Journal of Business, Industry, and Economics
Volume 27, Spring/Fall 2022, 112-144

Common Error in Brand Perception Survey Research: Cell Assignment Strategies

Timothy Norvell* and Erin Gillespie§

Abstract: Today marketing researchers are faced with many challenges in the pursuit of high quality, affordable survey data collection. Frequently researchers would like to measure the perceptions of multiple products or brands and then gather detailed information about each. However, when respondents qualify to answer questions about many of the entities (i.e., brands or products), there may not be enough time to answer questions about all of them. It becomes even more challenging to achieve the necessary sample size per entity when invariably some of the entities are of low incidence. Therefore, the researchers typically assign the respondents to answer questions about a subset of the entities using one of the three most common strategies. Given that incentives are frequently used to address non-response, the assignment strategy is used to minimize the sample size and cost as well as ensure the quality of the data. This paper examines the efficiency and effectiveness of the Random Assignment (randomly assigning respondents to one of the entities for which they qualify), Least-Filled Sampling (assigning respondents to the entity for which they qualify which has the fewest completed responses at the time) and Lowest-Incidence-First Sampling (assigning respondents to the lowest incidence entity for which they qualify) using a study from the Home Improvement Research Institute which evaluated products from fifteen different home improvement projects. The results will be discussed, including an evaluation of bias and error, and two possible solutions to address the error are presented.

Key Words: Sampling, Sampling Bias, Low-Incidence, Least-Filled, Random Assignment, Lowest Incidence First, Cell Assignment

* Associate Professor, Elon University
§ Associate Professor, University of North Alabama, Contact Author

Introduction

Many companies need to evaluate the perception of their brand or products versus their competition. This requires marketing researchers to collect in-depth information on customer preferences and attitudes about multiple entities of interest (e.g., brands, products, retail establishments, restaurants, etc.) within their industry through survey research (Dyson, Farr, and Hollis 1996). Typically, these surveys contain detailed measurements of items like awareness, incidence, usage, perceptions, satisfaction, and repurchase intent among those who are qualified to answer questions about one or more of the entities studied (Park and Srinivasan 1994). It is very common to see varied experience within any product or brand category where some customers may be knowledgeable of only one brand while others may be familiar with many or even all of the brands. But if those familiar with many brands answered all of the questions across all brands, the survey would be very long. This length will result in respondent fatigue and lower quality data (Lehmann, McAlister, and Staelin 2011; Petychev and Petychev 2017). It will also increase the cost of the data collection as many respondents will break off (Peytchev 2009; Stussman, Taylor, and Riddick 2003; Galesic 2006) or not participate at all (Vehovar, Zenel, Manfreda, and Zaletel 2002; Groves and Peytcheva 2008). Non-response, in general, has been increasing over time (De Heer and De Leeuw 2002). However, researchers are tasked with collecting very detailed information in a way to minimize total cost while ensuring high data quality. To address the issues, researchers are forced to limit the number of entities that a customer evaluates in a survey. If a customer uses many of the brands or products in the study, they must be assigned to answer questions about only a small subset of them.

The researchers will likely set quotas for minimum sample size per entity (e.g., brand or product) under study to ensure the ability to conduct proper analysis (Berinsky 2006). But in the

very likely scenario when one or more of the entities are of lower incidence (Berry, Chouhoud, and Junn 2016), research must efficiently assign customers to entities in the survey in order to prevent excessive sample size and cost issues, while still ensuring high quality data. Otherwise, if those who qualify for both low and high incidence entities are frequently assigned to the high incidence ones, a much larger sample will be needed to achieve the necessary sample size for the low incidence entities. But, this will result in unnecessarily large and costly oversampling of high incidence entities. Given that almost all brand perception studies will include lower incidence brands or products, this problem is very common.

Also, having to assign some respondents to answer questions about a subset of entities they use, instead of all, can introduce unintended error. For example, young male professionals tend to spend a great deal of money dining out at many different kinds of restaurants. If in a survey, these young professionals are only able to answer questions about a few, instead of all the restaurants they visit, then this group will be underrepresented in the data. Those brands that they used, but were not asked about in the survey, will not have the data about their experiences and perceptions (will only have data from the subset assigned to their brand).

To better understand how marketing researchers address this issue, qualitative in-depth phone interviews were conducted with senior leaders of four large marketing research firms. These interviews lasted approximately 30 minutes and covered three main topics: incidence of this problem and when it occurred, methods employed to address the problem, and any validation of the success of the different solutions. The first part of the interview confirmed that properly sampling of populations with high and low incidence subjects (brands) in the same study was a problem that was frequently encountered by all four firms. They indicated that this issue arose in

every brand tracking or brand evaluation study conducted in an industry with several or more competitors.

According to these research firms, one of three strategies is typically employed (random assignment, least-filled, and lowest incidence first) to assign respondents to entities (i.e., brands or products) in a survey order to address sample size, cost, and data quality issues. Although all four firms believed that the least-filled method would yield the smallest necessary sample size, none knew of the effects on sample representativeness. An extensive search of the academic literature failed to find any research comparing the three methods statistically. More specifically, does employing an assignment strategy result in a low cost, high quality sample or does it introduce unintended bias to the data? Therefore, the purpose of this research is to demonstrate and quantify the error associated with each of the three. As Sarstedt et al. (2018) note, sampling method and composition are essential building blocks to research design, but studies examining sampling issues are almost non-existent in marketing research. Thus, this paper aims to contribute to this important discussion.

The following section will review the common, but complex, situation involving a survey of multiple entities (e.g., brands or products) of interest with varying incidence rates where minimizing total sample size and cost while maintaining data quality is critical (Berry, Chouhoud, and Junn 2016). This paper addresses the three most common methodologies to handle this situation, testing the efficacy of each of the three in providing accurate results using the same data set while measuring the potential error associated with each approach. We then discuss potential solutions to remedy the shortcomings found.

