.

Table 2. Average Number of Reported Chronic Health Conditions for Children by Age and Household Size

	Household Size				
	2-3*	4-5	6-7	8+	Total
Children Less than 1	0.092	0.244	0.231	0.209	0.206
N	76	172	78	43	369
Children Aged 1 to 4	0.421	0.295	0.289	0.274	0.303
N	178	921	547	241	1887
Children Aged 5 to 12	0.239	0.222	0.200	0.137	0.199
N	201	1506	1513	730	3950
Children Aged 13 to 17	0.153	0.200	0.162	0.223	0.188
N	223	760	551	363	1897
All Children	0.242	0.238	0.211	0.186	0.221
N	678	3359	2689	1377	8103

ANOVA for all 4 household sizes $F_{3,8099} = 3.20$; p=0.03

ANOVA for 3 largest household sizes $F_{2,7422} = 4.25$; p=0.01

*Nearly all are only children.

Table 3. Average Number of Visits to the Doctor for Medical Advice Last Year for Children by Age and Household Size

	Household Size				
	2-3	4-5	6-7	8+	Total
Children Less than 1	3.16	3.02	2.87	2.83	3.00
N	76	172	78	43	369
Children Aged 1 to 4	3.45	2.40	2.22	2.11	2.41
N	178	921	547	241	1887
Children Aged 5 to 12	1.38	1.29	1.20	0.95	1.20
N	201	1506	1513	730	3950
Children Aged 13 to 17	1.03	1.33	1.06	1.10	1.17
N	223	760	551	363	1897
All Children	2.01	1.69	1.43	1.25	1.56
N	677	3354	2689	1375	8103

ANOVA for all 4 household sizes $F_{3,8099} = 21.9$; p=0.00

ANOVA for 3 largest household sizes $F_{2,7422} = 20.2$; p=0.00

*Nearly all are only children.

Table 4. Average Number of Overnight Hospitalizations Last Year for Children by Age and Household Size

	Household Size				
	2-3*	4-5	6-7	8+	Total
Children Less than 1	0.961	1.01	0.821	0.791	0.932
N	76	172	78	43	369
Children Aged 1 to 4	0.084	0.423	0.366	0.108	0.0530
N	178	921	547	241	1887
Children Aged 5 to 12	0.0199	0.0392	0.0278	0.0288	0.0319
N	201	1506	1513	730	3950
Children Aged 13 to 17	0.0448	0.0434	0.0327	0.0441	0.0406
N	223	760	551	363	1897
All Children	0.150	0.0905	0.0536	0.0704	0.0798
N	678	3359	2689	1377	8103

ANOVA for all 4 household sizes $F_{3,8099} = 15.9$; p=0.00

ANOVA for 3 largest household sizes $F_{2,7422} = 8.712$; p=0.00

*Nearly all are only children.

MULTIVARIATE ANALYSES

The negative relationship between reported number of health conditions and household size is complex and does not necessarily represent the adverse effects of response burden. Table 5 shows that several disease-based factors contribute to differences in reported health conditions. Table 5 reports two ordinary least squares equations which regress the number of all health conditions and chronic conditions on several indicators of respondent burden (family size, length of interview, whether the respondent was the mother or father) as well as socioeconomic status (total household income and respondent's education level).

The hypothesized relationship between household size and reports of child health conditions continues to find support after controlling for several confounding socioeconomic and demographic factors. For all conditions and chronic conditions, respondents who are asked to respond for larger numbers of household members tend to report fewer conditions. Could it be that household size represents an alternative effect, one which would lead us to expect that children from larger households should be in better health? For a Utah sample, larger households may be a rough indicator of membership in the Church of Jesus Christ of Latter-Day Saints (LDS). The LDS church encourages both large families and lifestyles (i.e., abstinence from caffeine and alcohol) that tend to produce positive health benefits. Other studies have shown that the prevalence rate for many chronic conditions among adults are lower among LDS members than others [3]. Children from LDS families may therefore benefit in this effect. To test for this possibility, LDS status of the household was included in the equations. LDS status of the household does not generate significant changes in the number of reported health conditions while household size continues to produce a downward impact on health reports holding LDS status constant.

