EUNKYU LEE, MICHAEL Y. HU, and REX S. TOH* 

Using a large-scale consumer database created by AT&T, the authors investigate how actual behavioral frequency and duration systematically affect the direction of errors in consumer survey responses. By analyzing errors in consumers’ reports on their frequency of using long-distance telephone calls, letters, cards, and visits for personal communication, the authors demonstrate that high-frequency groups underreported their behavioral frequencies, whereas low-frequency groups overreported them. Similarly, the results show that consumers underestimate the duration of lengthy telephone conversations, whereas they overestimate the duration of short ones. Overall, the authors find that people tend to overestimate both frequency and duration. These compressive regressive effects toward the mean and overall upward bias for both frequency and duration estimations result in a distorted view of the market, which will be incorrectly perceived to be more homogeneous and larger than it really is.

Are Consumer Survey Results Distorted? 
Systematic Impact of Behavioral Frequency and Duration on Survey Response Errors

Obtaining accurate information on the frequency and duration of consumer activities is important for marketing managers. For services provided by long-distance telephone companies, Internet service providers, rental car companies, amusement parks, hotels, and so forth, estimates of usage frequency and duration affect market size estimation and sales forecasts. The prominence of the 80/20 rule in marketing (Schmittlein, Cooper, and Morrison 1993) indicates marketing managers’ strong interest in enhancing their ability to identify heavy users accurately and zero in strategically on sizeable target markets. In contrast, the various pricing options offered by America Online Inc. (an Internet service provider) demonstrate the potential of employing consumer usage rate information for customized pricing and promotional policies targeted toward heavy and light user groups. In addition, marketing managers can better select effective advertising media by knowing how often and how long consumers read newspapers, watch television, listen to the radio, or surf the Internet.

To obtain such usage information, marketers often rely on surveys for their superior efficiency in time, cost, and effort relative to other methods of data collection, such as diaries and direct observation. However, there is abundant evidence in the literature that survey results on behavioral frequency are often subject to errors and biases (Blair and Burton 1987; Bradburn, Rips, and Sherwell 1987; Greene 1984; Hu, Toh, and Lee 1996; Loftus, Fieberg, and Tanor 1985; Menon 1993; Menon, Raghubir, and Schwarz 1995). In our article, we investigate how such errors and biases in consumer surveys might distort the marketing manager’s understanding of consumer usage frequency and duration.

Specifically, we examine the impact of both the actual frequency and duration of consumer activities on survey response errors. Our central research question is whether the consumer’s actual behavioral frequency and duration systematically affect the direction of survey response error. By analyzing a large database created by a major corporation, we show that systematic errors in surveys might lead marketing managers to overestimate the average usage fre-

*Eunkyu Lee is an assistant professor, Faculty of Commerce and Business Administration, University of British Columbia (e-mail: eunkyu.lee@commerce.ubc.ca) Michael Y. Hu is a professor, College of Business, Kent State University (e-mail: mhu@bsa.kent.edu). Rex S. Toh is a professor, Albers School of Business and Economics, Seattle University (e-mail: rextoh@seattleu.edu). The authors acknowledge helpful comments provided by Peter Darke, Vinay Datar, John Dickson, Sung-Tai Hong, Sharon Lobel, Geeta Menon, Carl Obermiller, Gregory Prussia, and Chuck Weinberg. They also thank the three anonymous JMR reviewers for their helpful comments. To interact with colleagues on specific articles in this issue, see “Feedback” on the JMR Web site at www.ama.org/pubs/jmr.

Copyright 2000. All rights reserved.
quency and duration and to underestimate the heterogeneity among consumers.

We seek to contribute to the literature on survey accuracy in the following ways: First, our study is based on a large-scale real-life database generated by industry. As such, this study has the advantage of enhanced realism over experimental studies using small student samples. Specifically, the data collected from a large, demographically proportional national sample of heads of households over a period of 14 months enable us to state our research findings with greater reliability and external validity. Furthermore, instead of measuring the effects of experimental treatments administered to randomly assigned subject groups, our data reflect the impact of different consumer activity levels in real life on systematic errors. Therefore, we can measure the actual heterogeneity of consumer usage without the contaminating effects of experimental controls.