Background

Sampling with the Presence of Low Incidence Entities

As noted earlier, marketing researchers often want to understand the customers' perceptions of their brands or products along with those offered by competitors. However, the desire for depth of research must be balanced with the need to minimize respondent fatigue. As such, in situations when respondents are qualified to answer questions about many entities (i.e., brands or products) covered in the survey, it is not feasible for them to answer detailed questions about all of them. Therefore, it is often necessary to limit the respondent to answering detailed questions about only one or two entities and to set quotas for a minimum sample size per entity for analyses purposes (Adiguezel and Wedel 2008). But because a respondent could qualify to answer questions about as few as one entity or as many as every entity in the study, split questionnaire design (dividing questionnaires into parts and administering each to a randomly selected group of entities) will not effectively assign respondents to cells (Peytchev and Peytchev 2017). Also, the incidence of usage of the entities in the study is likely to vary significantly from customer to customer as well. If the incidence for an entity is very low, assigning customers to answer questions about higher incidence entities may make it difficult to find enough respondents to meet the minimum sample size. The combination of the presence of low incidence entities and varied levels of customer entity experience require a specific cell assignment strategy.

Commonly Used Respondent Cell Assignment Strategies

According to the qualitative interviews with marketing research firms, they typically apply one of three strategies to assign respondents (i.e., customers) to one cell (or more, if time

allows) representing the detailed set of questions for one entity (e.g., product, brand) of the multiple for which they qualify.

The most commonly used method is the random assignment of respondents to a set of questions regarding an entity (Malhotra, 2012). In this approach the researcher randomly assigns respondents to a cell (set of questions) from the list of entities for which they qualify (henceforth "Random Assignment"). This process continues until every cell achieves the minimum targeted sample size.

Random assignment is very effective in reaching the sample size goal for each entity when the incidence of each entity in the study is fairly comparable. However, if the study includes one or more entities with low incidence and many respondents qualify for other entities (cells), random assignment will require a considerably larger sample size. That is because many of the respondents that qualify for the low incidence entities (cells) will be randomly assigned to answer questions about other entities at a rate of (n-1)/n) where n equals the number of entities for which the respondent qualifies. Therefore, the higher incidence cells will quickly exceed the minimum target sample size while researchers will need to continue sampling until they achieve the sample size quota for the low incidence cells. This can be particularly problematic when the availability of sample is limited (e.g., new brands, rare products) or is very costly (e.g., high-end products, surveying physicians).

To reduce the sample needed and the overall cost of the study associated with random assignment while maintaining data quality, researchers often employ two alternative sample assignment strategies. The first one is called Least-Filled Sampling (henceforth "Least-Filled") which has been used in the medical (e.g., Farrelly et al., 2015) and marketing research fields. In

this method, respondents are iteratively assigned to the cell, for which they are qualified, with the lowest number of completed samples at the time of the assignment. The process continues until all cells have their targeted sample size. This ensures that the lower incidence cells will get priority as the higher incidence cells naturally fill up sooner.

The second alternative to random sampling and our third strategy is called Lowest Incidence First Sampling (henceforth "Lowest Incidence First"). This method gives priority to low incidence entities (Berry, Chouhoud, and Junn 2016) and assigns respondents to the cell, for which they qualified, with the lowest incidence. If two or more cells for which the respondents qualified have the same low incidence rate, they are randomly assigned to one of the cells. Again, the process continues until all cells have reached the targeted sample size. The lowest incidence entities are typically identified through one of two ways. The first is the incidence rates gathered from previous research studies. If no such data exists, then firms will use market share or retail unit count, if applicable.

Each of the three approaches has different effects on the total sample size needed to achieve the targeted sample size within each of the cells which impacts the cost of the research. Additionally, when respondents qualify for a range in the number of entities from as little as one to as many as all entities, assigning respondents to just one cell potentially introduces bias into the data and reduces data quality through non-response error (Raghunathan and Grizzle 1995). This is because those who qualify for only one entity will be over-represented in each cell because those who qualify for multiple cells will often be assigned to other cells. Depending on the number of cells, the incidence of each, and the percentage of respondents that qualify for more than one, this bias and error can be quite significant. For example, those who only visit one restaurant and only qualify to answer questions about that one restaurant may be less affluent

or more brand loyal than someone who visits multiple restaurants. In the following sections, we demonstrate and quantify the error associated with each of the three common respondent assignment methodologies. We then offer possible suggestions to address the error caused by these strategies.

Study

Methodology

To evaluate the effect of the different methods used in practice to assign customers to a set of questions for a single entity when they qualify for multiple, we used the data from the Home Improvement Research Institute Study. Biennially, the Home Improvement Research Institute (HIRI) conducts a syndicated marketing research study for its member companies. These companies either own home improvement retail outlets or sell products within them. The study has two primary objectives: to obtain an overview of consumer behavior in the home improvement market and to gain a better understanding of the motivation for each of the 15 specific project and product related decisions. HIRI and its members use these findings to identify unmet needs, to formulate new marketing messages, and to identify new initiatives to drive industry growth. This data was provided to the authors to evaluate the potential error introduced by the chosen cell assignment strategy and to determine if other strategies are more effective.