The second important measure of respondent burden is the length of the interview. The argument is that longer interviews require a greater time commitment thereby taxing the respondent's good will and ability to concentrate. Alternatively, long interviews are long not because of the inherent length of the questionnaire but because of the established commitment of the respondent to expend greater effort in providing the requested information. Table 5 indicates that longer interviews significantly increase health reports rather than reducing them, controlling for household size. This relationship is somewhat knotty because it has a chicken-egg quality to it. Respondents that are asked about their children's health status require time to provide this information and as a consequence, the more health conditions to explain, the longer the interview. Moreover, only those who volunteer to participate to begin with are included in the sample. Those worried about the prospects of a long interview often do not agree to be interviewed to begin with.

Table 5. Effect of Child and Respondent Characteristics on Reporting the Number of Children's Health Conditions. Numbers are least squares regression coefficients and t-statistics (in parentheses).

	Total Current and Past Year Health Conditions	Total Current and Past Year Chronic Health Conditions
Constant	0.314 (6.44)	0.213 (5.08)
Household Size	-0.262 (5.95)	-0.244 (6.45)
Child's Age (yrs.)	-0.00420 (2.80)	-0.00322 (2.50)
Length of Interview (in Minutes)	0.0125 (11.2)	0.0102 (10.6)
Household Income	0.240 X (10-6) (0.398)	-0.850 X (10-6) (1.64)
R's Age (yrs.)	-0.00619 (5.15)	-0.00316 (3.07)
R's Education (yrs.)	-0.236 X (10-3) (0.344)	-0.109 X (10-3) (0.185)
R's Total Conditions	0.143 (16.9)	0.0639 (8.75)
R is Mother (=1)	0.0634 (3.46)	0.0576 (3.65)
LDS (=1)	-.014 (0.660)	-0.13 (0.688)
Mean of Dependent Variable	0.324	0.221
R^2	0.064	0.034
$\bar{R}^2$	0.063	0.033
N	8103	8103

Three important respondent characteristics were examined which are measures of the respondent's ability to respond about their children's health. First, respondents who are themselves in poor health may not be in a position to provide accurate accounts of their children's well-being. The equations reported in Table 5 clearly indicate that this is not the case and that quite the opposite is true: respondents with many health conditions report more health conditions for their children. This may be an indication that there are true but unobserved familial variables that lead both parent and child to have a greater number of health problems (e.g., shared environment and in some cases, shared genes). Or perhaps a respondent with more health conditions finds questions about his/her child's health to be more salient. With greater salience comes better reporting by the parent.

Not too surprising is the fact that mothers tend to report significantly more health conditions than fathers. While either a mother or father had to consider themselves knowledgeable about the health of all other household members, mothers could nonetheless recall more health events for their children. Assuming that forward telescoping of health events closer to the time of the interview is a problem for either mother or father, mothers appear to be better family spokespersons when dealing with their child's health.

Controlling for the child's own age, older parents recall fewer health conditions for their children than younger parents. The age of the parent was included to remove the possibility

that household size was an indirect indicator of the age of the respondent (i.e., older respondents have the opportunity to have more children). The negative effect of increasing respondent age is possibly due to reduced salience of this topic, declining memory, or competing concerns about other activities such as work or marriage.

Table 6 examines a slightly different health outcome: the respondent's qualitative assessment of his/her children's general health status. Note that the dependent variable is not strictly an interval level variable but instead takes on ordinal values of 1 if the child is reported to be in excellent health, 2 if very good, 3 if good, 4 if fair, and 5 if poor. Therefore, factors which tend to yield lower numbers on this health scale tend to encourage underreporting since health conditions which would lead a parent to evaluate their child as being in less-than-excellent health are somehow forgotten or supressed.

Table 6. Effect of Child and Respondent Characteristics on Reporting Children's General Health Status (1=Excellent, 2= Very Good, 3=Good, 4=Fair, 5=Poor). Numbers are least squares regression coefficients and t-statistics (in parentheses).