Second, although many studies have investigated the errors in survey responses on behavioral frequency, little work has been done on the accuracy of estimates of the duration of personally experienced activities or events. As previously discussed, frequency and duration are often the two key components of consumer usage rate, which is of interest to marketing managers. Consequently, our investigation of usage duration contributes to a more balanced understanding of errors and biases involved in estimating consumer usage rates with surveys. On a broader level, this article expands the literature on consumer perception of time duration, which has been largely limited to the issues surrounding waiting time and customer satisfaction (e.g., Carmon, Shanthikumar, and Carmon 1995; Kumar, Kalwani, and Dada 1997).

Third, we show that the previously observed regressive effect toward the mean also applies to reports of consumer usage. We also present some new insights into how this phenomenon manifests itself and into its overall impact on survey accuracy.

Note that our study is about activities for which the frequency and duration are reasonably under the consumer’s control (e.g., making long-distance telephone calls, visiting supermarkets, viewing television). Also note that these frequencies and durations vary substantially across consumers and occasions, respectively. Therefore, our findings might not apply to cases such as purchasing a Christmas tree or completing an income tax return (typically done once a year), having dinner (once a day for most people), or brushing teeth (rarely lasting for more than five minutes). Also note that activities involving socially undesirable activities, such as smoking, drinking, or gambling, are outside the scope of our findings, because the survey results on such activities have been found to have different systematic biases (Lansing, Ginsberg, and Braaten 1961; Wind and Lerner 1979).

The rest of the article is organized as follows: First, we generate research hypotheses based on relevant literature. Second, we describe the characteristics of the database and the methods of analysis for hypothesis testing. Third, we present the results of our analyses. Finally, we discuss some of the practical ramifications of our findings, recognize the limitations of our study, and then suggest avenues for further research.

RESEARCH HYPOTHESES

Existing studies on autobiographical frequency estimation are generally focused on cognitive processes used for storing and retrieving autobiographical information. Many researchers (Blair and Burton 1987; Brewer 1986, 1988; Jabin et al. 1984; Lessler et al. 1985; Menon 1993, 1997; Tourangeau, Lessler, and Salter 1985) have paid particular attention to the use of the episodic recall-and-count strategy (for example, I remember dining out on three specific occasions last week) or the rate-of-occurrence computational method (for example, if I brush my teeth twice a day, I brush my teeth 14 times a week). Bickart and Felcher (1996) and Conrad, Brown, and Cashman (1998) discuss a broader array of frequency estimation strategies, including converting a general impression of frequency into a numerical response. Many studies (Blair and Burton 1987; Burton and Blair 1991; Hu, Toh, and Lee 1996; Ross 1984; Schwarz 1990; Strube 1987) have found that frequent behavior generally leads to the use of the rate-based heuristic, which results in the possibility that actual frequency will systematically affect the direction of survey report errors. However, no study has directly examined the effect of actual frequency on the direction of survey report errors.

For nonautobiographical information, prior studies on frequency estimation provide evidence that there exists a general tendency for estimates to regress toward the mean (Berg et al. 1986; Fiedler and Armbruster 1994; Greene 1984; Hintzman 1988; Williams and Durso 1986). Furthermore, the psychometric formulation by Fiedler and Armbruster (1994) suggests that for similar levels of reliability, the regressive effect is strongest for extreme values. One explanation for such a regressive effect could be that, though actual frequencies vary over time stochastically, respondents might report averages, which leads to underreporting of high frequencies and overreporting of low frequencies. An alternative explanation is that low-frequency events tend to be more distinctive, causing respondents to use an availability heuristic (Tversky and Kahneman 1973, 1974) and overestimate infrequent events. In addition, Fiedler (1991) provides some plausible theories for the cognitive processes leading to regressive frequency estimation even in the absence of the distinctiveness effect. All these studies suggest a strong possibility that high-frequency consumers underreport their usage frequencies whereas low-frequency consumers overreport.