A total of approximately 2,700 people complete the survey biennially. In the data set used for our analysis, 2,680 finished the survey. These sample members were selected from a representative sample drawn from a TNS (a large survey marketing research firm) online panel of U.S. citizens. To qualify for the study, the sample members must: be between 25 and 70 years

old, own their residence, be the primary decision maker for home improvement projects, and have completed at least one of 15 home improvement projects in the past twelve months. About 35% of the panel population qualified for the study and the response rate was estimated at 40%. Each respondent was compensated $4 for their participation based on the suggestion from the panel company given the length of the survey and the desired response rate. The survey consisted of 25 general home improvement questions (answered by all respondents) and then followed with a series of 28 to 30 specific questions for each project and the products used to complete it (answered only by those assigned to that cell). The general questions included identifying which projects they completed, how much money they spent on home improvement in total, and demographics. This information was used to assign participants to specific cells. But since the general questions identified every project they completed, we were able to determine the effect of the three assignment strategies on the composition of those assigned to answer questions about a project versus all of the respondents who were qualified to answer those questions. Many respondents conducted more than one project (M = 2.9 projects), but given the large number of specific product questions for each project (28 – 30), there was only enough time for each person to cover one project in detail (See Table 1, below).

Robustness Checks

The sample was randomly selected from a nationally representative subgroup of the TNS online panel. The panel company used both probability and non-probability procedures to recruit the members. Panel members are randomly intercepted through an initial screening survey placed on numerous websites. After the initial survey, they are added to the panel with preference given to underrepresented demographics in the existing panel. Therefore, we examined the sample composition relative to known market information and then performed

robustness checks of our findings within four key demographic variables and two dependent measures (Berrens et al., 2003).

Table 1-Project Listing, Incidence and Projects Per Cell

	Projects Completed By Respondents	**Project Incidence**	**Average Number of Projects for Respondents Within each Cell**
1	Added a Bathroom	5.7%	6.1
2	Replaced Siding	6.7%	5.6
3	Added a Room	8.7%	5.5
4	Rewired Electric	11.0%	6.2
5	Remodeled Room	11.2%	5.7
6	Replaced Roof	11.5%	4.9
7	Built a Deck/Patio	14.9%	4.7
8	Painted Exterior	16.3%	4.9
9	Replaced Windows	16.7%	5.2
10	Remodeled Kitchen	17.1%	5.6
11	Remodeled Bath	18.8%	5.2
12	Replaced Floor	22.0%	5.2
13	Plumbing Work	30.0%	4.7
14	Landscape Work	39.0%	4.1
15	Painted Interior	45.1%	4.1

First, there are no significant differences in the age categories of our sample compared to the age categories of those who did home improvement projects listed in the American Housing Survey conducted by the U.S. Census Bureau. Although our sample is slightly more affluent (42% with HH income over $75K versus 39% for the American Housing Survey home improvement section), this may be due to our survey excluding some small, less expensive home improvement projects. The American Housing survey does not collect gender information, but the U.S. Bureau of Labor Statistics indicates that women are more likely to be homeowners than men, driven primarily by a higher ownership rate among single, divorced, widowed or separated women. Therefore, our sample composition of 54% female appears reasonable.

We also conducted robustness checks on our findings within the four key demographic variables of gender, age, income, and education. To enable enough sample size within each of the 15 projects for analysis, we first created dichotomous variables for age (25 – 49 and 50 – 70), income (under $75,000 and $75,000+) and education (no college degree and college degree). Then we ran analyses to determine if our findings and conclusions were consistent among two critical dependent measures (average number of projects per cell and average total cost of each project) within each of these newly created dichotomous variables. We found that the results were very consistent throughout all of the demographics and the two dependent variables (Please see Appendix Tables A through H). These findings provide even more evidence to support our assertion that any respondent cell assignment strategy that assigns a respondent to a single cell (or two) when they qualify for more introduces bias. This is because those who completed multiple projects are underrepresented in every cell because they only answer questions about one entity instead of all for which they qualify. Both probability and non-probability samples will suffer from this bias.

Project Sample Design and Assignment

HIRI set the target sample size for each of the 15 cells (projects) at 155 because they felt this enabled enough statistical precision for their planned sub-group data analyses and kept the total cost of the study within the approved budget. Through a prior study, HIRI learned that the incidence for each of the projects was different and varied significantly from a low of about 6% for those who "Added a Bath" to a high of 45% of those who "Painted the Interior." Therefore, some cells are much more difficult to fill than others. Given that many respondents completed multiple projects, a simple random assignment of respondents to one of the cells for which they qualified would result in frequently assigning low incidence participants to higher incidence

cells, making it even more difficult to find those who qualify for the low incidence projects. Ultimately, this would lead to a much larger sample size and a much more expensive study ($7 per completed survey to cover the incentive and panel acquisition costs).

To minimize the sample size needed and the cost of the study, while still obtaining high quality data, HIRI employed the Least-Filled methodology. Each respondent was assigned to the project for which they qualified that had the lowest number of completes at that time in the data collection process. Then they answered the set of questions for only that specific project. In cases when respondents qualified for two or more cells that were tied for the fewest completes, they were randomly assigned to one of them. This process continued until each cell had at least 155 completes.

Because we have the incidence for all 15 projects and know which projects each respondent completed, we were able to use this data to simulate the other two approaches: Lowest Incidence First and Random Assignment. For the Lowest Incidence First approach, the respondent was assigned to the lowest incidence project (based on the data from this survey) for which they qualified. All of the projects had different incidence rates, so there was no need for a "random assignment tie-breaker" if two groups had the same incidence.

In the Random Assignment approach, we randomly assigned the participants to one of their qualified cells. This continued until all projects have reached a sample size of at least 155. But because Random Assignment approaches typically require a larger sample size, we used the entire sample of the study before we achieved the targeted sample size of 155 per project. Therefore, we produced an additional 3,100 cases by randomly drawing an additional 3,100 from

the original 2,680 to use for the remainder of our random assignment. This resulted in a final sample size of 5,780 for this analysis, which included using some cases more than once.