	Dependent Variable - Children's General Health Status
Constant	1.81 (33.2)
Household Size	-0.0242 (4.89)
Child's Age (yrs.)	-0.00327 (1.94)
Length of Interview in Minutes	0.0102 (8.08)
Household Income	-0.640 X (10-5) (9.45)
R's Age (yrs.)	-0.005 (3.70)
R's Education (yrs.)	-0.001 (1.30)
R's Total Conditions	0.101 (10.6)
R is Mother (=1)	-0.004 (0.188)
LDS (=1)	-0.0364 (1.53)
Mean of Dependent Variable	1.55
R^2	0.043
$\bar{R}^2$	0.042
N	8103

A second feature about this general health status question is that it was asked person-by-person rather than household-then-person. That is, after respondents provided information about the age, sex, and relationship of all household members, they were then asked about each individual's general health status. Because each member was mentioned explicitly in this design, the respondent may be in a better position to pause and think about the health condition of an individual, more so than in the household-then-person design.

Table 6 shows that the continued effects of household size persist: respondents from larger households tend to report their children as being in significantly better health than respondents from smaller households. This partially addresses the possibility that underreporting due to many proxy interviews is not largely attributable to the household-then-person design. Instead, a questionnaire that asks for many proxy interviews simply makes large demands on a respondent under either design (e.g., household-then-person or person-then-person) and meeting these demands is an inherently difficult task.

As before, the estimates in Table 6 show that longer interviews are associated with reports of lower health status. Additionally, older parents continue to rate their children as being in better health than their younger counterparts. The respondent's own health condition once again affects reports of their children's general health status: parents with many health problems tend to report their children as having significantly poorer health relative to parents with fewer health conditions.

Table 7 reports two equations for the effects of household size on health care utilization patterns for children. The negative influence of household size on doctor visits and hospitalization in the past year is evident although the usually high significance of level for the household size effect is somewhat reduced in the hospitalization equation (significance levels are actually not too informative for these equations since the sample size is so large) relative to the other equations. It appears that even memorable health events such as hospitalizations are not immune from underreporting bias.

Table 8 examines two health conditions that are unlikely to be underreported or forgotten by the respondent even under the most adverse survey conditions: a heart condition and hearing loss. The dependent variable is a dichotomy (presence or absence of a condition) and therefore the equations were estimated by logistic regressions. Unlike the findings described earlier, household size does not influence the likelihood that these conditions would be mentioned. It is possible that the disappearance of this respondent burden effect is due to the rarity of these conditions thereby minimizing the likelihood that household size would generate a significant impact on reporting their child's health condition. If this were the case, other variables previously found to be influential in affecting health reports would also disappear. This does not seem to be the case since interview length produces more reports of heart disease and hearing loss, higher levels of household income are negatively associated with reported childhood heart disease, and respondents in poorer health significantly report more hearing loss problems for their children compared to respondents in better health. Thus, even when the number of child proxy interviews increase, memorable health events are not differentially recalled by the parent respondent.

Table 7. Effect of Child and Respondent Characteristics on Reporting Children's Health Care Utilization Patterns. Numbers are least squares regression coefficients and t-statistics (in parentheses).

	Number of Times Received In-Person Advice from M.D. in Past Year	Number of Times* Hospitalized in Past Year
Constant	1.58 (9.40)	0.0428 (2.21)
Household Size	-0.209 (13.5)	-0.00401 (2.27)
Child's Age (yrs.)	-0.0887 (17.3)	-0.904 x (10-3) (1.22)
Length of Interview (in Minutes)	0.0512 (13.4)	0.00248 (5.66)
Household Income	0.107 x (10-4) (5.18)	-0.183 x (10-6) (0.779)
R's Age (yrs.)	-0.0887 (17.3)	-0.669 x (10-3) (1.38)
R's Education (yrs.)	-0.00158 (0.673)	-0.197 x (10-3) (0.733)
R's Total Conditions	0.239 (8.25)	0.00917 (2.79)
R is Mother (=1)	0.443 (7.08)	0.00747 (1.04)
LDS (=1)	0.303 (4.18)	0.00462 (0.555)
Mean of Dependent Variable	1.56	0.039
R^2	0.101	0.0071
$\bar{R}^2$	0.100	0.0059
N	8103	7734

*Excludes children born during the year prior to the interview because their birth was recorded as a hospitalization.