In contrast to the many studies conducted on frequency estimation errors, there have been virtually no studies on errors and biases associated with duration estimations of personally experienced events and activities. In psychology, the majority of prior studies on duration estimations involve the courts’ interest in estimating the duration of witness testimonies (e.g., Loftus et al. 1987; Yarmey 1990; Yarmey and Matthis 1990), but these studies provide few insights on the accuracy of consumer estimations of usage durations. In marketing, the focus has been on waiting time, that is, how to reduce the perceived waiting time and its impact on the consumer (Carmon, Shanthikumar, and Carmon 1995; Kumar, Kalwani, and Dada 1997).

1 In this study, respondents were asked about their behavior in an average or typical month. Consequently, this is not a valid explanation for the regression observed in frequency reports in our data.
Nevertheless, we find some valuable hints in Burt's (1993) study, which suggests a possible systematic relationship between duration estimation accuracy and actual duration. When examining students' retrospective duration estimations of 20 public events (such as the 1985 hijacking of TWA Flight 847 and the Iranian students' occupation of the U.S. Embassy in Teheran), he found that when the actual durations of events were atypically long (short), respondents generally underestimated (overestimated) their duration. Thus, for nonautobiographical events, respondents' estimations of duration tend to show the regressive pattern similar to what is observed in frequency estimations. On the basis of the implications of the previously discussed studies, we suspect that this persistent regressive phenomenon might also exist in survey reports of both frequency and duration for consumer activities. Thus we postulate the following:

\[ H_1: \text{Higher frequency and longer duration tend to be underestimated, whereas lower frequency and shorter duration tend to be overestimated.} \]

As a corollary,

\[ H_2: \text{The higher the frequency and the longer the duration, the greater is the number of people who will underestimate rather than overestimate, whereas the lower the frequency and the shorter the duration, the greater is the number of people who will overestimate rather than underestimate.} \]

\[ H_1 \] and \[ H_2 \] might give the wrong impression that the underestimation by the high-frequency and long-duration groups will more or less cancel out the overestimation by the low-frequency and short-duration groups, which would result in reasonably accurate estimations of the overall mean. However, Fiedler (1991) shows that the inclusion of a null category in a frequency judgment task produces a considerable upward regressive effect, because zero is an extreme value that results in pronounced overestimation. Note also that the maximum degree of overestimation is theoretically infinite, whereas the degree of underestimation is zero-bounded downward. Empirically, many studies have reported a general tendency for respondents to overreport in surveys (Burton and Blair 1991; Hu, Toh, and Lee 1996; Menon 1993; Neter and Waksberg 1961; Parfitt 1967). Therefore we postulate the following:

\[ H_3: \text{When regressing toward the mean, the degree of overestimation will exceed the degree of underestimation, and therefore on average survey estimates will be exaggerated.} \]

**DATABASE AND METHODOLOGY**

We tested the hypotheses by analyzing the data set collected by AT&T. The data were collected from a demographically proportional national sample of 3990 households stratified on the basis of income, marital status, age, sex, population density, and geographical region. The company collected information on the participants' personal communication activities for four different modes (long-distance telephone calls, letters, cards, and visits). The data collection process consisted of the following three stages: the initial survey stage, the diary-keeping stage, and the final survey stage (see Figure 1).

In the initial survey stage, the sample participants were administered a mail questionnaire survey that asked, "Dur-
then permitted comparisons of the duration of calls reported with actual billing records at the individual call level. Using the responses obtained in the initial survey and the measures of actual frequency and duration obtained from the diaries and company records, respectively, we computed the survey frequency error in the following two ways:

\[ SEF = EF - AF \]

and

\[ SPEF = (SEF/AF) \times 100. \]

where

- \( SEF \) = signed error in survey report of frequency,
- \( SPEF \) = signed percentage error in survey report of frequency,
- \( EF \) = estimated monthly frequency in the initial survey report, and
- \( AF \) = actual monthly frequency from average of diary records.