Method Effect on Total Sample Size

Of the three methods, Least-Filled is able to fill the cells with the targeted sample size more efficiently than the other methods. In our study, Least-Filled needed a total sample size of 2,680 to achieve a minimum of 155 completes for each project. Lowest Incidence First needed about 100 more respondents (2,784) to achieve the same minimum sample size per project. Although these results may differ from sample to sample, Least-Filled will better address random variation that would lead to over-assignment of respondents to the lowest incidence cell. For example, if the first four respondents qualified for the two lowest incidence entities, Lowest Incidence First would assign all four to the cell with the lowest incidence. But, Least-Filled would assign two to each entity, which better represents the ratio of their incidences in the total population. This will prevent the lowest incidence cells from filling too quickly with respondents who also qualify for other low incidence entities.

Because the Random Assignment approach continues to assign respondents to cells regardless of incidence, a much larger sample size is needed (5,746) to acquire minimum 155 completes in the low incidence projects (See Table 2). This results in a much larger sample size for the more common projects and a significantly more costly study.

Table 2: Sample Per Project by Methodology

			Lowest	Random
		Least-Filled	Incidence	Assignment
	Incidence	Sample Size	Sample Size	Sample Size
Added a Bathroom	5.8%	156	160	155
Replaced Siding	6.7%	163	164	170
Added a Room	8.7%	156	163	214
Rewired Electric	11.0%	157	175	222
Remodeled Room	11.2%	163	172	222
Replaced Roof	11.5%	167	178	283
Built a Deck/Patio	14.9%	177	232	337
Painted Exterior	16.3%	167	227	364
Replaced Windows	16.7%	178	177	344
Remodeled Kitchen	17.1%	183	158	282
Remodeled Bath	18.8%	163	172	388
Replaced Floor	22.0%	184	155	351
Plumbing Work	30.0%	155	193	572
Landscape Work	39.0%	185	239	910
Painted Interior	45.1%	326	208	956
		2680	**2773**	**5770**

Methodology Effect on Key Dependent Variable

One of the key dependent measures gathered during the survey was the total dollar amount the customer spent on the project. Because the survey identified every project the customer was involved in, we can compare the average dollar amount spent by those who qualified for each project versus those who were actually assigned to it. Our modeling of Lowest Incidence First and Random Assignment also enables us to compare the effect of those two assignment methods on the accuracy of the measurement of the total amount spent. Here is where the introduced bias and error associated with each assignment methodology becomes very apparent. The Least-Filled approach essentially gives priority to the low incidence cells because it will continue to fill the lower incidence cells (where qualified) until they have the same sample

size as the higher incidence cells. Because of this, the spending estimates for the three lowest incidence cells are fairly accurate. However, the estimates for the eleven out of the remaining twelve cells show a significantly lower amount ($p < .05$) for those who were assigned to the cell versus all of those who actually qualified. The estimates using Least-Filled were lower by an average of 56% with a range from 2% to 90% lower. The Lowest Incidence First approach was very similar in accurately representing the overall project spending for the two lowest incidence projects and inaccurately for the remaining ones. The spending estimate for thirteen of the fifteen projects was significantly lower ($p < .05$) than the true expenditure for everyone who qualified for the projects. This approach underestimated total spending by an average of 58% (See Table 3).

Table 3: Average Spend Per Project by Methodology

				Low	Low		
	TRUE	Least Fill	Least Fill	Incidence	Incidence	Random	Random
Project/CELL	Avg Spend	Avg Spend	Spend Diff	Avg Spend	Spend Diff	Avg Spend	Spend Diff
Added a Bathroom	$18,527	$18,182	2%	$18,527	0%	$11,791	36%
Replaced Siding	$14,770	$11,829	20%	$12,573	15%	$15,981	+8%
Added a Room	$19,867	$17,022	14%	**$13,539***	32%	**$12,956***	35%
Rewired Electric	$8,390	**$1,730***	79%	**$1,662***	80%	**$4,750***	43%
Remodeled Room	$9,272	**$3,027***	67%	**$3,260***	65%	$6,003	35%
Replaced Roof	$10,563	**$5,636***	47%	**$5,352***	49%	$8,954	15%
Built a Deck/Patio	$8,920	**$3,693***	59%	**$3,695***	59%	**$5,069***	43%
Painted Exterior	$5,310	**$1,226***	77%	**$1,750***	67%	**$2,406***	55%
Replaced Windows	$9,721	**3,775***	61%	**$3,493***	64%	**$4,908***	50%
Remodeled Kitchen	$8,745	$6,840	22%	**$5,169***	41%	$6,350	27%
Remodeled Bath	$6,480	**$2,887***	55%	**$2,652***	59%	**$3,506***	46%
Replaced Floor	$7,112	**$2,262***	68%	**$1,849***	74%	**$2,643***	63%
Plumbing Work	$5,746	**$235***	96%	**$251***	96%	**$2,622***	54%
Landscape Work	$4,232	**$858***	80%	**$801**	81%	**$1,801***	57%
Painted Interior	$5,025	**$524***	90%	**$329***	93%	**$2,512***	50%
			56%		**58%**		**44%**

*** Significantly Different from True Spend at 95% Confidence Level**

The Random Assignment approach was slightly more accurate in measuring total spending with ten out of the fifteen projects values significantly below ($p < .05$) the true means.

The average amount of error across the projects was 44% with a range of an 8% overestimate to a 57% underestimate. However, overall, none of the three methods were effective in measuring the total amount spent.