Table 8. Effects of Respondent and Child Characteristics on Reports of Child Health Conditions That Are Unlikely to Be Underreported. Figures are Logistic Regression Coefficients and t-Statistics (in Parentheses)

	Has Child Ever Had a Heart Condition (Yes=1, No=0)	Does Child Now Have Any Hearing Loss (Yes=1, No=0)
Constant	-3.86 (4.44)	-4.35 (6.73)
Household Size	0.00923 (0.121)	-0.0700 (1.24)
Child's Age (yrs.)	-0.00911 (.333)	-0.0214 (1.01)
Length of Interview (in Minutes)	0.0191 (2.05)	0.0190 (2.75)
Household Income	-0.0000294 (2.59)	-0.00000557 (1.22)
R's Age (yrs.)	-0.0247 (1.01)	-0.0295 (1.73)
R's Education (yrs.)	-0.00432 (0.380)	-0.00660 (0.867)
R's Total Conditions	0.0256 (0.169)	0.0258 (3.15)
R is Mother (=1)	0.0515 (0.147)	0.0375 (1.34)

CONCLUSION

The impact that household size has on reports of child health status made by parents has been examined for several different health outcomes. The motivation for looking at this relationship comes from the common practice among survey researchers to use a single respondent as the source of information for all household members. Respondents from larger households are therefore asked to do more work during the course of the interview than respondents from smaller households.

Based on these results, parents report significantly fewer health conditions and rate their children higher in terms of general health status when they conduct more proxy interviews. It is possible that this relationship might be spurious and that the real mechanism at work was the length of the interview. Arguably, interviews may take more time because there are more children in the home to either think about or to be distracted by leading to lower reporting. The data do not support this position because the inverse relationship between health events and household size exists once controlling for interview length. In fact, longer interviews have quite the opposite effect in that greater mentions of health conditions come from lengthier interviews rather than from briefer ones.

Based on these data, what recommendations are there for future surveys? Several options exist which vary in their degree of payoff and which are appropriate only under particular circumstances.

1) Use only self-reported data from a randomly selected adult within the household: If the burden on the respondent increases as the number of proxy interviews increases, it is possible that the respondent should only answer for him/herself. This is appropriate in only certain settings where strictly individual information is being sought. When family or household information is needed, the respondents must either rely on their own recollections of other's behavior (which in some cases may be desirable) or the questionnaire should exclude such items from the survey. Obviously, reliance on self-reports is not feasible when information about children is sought. But unfortunately, the only reliable and practical method for obtaining this information is through the parent.

2) Interview all household members: Ideally one would probably prefer to obtain information about each household member directly from the member in question. Unfortunately, this design can be extremely time consuming and costly. Even large survey operations attempt to achieve this level of contact among all adult household members but must live with some proxy interviews out of practical necessity [7]. Even with the National Health Interview Survey, information on children is always obtained by proxy.

3) Control for the number and type of proxy information if individual information is being sought: To minimize the differential amount of proxy interviewing performed by any one respondent, it is possible to fix the number of household members. This approach is tenable when the survey seeks to obtain information on individuals within the household. Some studies have used this design [11] although it has the obvious limitation of reducing the generalizability of the surved to only households of a

given size. Furthermore, it does not alleviate the potential proxy interview bias but merely standardizes the amount of respondent burden.

4) Statistically control for the household size and proxy interview status: With this approach, the impact of proxy interviews and the number of other interviews proxied by the respondent could be adjusted for in the analysis (see, for example, [4, p. 643]). This strategy has an advantage in that data already collected with proxy interviews can be analyzed after adjustments have been made for possible misreporting bias. However, in the case of proxied children interviews, it is impossible to separate out the effect of a child's proxied response from any child's response since there can typically be no comparisons between proxied information and self-reports (i.e., child reports and proxied reports are the same).

5) Continue to use proxy interviews but with more extensive probes: It is possible that underreporting occurs more frequently among larger households, particularly for less memorable health events, because the survey instrument does not contain sufficient probes. More follow-up questions may help to alleviate underreporting by the resondent [6]. Moreover, additional probes may also encourage interviewers to "stay with" the respondent longer to obtain the information for numerous individuals.