Errors in duration reports were computed in the same way. Specifically,

\[ SED = ED - AD \]

and

\[ SPED = (SED/AD) \times 100. \]

where

- \( SED \) = signed error in the estimation of duration,
- \( SPED \) = signed percentage error in the estimation of duration,
- \( ED \) = estimated duration as reported in the diary, and
- \( AD \) = actual duration from billing record.

Note that we have two measures of error: signed error and signed percentage error. We used the percentage error as a measure of survey accuracy to eliminate the level effect, because people associated with high frequency or duration tend to make larger estimation errors, and vice versa. The use of percentages or ratios (same concept) has long-standing precedence (McKenzie 1983; Neter 1970; Parfitt 1967; Sudman 1964a, b; Wind and Lerner 1979).

One month after the end of the 12-month diary-keeping period, the respondents again were asked in the final mail survey to indicate their perceived frequency by category (heavy, medium, and light users and nonusers) for the four communication modes. To avoid confounding estimation errors with substantial behavioral changes when computing SEF, our analysis is based on only those who remained in the same self-reported categories before and after the 12-month diary-keeping period for each communication mode. This match-screening process minimizes discrepancies between the survey estimates and subsequent diary averages because of secular trend. Panel attrition due to natural mortality and participation fatigue (Tolh and Hu 1996) reduced the sample size from 3990 to 1108. The match-screening process resulted in even smaller but cleaner samples of sizes that ranged from 307 for cards to 659 for long-distance telephone calls.

**RESULTS**

Our findings regarding the impact of frequency and duration on the direction of survey response errors are summarized in Table 1. We demarcated the high/low frequencies and long/short durations by their respective sample means. Examining Table 1, we note that in 19 of 20 cases, higher frequency and longer duration lead to underestimation, whereas lower frequency and shorter duration lead to overestimation. This is true for both signed error and signed percentage error, and for 16 of the 19 cases that are in the predicted direction, the results are significant at \( p < .0200 \).

Then we plotted the signed errors against the frequencies and durations and got the expected results for all four modes of communication as well as for the duration of long-distance telephone calls. In the interest of parsimony, we present the results only for the frequency and duration of long-distance telephone calls (see Figure 2). Observe that low frequencies and short durations lead to overestimations, whereas high frequencies and long durations lead to underestimations. The figure shows that the relationships are almost linear monotonic.

Thus our results not only support \( H_1 \), which holds that higher frequency and longer duration tend to be underestimated whereas lower frequency and shorter duration tend to be overestimated, but also show that there is an almost monotonic and linear relationship between frequency and duration and signed error.

We suspect that there are at least two reasons diary reports ostensibly filled out at the time of the calls also exhibit a regression effect. First, the diaries were recorded weekly, and the elapsed time between the telephone call and the diary entry might have caused some memory loss. Second, even if some diary records were entered right after the calls, few respondents are expected to have measured the actual durations. Consequently, we speculate that the uncertainty involved in recalling the durations caused the respondents to gravitate toward the average.

Recall that \( H_2 \) holds that the higher the frequency and duration, the greater is the number of people who will underestimate rather than overestimate, and vice versa. Examining Table 2, we find support for \( H_2 \) in nine of ten cases (high- and low-frequency groups for four modes of communication plus long- and short-duration calls for long-distance telephone). For example, in the high-frequency group for long-distance telephone calls, 130 respondents underestimated whereas 99 overestimated. In contrast, in the low-frequency group more participants overestimated than underestimated (237 versus 182). Similarly, for long-duration calls the durations were underestimated more often (997) than overestimated (513), whereas for short-duration calls the durations were more likely to be overestimated (1635) than underestimated (708). The empirical support for \( H_2 \) and \( H_3 \) strongly suggests that surveys of consumer activity levels tend to underreport the actual heterogeneity among consumers by making the distribution of activity frequencies and durations appear to be more concentrated around the mean.