Method Effect on the Average Number of Projects per Cell

Another significant effect of using an assignment methodology, when respondents qualify for multiple entities in a survey but are assigned to one cell, is the difference observed in the assigned versus the true average number of projects per cell. For example, in our study if some respondents only qualified for "Remodeled Kitchen," then all of them would obviously be assigned to that cell. However, if some of those respondents qualified for other cells, then some or all could be assigned to other cells (depending on the methodology). Because everyone who only completed one project will be assigned to that respective cell and others who completed multiple projects could be assigned to other cells, the use of an assignment strategy will result in cells containing more respondents that completed only one project than the true sample. Also, when respondents complete a large number of projects, that project number is used in the calculation of the true average of all projects in which they were involved. Yet, in our simulations that project number is only assigned to one cell. To clarify, let's take an example of when someone completed all 15 projects. In the calculation of the true mean number of projects, the value of 15 would be in every project. However, when we assigned respondents, the value of 15 is only placed in one cell (project), instead of all fifteen. Therefore, the average number of projects calculated after our assignment strategy will be lower than the true average in every instance with one exception (See Table 4). The one exception is under the Lowest Incidence First strategy for the "Added Bathroom" cell. Given that this cell has the lowest incidence and that the methodology fills the lowest incidence category first, this cell is assigned every

respondent who has completed that project. Therefore, the true sample and the assigned sample are identical. The average number of projects for the "Added Bathroom" using the Least Fill approach is also not statistically different. For every other cell, the assigned average number of projects is different than the true sample average. Although the average error is almost identical across all three methodologies (43% to 44%), the error in the averages tends to increase as incidence increases for the Least Fill and Low Incidence First approaches. For Random Assignment, the amount of error is fairly consistent from project to project.

Table 4: Average True Number of Projects vs. the Assigned

Project/Cell	True Total Avg. No. Projects	Least Fill Avg. No. Projects	Least-Fill Diff	Low Incidence Avg. No. Projects	Low Incidence Diff	Random Avg. No. Projects	Random Diff
Added Bathroom	6.11	5.93	3%	6.11	0%	3.34*	45%
Replaced Siding	5.58	4.71*	16%	4.72*	15%	2.95*	47%
Added a Room	5.47	3.67*	33%	3.89*	29%	3.26*	40%
Rewired Electric	6.18	4.45*	28%	4.51*	27%	3.09*	50%
Remodeled Room	5.74	3.64*	37%	3.81*	34%	2.99*	48%
Replaced Roof	4.87	2.70*	45%	2.86*	41%	2.71*	44%
Built a Deck/Patio	4.72	1.92*	59%	2.67*	43%	2.77*	41%
Painted Exterior	4.93	2.26*	54%	2.82*	43%	2.85*	42%
Replace Windows	5.18	2.28*	56%	2.31*	55%	2.99*	42%
Remodel Kitchen	5.57	3.60*	35%	2.89*	48%	2.67*	52%
Remodeled Bath	5.17	2.17*	58%	2.12*	59%	2.87*	44%
Replaced Floor	5.22	2.59*	50%	1.93*	63%	2.87*	45%
Plumbing Work	4.70	1.66*	65%	1.82*	61%	2.85*	39%
Landscape Work	4.05	1.36*	66%	1.43*	65%	2.89*	29%
Painted Interior	4.11	2.36*	43%	1.11*	73%	2.91*	29%
			43%		**44%**		**43%**

*** Significantly Different from True Average at 95% Confidence Level**

*** Significantly Different from True Spend at 95% Confidence Level**

Again, the key driver of the error is the difference in the percentage of respondents assigned to each cell who have only completed that one project versus the percentage of those of the true sample that have only completed one project (See Table 5). Similar to the error

observed in the number of projects, the amount of error generally increases as incidence increases for the Least Fill and Lowest Incidence First methods, and the error remains fairly consistent with Random Assignment. Given that we identified that the composition of those assigned to each cell is different than the true sample and that three other measures are significantly different for many of the cells, it is highly likely that the results for many of the project-specific questions are also inaccurate.

Table 5: Percentage of Respondents That Did Only One Project

Projects Completed By Respondents	True Sample Did 1 Project	Least-Filled Did 1 Project	Lowest Incidence Did 1 Project	Random Did 1 Project
Added a Bathroom	16.3%	17.3%	16.3%	27.7%*
Replaced Siding	10.8%	12.9%	12.4%	37.6%*
Added a Room	8.9%	31.4%*	28.8%*	30.4%*
Rewired Electric	7.5%	14.0%*	12.8%	35.6%*
Remodeled Room	8.9%	17.8%*	15.8%*	36.0%*
Replaced Roof	18.3%	36.5%*	31.9%*	40.3%*
Built a Deck/Patio	19.4%	47.5%*	33.6%*	37.7%*
Painted Exterior	12.2%	34.1%*	24.3%*	38.7%*
Replaced Windows	15.1%	40.4%*	39.7%*	38.7%*
Remodeled Kitchen	8.5%	20.8%*	26.8%*	37.6%*
Remodeled Bath	14.2%	46.6%*	43.3%*	35.8%*
Replaced Floor	8.6%	29.9%*	40.4%*	37.0%*
Plumbing Work	10.6%	54.8%*	46.2%*	37.9%*
Landscape Work	14.2%	68.1%*	62.6%*	36.2%*
Painted Interior	15.0%	53.4%*	88.7%*	35.7%*

*** Significantly Different from True Spend at 95% Confidence Level**

Discussion

Often in surveys the researcher's objectives include obtaining awareness and incidence of a large number of relevant entities followed by detailed information about each of them. However, if a respondent uses multiple entities, there may not be enough time to cover all of the detailed questions in the survey (Galesic and Bosnjak 2009). Therefore, the researcher will typically employ one of the three respondent cell assignment strategies discussed here. In this

paper, we tested the efficacy of each of the three methods in providing accurate results using the data from the Home Improvement Research Institute's biennial study.