(6) Randomly Select One child: This would standardize the possible burden to all parents. However, the objective of the survey must not be a complete accounting of household health status.

Further studies of the relationship between household size and respondent burden with other widely available data sets covering different topics will help to increase our understanding of this survey design. Studies that would be the most useful to examine for household size-respondent burden effects include the National Health Interview Survey, University of Michigan's Panel Study of Income Dynamics and the Health and Nutrition Examination Surveys. Each of these surveys have used different designs

and focus on a variety of topics that may either enhance or suppress the adverse burden effects examined in this paper.

APPENDIX A

The health conditions asked in the Utah's Health Status Survey: 1986 are listed below.

A. Has anyone in the household ever had any of the following conditions (that was verified by a medical doctor)?

1. Arthritis or rheumatism
2. High blood pressure or hypertension
3. A stroke
4. Chest pain caused by angina
5. Any other heart condition

B. Has anyone in the household had any of the following conditions during the past 12 months (that was verified by a medical doctor)?

6. Diabetes
7. Bronchitis, emphysema, or asthma
8. Cirrhosis or enlarged liver
9. Cancer
10. Influenza or pneumonia

C. Does anyone in the household now have any of the following conditions (that was verified by a medical doctor)?

11. Psychiatric disorders
12. Hearing
13. Wear a hearing aid
14. Ear infections or other problems with one or both ears.
15. Blindness in one or both eyes
16. Problems seeing even when wearing glasses
17. Any speech or language requirements
18. Palsy or cerebral palsy
19. Paralysis of any kind
20. Difficulty with balance and/or walking
21. Reading problems
22. Learning disability
23. Mental retardation

D. Was anyone in the household a patient in a hospital overnight or during the past 12 months? How many different times did (___) stay in a hospital overnight during the past 12 months?

E. Within the past 12 months, did any member of the household get medical advice from a medical doctor or assistant either in person or over the telephone (do not include overnight hospital visits)? How many times did (___) get medical advice in person from a medical doctor (within the past 12 months)?

REFERENCES

1. Andersen, Ronald, Kooper, Judith, Frankel, Martin R., and Associates. Total Survey Error: Applications to Improve Health Surveys. San Francisco, CA.: Jossey-Bass Publishers, 1979

2. Bradburn, Norman. "Respondent Burden," in L. Reeder (Ed.) Health Survey Research Methods: Second Biennial Conference, Williamsburg, VA. Washington, D.C.: U.S.G.P.O., 1979

3. Lyon, Joseph L., Gardner, John W., West, Dee W. "Cancer Incidence in Mormons and Non-Mormons in Utah During 1967-75", J. National Cancer Institute 65:1055-1061, 1980.

4. Mathiowetz, Nancy A., and Groves, Robert M.. "The Effects of Respondent Rules on Health Survey Reports," American Journal of Public Health (v. 75(6):639-644, 1985..

5. Mosely, Ray R., II, and Wolinsky, Frederic D. "The Use of Proxies in Health Surveys: Substantive and Policy Implications," Medical Care v. 24(6):496-510, 1986.

6. National Center for Health Statistics. Cannell, Charles F. and Fowler, Floyd. Comparison of Hospitalization Reporting In Three Survey Procedures. Public Health Service Publication No. 1000-Series 2, No. 8, Washington D.C. Government Printing Office, 1965.

7. National Center for Health Statistics. Kovar, M. G., and Poe, G. S. The National Health Interview Survey Design, 1973-84, and Procedures, 1975-83. Vital and Health Statitics, Series 1, No. 18. DHHS Publication No. (PHS) 85-1320. Public Health Service. Washington, U.S. Government Printing Office, August, 1985.

8. Sharp, Laure E., and Frankel, Joanne. "Respondent Burden: A Test of Some Common Assumptions," Public Opinion Quarterly v. 47:36-53, 1983.

9. Sheatsley, P. "The Harassed Respondent: II. Interviewing Practices," in L. Bogart (Ed.) Current Controversies in Market Research. Chicago: Markham, 1969.