Further examining Table 2, we arrive at an unexpected finding. Note for example that in the case of the frequency of long-distance telephone calls, the average magnitude of overestimation (in absolute terms) is greater than that of underestimation for both the high-frequency group (i.e.,
Table 1
EFFECTS OF FREQUENCY AND DURATION ON THE DIRECTION OF SURVEY RESPONSE ERRORS

A. Effects of Frequency on Average SEF and Average SPEF

<table>
<thead>
<tr>
<th>Mode of Communication</th>
<th>Dependent Variable</th>
<th>High Frequency</th>
<th>Low Frequency</th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-distance telephone</td>
<td>Average SEF</td>
<td>-3.3</td>
<td>.94</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.2129)</td>
<td>(.0001)</td>
<td>(.0011)</td>
</tr>
<tr>
<td></td>
<td>Average SPEF</td>
<td>6.15*</td>
<td>142.90</td>
<td>94.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.1230)</td>
<td>(.0001)</td>
<td>(.0001)</td>
</tr>
<tr>
<td>Letters</td>
<td>Average SEF</td>
<td>-1.54</td>
<td>70</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0084)</td>
<td>(.0001)</td>
<td>(.2380)</td>
</tr>
<tr>
<td></td>
<td>Average SPEF</td>
<td>-2.19</td>
<td>141.80</td>
<td>97.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.3445)</td>
<td>(.0001)</td>
<td>(.0001)</td>
</tr>
<tr>
<td>Cards</td>
<td>Average SEF</td>
<td>-76</td>
<td>76</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0171)</td>
<td>(.0001)</td>
<td>(.0228)</td>
</tr>
<tr>
<td></td>
<td>Average SPEF</td>
<td>-5.06</td>
<td>189.36</td>
<td>123.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.2676)</td>
<td>(.0001)</td>
<td>(.0001)</td>
</tr>
<tr>
<td>Visits</td>
<td>Average SEF</td>
<td>-3.85</td>
<td>.30</td>
<td>-12*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0001)</td>
<td>(.0001)</td>
<td>(.0519)</td>
</tr>
<tr>
<td></td>
<td>Average SPEF</td>
<td>-23.63</td>
<td>129.84</td>
<td>73.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0002)</td>
<td>(.0016)</td>
<td>(.0022)</td>
</tr>
</tbody>
</table>

B. Effects of Duration on Average SED and Average SPED. Long-Distance Telephone Calls

<table>
<thead>
<tr>
<th>Mode of Communication</th>
<th>Dependent Variable</th>
<th>Long Duration</th>
<th>Short Duration</th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-distance telephone</td>
<td>Average SED</td>
<td>-3.03</td>
<td>2.93</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0001)</td>
<td>(.0001)</td>
<td>(.0001)</td>
</tr>
<tr>
<td></td>
<td>Average SPED</td>
<td>-9.92</td>
<td>119.18</td>
<td>70.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0001)</td>
<td>(.0001)</td>
<td>(.0001)</td>
</tr>
</tbody>
</table>

* = misdirection.
Numbers in parentheses are p-values

4.070 > 3.688) and the low-frequency group (i.e., 2.216 > .688). This phenomenon can be observed in eight of ten cases for the frequencies of all four modes of communication and the durations of long-distance telephone calls. This is in contrast to our validation of H2, which holds that the higher the frequency and the longer the duration, the greater is the number of people who will underestimate rather than overestimate, and vice versa. This consistent pattern of greater magnitude of overestimation relative to that of underestimation might be due partly to overestimation having no theoretical upper limit, whereas underestimation is zero-bounded downward. In the case of frequency as opposed to duration, this might also be due to excessive telescoping ( Sudman, Bradburn, and Schwarz 1996) by those who used the recall-and-count strategy of memory retrieval.