The Least-Filled and Lowest Incidence methodologies enable the researcher to achieve the targeted sample size per cell with the lowest overall sample size. These two methods needed a sample size less than half of Random Assignment (See Table 6). But, Random Assignment is slightly better at measuring the average total spending for each cell, and comparable to the other two methodologies on measuring average number of projects and the percentage of respondents within each cell that only did one project. Since Least-Filled and Lowest Incidence gave priority to low incidence cells, those two methods were much more accurate among the lowest incidence projects. This accuracy came at the expense of the accuracy of the high incidence projects where Random Assignment was more accurate.

Table 6-Summary of Methodologies

Measure	**Least-Filled**	**Lowest Incidence**	**Random Assignment**
Sample size needed to achieve a minimum cell size of 155	2680	2773	5770
Average total spending error percentage	56%	57%	42%
Average total spending error trend	Increases as incidence increases	Increases as incidence increases	Relatively flat across projects
Average number of projects error percentage	43%	44%	43%
Average number of projects error trend	Increases as incidence increases	Increases as incidence increases	Relatively flat across projects
Average error of the percentage of respondents that only did one project/cell	23%	22%	24%
Average error of the percentage of respondents that only did one project per cell trend	Slight increase as incidence increases	Slight increase as incidence increases	Relatively flat across cells

However, regardless of methodology, a significant amount of error is introduced when respondents are assigned to one (or a subset) of the projects when they qualify for more. All three methodologies create a group composition within each cell that is different than the true sample by assigning a higher percentage of respondents who only did one project and only qualified for that one cell. The assignment methodology essentially creates a non-response error by preventing a specific group of respondents from answering some questions. As we see in our data, the respondents assigned to the detailed questions for each product do not spend as much or complete as many projects as all of those who qualify. This potential for error extends to numerous other research situations. For example, perceptions and purchase behavior could be significantly affected in product or brand studies when the cells are oversaturated with customers who only use that product or brand. These customers may have more loyalty, have stronger perceptions, and have a better overall understanding of the brand than the average user. Another possibility is that they are less affluent customers who infrequent the category and do not purchase multiple brands. If those assigned spend less than the true total, like we observed in our study, then market size and market share estimations will also be significantly inaccurate.

Using these methodologies in pharmaceutical studies may underestimate the interaction and coexistence with other conditions when those who only have one condition are over-represented in each condition under study. In studies to understand the effect of social issues on purchase behavior, an assignment strategy is likely to underrepresent those who are most knowledgeable because they will answer questions about only one topic instead of all for which they qualify. This could lead to an error in the measurement of the public's understanding of and attitudes toward certain topics.

Although two of the methodologies (Least Filled, Lowest Incidence First) are very efficient in filling the cells at a minimum sample size, all three yield a composition of the assigned groups which are not representative of the true sample leading to significant errors on several measures. The error is driven by assigning respondents to answer questions about only one entity when they often qualify for many which results in an overrepresentation of those who did qualify for only one entity and an underrepresentation of those who qualify for many entities in each cell. Therefore, we suggest that researchers avoid using the three methods studied here and, instead, consider using one of two potential approaches below.

Potential Solutions

Initially, we tested two data weighting schemes in an effort to address the differences in the number of projects completed by true sample versus the sample actually chosen given that this appeared as the primary source of the error. The first was an incidence based weighted random assignment where we randomly assigned respondents based on a weighted value which resulted in an equal probability of selection for all projects. The weighting reduced the assignment of those who also qualified for low incidence projects to a high incidence project. Thus, we saw the expected decrease in necessary sample size versus the Random Assignment methodology (4,476 vs. 5,770), but this did not have a measurable impact on accuracy (See Appendix Tables I and J). The second weighting scheme involved weighting each respondent based on the number of projects they completed. This essentially weights those who did multiple projects (heavy users) more which increased their representation across the entities. Given this, we expected to see a decrease in error. This post hoc weighting scheme did reduce the average error in the Total Amount Spent variable for Random Assignment methodology (41% down to

29%, See Appendix Table K). This amount of error may be an acceptable tradeoff for companies with budget concerns.

To take the weighting scheme a step further, a company may wish to increase the sample size to create more a complex and, hopefully, more accurate scheme. This weighting scheme could account for different combinations of projects that drive the error in the key dependent measures. In our example, a scheme that assigned different weights to each project based on complexity may yield greater accuracy.

The final possible solution is one that addresses the source of the problem: assigning respondents to answer questions about a subset of entities instead of all for which they qualify. To accomplish this would require the survey to be shorter and to cover fewer entities at one time. In our example, it would not be possible to answer questions about all 15 projects. The same is true in any industry or sector where awareness and usage of a large number of relevant entities (e.g., competitors) differs across the target population, as it often does. This is why a cell assignment strategy is used in the first place. Therefore, in our study, the researcher must streamline the survey to where the respondent would be able to answer questions about 4 - 5 projects. Then the study would have to be repeated two more times to cover the other projects of interest. To ensure a large enough sample in the low incidence cells, they could be included in the first study and then added to one or both of the next two waves to get to the appropriate sample size. This would clearly increase the total cost of the research but would ensure accuracy.

Conclusion

The purpose of this paper was to demonstrate and quantify the error associated with three common respondent assignment methodologies. Through the analysis of an extensive dataset,

we were able to compare Random Assignment, Least-Filled, and Lowest Incidence First methodologies. We found that while Random Assignment tended to have lower error, it required a far larger sample size, which reduces the practicality of this method. But, ultimately, any time researchers assign respondents to only a subset of entities for which they qualify, they introduce bias and error. The composition of the resulting assigned groups will always differ from the actual ones. Those who qualify for only one or two entities will be overrepresented and those who qualify for many or all will be underrepresented in the assigned groups. As such, we have proposed a few solutions that can reduce and possibly eliminate the error associated with the assignment methodologies. When researchers seek to investigate multiple entities but also need to restrict sample size due to associated costs, we believe these suggestions will allow them to (practically) obtain quality data.