10. Turner, Charles F., and Martin, Elizabeth (Ed.) Surveying Subjective Phenomenon, Vol. 1, Chapter 8. Russell Sage Foundation: New York, NY, 1984.

11. Virginia Agricultural Experiment Station. Family Time Use: An eleven-state Urban/Rural Comparison, Bulletin VPI-2, Virginia Agricultural Experiment Station, Blacksburg, Virginia, 1981.

12. Waksberg, Joseph. "Sampling Methods for Random Digit Dialing," Journal of the American Statistical Association 73:40-46, 1978.

GHOSTBUSTING: IN SEARCH OF THE RESERVATION WAGE

Janis Kohanski Pappalardo, The Federal Trade Commission[1]

ABSTRACT

Two techniques for estimating reservation wage rates for non-workers are compared in this paper: the Heckman technique and a direct method based upon responses to a survey question. A low correlation coefficient between the two suggests that the two approaches are not measuring the same underlying variable.

INTRODUCTION

Social science is overrun by ghosts: elusive variables that exist in theory but are unobserved in reality. Researchers who attempt to estimate models of consumer behavior consistent with economic theory often face dilemmas about how to capture these ghosts. Can survey respondents be relied upon to reveal ghost measurements? Or, is it better for researchers to sneak up on the ghosts, by estimating measures of unobserved variables from related data? How much difference is there between a direct approach and an indirect approach to ghostbusting?

One ghost that pervades the labor supply literature is the opportunity cost of time for the voluntarily unemployed (non-workers.) Estimation of this unobservable variable is further complicated by apparent confusion about its proper conceptual measure. For example, Ferber and Green (1985) appear to suggest that potential market wage rates (or perceived potential market wage rates) measure the opportunity cost of time for non-workers. Zick and Bryant (1983), on the other hand, explain why the reservation wage rate is a more appropriate measure. Potential wages, as noted by Zick and Bryant, "only provide us with a lower limit of the price non-employed individuals place on their home work time" [7, p.136].

Despite some apparent conceptual problems with Ferber and Green's work, they have contributed to the literature by exposing discrepancies between direct and indirect approaches to ghostbusting. Their direct approach is to ask non-workers for their own estimates of the wages they could command in the market. Their indirect approach is to generate potential market wage estimates via econometric techniques and data on related variables. (Potential market wages are estimated with and without sample selection bias corrections.) Ferber and Green find a substantial difference between women's own estimates and those derived from the econometric procedures. They therefore suggest that "data adjusted for selection bias may not be useful for the opportunity cost approach to valuation of housework"[2] and call for further research on the differences between respondent's own value of time estimates and those produced by econometrics.

In this paper I answer Ferber and Green's call for more comparisons between direct and indirect value of time estimates. However, my work differs from theirs in two ways. First, I estimate the value of time for full-time college students[3] rather than full-time homemakers. Second, following Zick and Bryant's approach, I use reservation wages (not potential market wages) to measure the value of time for non-workers.

A comparison between the direct and indirect approaches to time valuation will not indicate which procedure is superior because the true values for such ghost variables are never known. However, a comparison can reveal how close indirect estimates are to respondent's own estimates. If the two approaches yield significantly different measures, then this would suggest that researchers should use both approaches to estimate a range of reservation wage rates when measures of the value of time for non-workers are needed.

RESERVATION WAGES AND THE VALUE OF TIME

Time is money. It is therefore not surprising that neoclassical economists have shown how wages can be used to measure the value of time. An understanding of time valuation for full-time students can be gleaned from an application of household production theory to student academic performance production.[4] The theory indicates that the value of time for students who work is approximated by their market wage rates and the value of time for students who do not work is approximated by their reservation wage rates.

---

[1] Staff Economist, Bureau of Economics. The views expressed in this paper are the author's and may not be shared by any FTC Commissioners or other members of the Commission Staff.