The previous discussion indicates that the survey method for estimating the average usage rate might be subject to greater errors in the case of light users compared with that of heavy users. That is because for low-frequency (short-duration) cases, not only is the majority of respondents expected to overreport, but also the magnitude of overreporting will be relatively high, which will result in a serious overreporting problem. In contrast, the underestimation problem in high-frequency (long-duration) cases is somewhat mitigated by the fact that, though a greater proportion of respondents are expected to underreport, the impact is partially canceled out by the greater magnitude of overreporting. In the AT&T data, this asymmetry between overestimation and underestimation caused overall overestimation.
in nine of ten cases for both actual signed error and signed percentage error (see the right-hand column of Table 1). In the nine cases in which the results were in the predicted direction, eight are significant at \( p < .0300 \). Thus the results support \( H_1 \), which holds that when regressing toward the mean, the degree of overestimation will exceed the degree of underestimation, and therefore on average survey estimates will be exaggerated.

Finally, we discuss how the combination of the previously reported effects can produce a distorted view of the market when consumer surveys are used. With respect to long-distance telephone calls, we note that the distortions in surveys are as follows:

- The frequency means shifted from 3.27 to 3.76, which represents a 14.98\% upward shift.
- The duration means shifted from 13.40 to 14.06, which represents a 4.93\% upward shift.
- The frequency range was compressed from 43.09 to 30.00.
- The frequency range was compressed from 120.00 to 98.00.
- The frequency coefficient of variation was compressed from 124\% to 118\%, which represents a compression of 4.84\%.
- The duration coefficient of variation was compressed from 85.22\% to 76.74\%, which represents a compression of 9.95\%.

These distortions suggest that autobiographical survey reports for these kinds of behavior tend to be shifted upward and compressed toward the mean, which leads to overestimation of the average and underestimation of the heterogeneity. This gives the marketing manager the false impression that in the case of usage rates there is a larger and more homogeneous market than is actually the case. For example, if AT&T had used the survey results to estimate the demand for the long-distance telephone services in number of minutes, the 14.98\% overestimation of the average frequency combined with the 4.93\% overestimation of the average duration would have led to a total overestimation of 20.65\%.

**DISCUSSION**

In our study, we seek a deeper understanding of systematic errors and biases involved in survey responses to questions on consumer activity levels. We analyzed the problem by decomposing activity level into frequency and duration. From the literature on errors made in frequency and duration estimations, we generated hypotheses relating to survey reports on consumers' activities. Using our access to a real-life, demographically and geographically representative database created by AT&T, we were able to test and validate our hypotheses. Additional information from the company's billing records further provided us with an exceptional opportunity to compare autobiographically reported durations with immaculately accurate electronic records. Consequently, our findings provide rich insights into the syst-

\[ \text{The computation is as follows: } (1.498 \times 1.0493) - 1 \times 100\% = 20.65\% \]
tematic errors and biases in survey responses at the individual as well as the aggregate level.

**Findings and Marketing Implications**

As the marketplace becomes more diverse and competitive, it is now more important for marketers to have an accurate picture of actual consumer behavior. In this regard, we believe that our findings offer some interesting practical implications. First, our study shows that consumer survey reports are, on average, exaggerated relative to actual frequency and duration. This implies that marketers relying on survey methods must be aware of the danger of making overly optimistic market size estimations or sales forecasts.

Second, we confirm that the often observed regressive effect also exists in the context of autobiographical recollections of both behavioral frequency and duration. This implies that the distribution of behavioral frequencies and durations will appear more concentrated in surveys than they actually are. Consequently, marketing managers using survey reports should be cautioned that they might be underestimating the degree of consumer heterogeneity for both the frequency and duration of usage. These observations pertain mainly to reasonably frequent behaviors that are likely to be recalled using the rate-based estimation method.

Third, although we find that survey reports from both high- and low-frequency consumers are subject to errors as expected, we make the additional discovery that survey reports are less accurate among low-frequency respondents at the aggregate level. Similar results were found for duration. This means that estimating the frequencies and durations is more challenging in low-frequency and short-duration cases than in high-frequency and long-duration cases and that the survey method is less reliable for obtaining such information from the former than from the latter. If surveys are inevitable, marketing managers should know, at the very least, that heavy users tend to underestimate whereas light users tend to overestimate and that the reports from light users tend to be more inaccurate. We also note that, for low-frequency cases, it might be better to ask respondents to count specific occurrences in a given reference period.