From a practical standpoint, a researcher must always balance the cost of the project with the quality of the data. Having respondents answer all questions about all entities will eliminate the error discussed in this paper, but may not be feasible from a survey length or cost perspective. Therefore, if cell assignment is necessary, a category usage (number of projects/brands used) weighted random assignment provides the highest quality data with a slightly lower required sample than simple random assignment. However, this comes at a higher cost than Least Filled Quota or Lowest Incidence First. If cost is a major concern, of the latter two, Least Filled Quota sampling is superior in both quality and sample size efficiency.

Above all else, when the number of entities (e.g., brands) expand to where an assignment strategy is necessary, the researcher must examine each entity to determine its importance and relevance. As the number of entities increase, fewer qualified respondents will answer questions about each entity, increasing the error. Reducing unnecessary entities (brands), will increase the

quality of the data collected of those that remain in the study. If an assignment strategy is still needed after a reduction in entities, then a category usage weighted random assignment, which provides the highest quality data, may be affordable.

References

Adigüzel, Feray, and Michel Wedel, 2008. Split questionnaire design for massive surveys. *Journal of Marketing Research* 45 (5): 608-617.

Berrens, Robert P., Alok K. Bohara, Hank Jenkins-Smith, Carol Silva, and David L. Weimer, 2003. The advent of Internet surveys for political research: A comparison of telephone and Internet samples. *Political analysis* 11 (1): 1-22.

Berinsky, Adam J, 2006. American public opinion in the 1930s and 1940s: The analysis of quota-controlled sample survey data." *International Journal of Public Opinion Quarterly* 70, (4): 499-529.

Berry, Justin, Youssef Chouhoud, and Jane Junn, 2016. Reaching Beyond Low-Hanging Fruit: Surveying Low Incidence Populations. *The Oxford Handbook of Polling and Polling Methods*.

De Heer, W. and E. De Leeuw, 2002. Trends in household survey nonresponse: A longitudinal and international comparison. *Survey Nonresponse,* 41: 41-54.

Dyson, Paul, Andy Farr, and Nigel S. Hollis, 1996. Understanding, measuring, and using brand equity. *Journal of Advertising Research*, 36 (6): 9-22.

Farrelly, Matthew C., Jennifer C. Duke, Erik C. Crankshaw, Matthew E. Eggers, Youn O. Lee, James M. Nonnemaker, Annice E. Kim, and Lauren Porter, 2015. A randomized trial of the effect of e-cigarette TV advertisements on intentions to use e-cigarettes. *American journal of preventive medicine* 49 (5): 686-693.

Galesic, Mirta, 2006. Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics* 22 (2): 313.

Galesic, Mirta, and Michael Bosnjak, 2009. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly* 73 (2): 349-360.

Groves, Robert M., and Emilia Peytcheva, 2008. The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public Opinion Quarterly* 72 (2): 167-189.

Lehmann, Donald R., Leigh McAlister, and Richard Staelin, (2011). Sophistication in research in marketing. *Journal of Marketing* 75 (4): 155-165.

Malhotra, Naresh K. *Basic marketing research: Integration of social media*. (Boston: Pearson, 2012).

Park, Chan Su, and Vern Srinivasan, 1994. A survey-based method for measuring and understanding brand equity and its extendibility. *Journal of Marketing Research* 31 (2): 271-288.

Peytchev, Andy, 2009. Survey breakoff. *Public Opinion Quarterly* 73 (1): 74-97.

Peytchev, Andy, and Emilia Peytcheva, 2017. Reduction of measurement error due to survey length: Evaluation of the split questionnaire design approach. *Survey Research Methods* 11 (4): 361-368.

Raghunathan, Trivellore E., and James E. Grizzle, 1995. A split questionnaire survey design. *Journal of the American Statistical Association*, *90*(429), 54-63.

Sarstedt, Marco, Paul Bengart, Abdel Monim Shaltoni, and Sebastian Lehmann, 2018. The use of sampling methods in advertising research: A gap between theory and practice. *International Journal of Advertising* 37 (4), 650-663.

Stussman, Barbara J., Beth L. Taylor, and Howard Riddick, 2003. Partials and break-offs in the National Health Interview Survey, 2002. In *Federal Committee on Statistical Methodology (Ed.), Federal Committee on Statistical Methodology Research Conference (Proceedings—Wednesday sessions)*, pp. 21-26.

Vehovar, Vasja, Zenel Batagelj, Katja Lozar Manfreda, and Metka Zaletel. Survey Nonreponse. *Nonreponse in web surveys.* (New York: Wiley,2002).