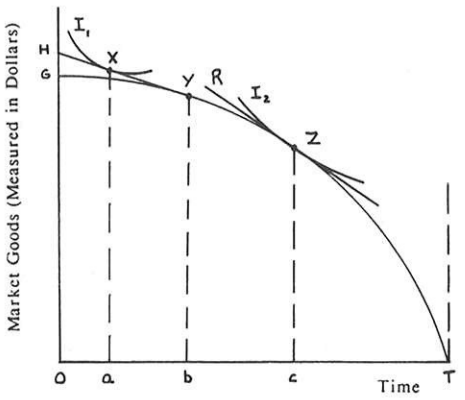
[2] Ferber and Green (1985) at 99.

[3] I use "full-time" student to indicate that students are enrolled in a full-time program, and not to indicate that they do not work for pay.

[4] The graphical approach follows Groneau's (1978) model. To represent the model in two dimensions requires an assumption that academic performance is a perfect substitute for market goods. A more general approach, with less restrictive assumptions, appears in Pappalardo (1986)

Students, like homemakers, combine time and market inputs to produce outputs; homemakers engage in household production and students engage in academic performance production. The academic performance production process is sketched in Figure 1. Time is measured along the horizontal axis and market goods (in dollars) are measured along the vertical axis.[5] The total time available during the period under study is T.

FIGURE 1. Graphing the Trade-offs between Work, Study, and Leisure Time.



In this model students choose among three time uses: leisure time, study time, and work time. Time allocation is determined by the student's academic performance production function, the market wage rate, and his preferences[6]. Arc TG, the academic performance production function, shows how time is related to the present value of academic performance[7]. The slope (in absolute value terms) of line HY reflects the real market wage rate available to the student. Preferences are reflected by indifference curves I_1 and I_2.

If the student has the preferences depicted by I_2 then his utility is maximized at point z, where the rate at which he willing to trade goods for time (depicted by the slope of I_2 at z) is equal to the rate at which he is able to produce academic performance units (measured as the slope of the production function at z.) Thus, this student devotes CT hours to study and OC hours to leisure. The slope of RZ (in absolute value terms) illustrates the value of time for this student who chooses not to work. He does not work because the market's valuation of his time is lower than his own valuation. (Note that RZ is steeper than HY.) The slope of RZ (in absolute value terms) measures the student's reservation wage rate: the minimum wage he would have to be paid for him to be indifferent between working and not working.

If the student has indifference curve I_1, then he will devote OA hours to leisure, AB hours to work, and BT hours to study. Notice that this is the combination of time uses that maximizes his utility, given his income and market prices. The value of time to this worker is reflected by his market wage rate, which is measured by the slope of HY, (and is equal to his reservation wage.) At point x, the student equates his willingness to trade market goods for time (as depicted by the slope of the indifference curve at x) with (1) his ability to trade market goods for time (as depicted by the slope of HY) and (2) the rate at which he is able to produce academic performance units (in present value terms and measured as the slope of the production function).

DATA

The data used for this analysis are from a random survey of full-time Cornell University sophomores, juniors, and seniors conducted during the fall, 1984 semester.[8] The first wave of the survey consisted of 183 personal interviews, which were conducted prior to Thanksgiving break. Data were gathered on student time-use and demographic characteristics. Those who agreed to release their grades at the end of the term were surveyed again by telephone after Thanksgiving break. The sample used for this analysis consists of 143 students. Observations with missing values were dropped. This study is limited to an analysis of behavior during the post-Thanksgiving period because the reservation wage rate question was asked during the follow-up interview. Variables used in the analysis and their descriptive statistics are presented in Table 1.

[5] Assume that academic performance is measured as the student's grade point average (GPA) multiplied by the number of completed credit hours. The resultant measure, which I call "grade point" or "GP" takes into account the quality as well as the quantity of academic performance. For the graphical analysis the value of academic performance can be viewed as the present value of the student's grade point. By translating GP into a dollar amount, it is possible to measure GP and market goods along the same axis.

[6] Non-wage income also affects time allocation, however, for simplicity, this aspect is not incorporated into Figure 1. Prices, including the interest rate, also enter into the analysis by affecting the magnitudes depicted by the vertical axis.

[7] Arc TG is therefore a transformation of the academic performance function that strictly relates time and grade points.

[8] See Pappalardo (1986) for further details.