Finally, our investigation yields two additional interesting insights. First, we showed that the regressive effect occurs because a greater proportion of consumers report their behavioral frequencies and durations as if they were closer to the mean than they actually are. Second, we showed that overall exaggeration in surveys occurs primarily because the degree of overestimation exceeds that of underestimation, and surprisingly, this is true for both the high-frequency/long-duration groups and the low-frequency/short-duration groups. In other words, regression toward the mean is characterized by proportion, whereas exaggeration in surveys is driven by magnitude.

In summary, we show that systematic biases and errors distort consumer survey results predictably. Specifically, all our results suggest that autobiographical survey reports tend to be shifted to the right and compressed to the mean, which leads to overestimation of the mean and underestimation of heterogeneity. These results do not necessarily imply that such biases and errors are so great as to make consumer surveys totally useless for marketing research purposes. We also find it comforting that the inaccuracy is more problematic for light users, who are generally of less interest to managers. Nevertheless, our findings strongly suggest that usage rate information in consumer survey reports must be used with caution.

**Limitations and Future Research Directions**

Despite the many advantages of using the AT&T data, they also pose some limitations to our study. In the case of frequency, because the survey and diary data were collected at the individual level and the billing records are at the household level, we could not identify the long-distance telephone callers from billing records. Otherwise we could have made direct comparisons between the survey reports and actual frequencies of usage. We therefore had to resort to the use of less-than-perfect but relatively accurate diary entries as anchors to measure accuracy. In the future, perhaps studies can be done on activities such as using private pagers or logging onto the Internet, for which electronic records of usage particular to the individual can be retrieved by user name. In the case of duration, this problem did not arise, because we were able to match exactly the reported calls by time and date and telephone number with actual billing records to measure the accuracy of reported durations.

Also, the database did not provide enough information to identify the underlying cognitive processes and various mediating factors that led to the observed effects. Although there have been many studies that have demonstrated inaccuracies in surveys, only a few have begun to suggest explanations (e.g., Blair and Burton 1987; Burton and Blair 1991; Felcher and Calder 1991; Hu, Toh, and Lee 1996; Menon 1993, 1997; Menon, Raghunib, and Schwarz 1995). Further research in the direction of recent studies by Conrad and Brown (1994) and Schwarz and Sudman (1994) should provide additional insights into the cognitive processes involved in accurately estimating behavioral frequency and duration.

In our study, the upward bias in survey reports is greater for frequency than for duration, whereas for compression the reverse is true. It would be interesting to study the degrees of mean shifts and compressions in consumer surveys of usage frequency and duration across different product and service categories as well as activities. Armed with empirical results, the marketing manager can then make the necessary shifts and decompressions to adjust for the varying degrees of systematic distortions in survey data.

Finally, one natural extension of our study is to investigate other categories in which a key subcomponent of consumption rate is either consumption amount (how many units) or spending amount (how many dollars), instead of usage duration. Another tangential research direction is to investigate the impact of self-perception of usage rates on actual usage behavior, the reverse of the relationship investigated in this article. For example, researchers could test whether the underestimation of consumption among heavy users is the mechanism that sustains such behavior and whether making the actual usage rate more cognizant to heavy users can reduce the problem of excessive consumption.

---

4Note that in the case of very low frequencies, people who use the recall-and-count strategy of memory retrieval may underestimate actual frequencies because of recall loss (Sudman, Bradburn, and Schwarz 1996)
REFERENCES


Huang, Michael Y., Rex S. Tih, and Eunkyu Lee (1996), "Impact of the Level of Aggregation on Response Accuracy in Surveys of Behavioral Frequency," Marketing Letters, 7 (October), 371–82.


——— and Seymour Sudman (1994), Autobiographical Memory and the Validity of Prospective Reports. New York: Springer.