Appendix

Table A: Robustness Check: Gender by Number of Projects

APPENDIX TABLE A: AVERAGE NUMBER OF PROJECTS IN TOTAL VERSUS THE ASSIGNMENT METHODOLOGIES - GENDER

Project	True Total # Projects Male	Least Fill Male	Low Incidence Male	Random Assign Male	Wtd Rand Assign Male	True Total $ Projects Female	Least Fill Female	Low Incidence Female	Random Assign Female	Wtd Rand Assign Female
Added a Bathroom	6.0	5.9	6.0	**4.0**	**4.2**	6.3	6.0	6.3	5.3	**4.8**
Replaced Siding	5.6	**4.6**	4.7	**3.8**	**4.1**	5.6	4.9	4.8	**3.9**	**4.3**
Added a Room	5.9	**4.3**	**4.3**	**4.1**	**4.1**	6.6	**4.8**	**4.8**	**5.1**	**4.1**
Rewired Electric	5.8	**3.7**	**2.8**	**3.5**	**3.5**	5.4	**3.5**	**3.0**	**3.0**	**3.3**
Remodeled Room	5.7	**3.7**	**4.1**	**3.8**	**4.0**	5.3	**3.7**	**3.7**	**3.1**	**2.3**
Replaced Roof	6.0	**3.5**	**3.8**	**3.8**	**3.3**	5.5	**3.7**	**3.9**	**4.0**	**3.4**
Built a Deck/Patio	4.9	**2.8**	**2.8**	**2.9**	**2.8**	4.9	**2.7**	**2.9**	**2.8**	**2.7**
Painted Exterior	5.4	**2.5**	**1.9**	**2.8**	**3.5**	5.1	**2.7**	**2.0**	**2.9**	**3.4**
Replaced Windows	5.2	**2.2**	**2.2**	**2.9**	**3.0**	5.1	**2.2**	**2.0**	**2.5**	**3.0**
Remodeled Kitchen	5.4	**2.4**	**2.3**	**3.0**	**2.7**	5.0	**2.2**	**2.3**	**2.4**	**2.7**
Remodeled Bath	5.0	**2.1**	**2.9**	**3.0**	**2.8**	4.9	**2.4**	**2.7**	**2.6**	**3.0**
Replaced Floor	5.1	**2.0**	**2.7**	**2.7**	**2.5**	4.4	**1.8**	**2.6**	**2.4**	**2.0**
Plumbing Work	4.7	**1.7**	**1.8**	**2.9**	**2.5**	4.7	**1.6**	**1.8**	**2.9**	**2.7**
Landscape Work	4.3	**1.4**	**1.5**	**2.7**	**2.3**	3.8	**1.3**	**1.4**	**2.4**	**2.2**
Painted Interior	4.4	**2.8**	**1.2**	**2.5**	**2.4**	3.9	**2.1**	**1.1**	**2.3**	**2.2**

Bolded is Significantly Different from True Spend at 95% Confidence Level

Table B: Robustness Check: Education by Number of Projects

APPENDIX TABLE B: AVERAGE NUMBER OF PROJECTS IN TOTAL VERSUS THE ASSIGNMENT METHODOLOGIES - EDUCATION

Project	True Total # Projects No Degree	Least Fill No Degree	Low Incidence No Degree	Random Assign No Degree	Wtd Rand Assign No Degree	True Total No. Projects Degree	Least Fill Degree	Low Incidence Degree	Random Assign Degree	Wtd Rand Assign Degree
Added a Bathroom	6.1	6.1	6.1	**4.9**	**4.9**	6.1	5.9	6.1	**4.6**	**4.3**
Replaced Siding	5.2	4.5	4.6	**3.7**	**4.0**	6.1	**5.0**	**5.0**	**4.2**	**4.6**
Added a Room	6.1	**4.6**	**4.6**	**4.6**	**4.0**	6.3	**4.3**	**4.5**	**4.4**	**4.4**
Rewired Electric	5.4	**3.4**	**2.8**	**3.0**	**3.3**	5.8	**3.8**	**3.0**	**3.6**	**3.6**
Remodeled Room	5.1	**3.5**	**3.8**	**2.7**	**2.6**	5.9	**4.0**	**4.0**	**4.1**	**3.5**
Replaced Roof	5.7	**3.2**	**3.7**	**3.9**	**3.1**	5.8	**4.0**	**4.0**	**4.0**	**3.6**
Built a Deck/Patio	4.8	**2.7**	**2.8**	**3.0**	**2.6**	5.0	**2.7**	**2.9**	**2.7**	**2.8**
Painted Exterior	5.0	**2.6**	**1.9**	**2.8**	**3.4**	5.4	**2.6**	**2.0**	**3.0**	**3.5**
Replace Windows	5.2	**2.1**	**2.0**	**2.7**	**3.3**	5.1	**2.2**	**2.2**	**2.7**	**2.7**
Remodel Kitchen	5.1	**2.1**	**2.2**	**2.4**	**2.8**	5.3	**2.4**	**2.3**	**2.9**	**2.6**
Remodeled Bath	4.8	**2.3**	**2.8**	**2.8**	**2.9**	5.1	**2.3**	**2.8**	**2.8**	**2.9**
Replaced Floor	4.6	**2.0**	**2.6**	**2.6**	**2.1**	4.9	**1.9**	**2.7**	**2.5**	**2.4**
Plumbing Work	4.7	**1.6**	**1.8**	**3.0**	**2.8**	4.8	**1.7**	**1.9**	**2.7**	**2.4**
Landscape Work	4.0	**1.3**	**1.4**	**2.4**	**2.3**	4.1	**1.4**	**1.5**	**2.7**	**2.3**
Painted Interior	4.2	**2.6**	**1.1**	**2.5**	**2.2**	4.0	**2.2**	**1.1**	**2.3**	**2.3**

Bolded is Significantly Different from True Spend at 95% Confidence Level

Table C: Robustness Check: Age by Number of Projects

APPENDIX TABLE C: AVERAGE NUMBER OF PROJECTS IN TOTAL VERSUS THE ASSIGNMENT METHODOLOGIES - AGE

Project	True Total # Projects Younger	Least Fill Younger	Low Incidence Younger	Random Assign Younger	Wtd Rand Assign Younger	True Total # Projects Older	Least Fill Older	Low Incidence Older	Random Assign Older	Wtd Rand Assign Older
Added a Bathroom	5.6	5.5	5.6	**4.3**	**3.6**	6.9	6.6	6.9	**5.2**	**5.7**
Replaced Siding	5.9	5.0	5.1	**4.4**	**4.4**	5.5	4.7	4.6	**3.5**	**3.9**
Added a Room	7.1	**5.6**	**5.4**	**6.6**	**5.2**	5.6	**3.9**	**4.1**	**3.6**	**3.6**
Rewired Electric	5.9	**3.8**	**2.8**	**3.1**	**3.5**	5.4	**3.6**	**3.0**	**3.4**	**3.5**
Remodeled Room	5.7	**3.7**	**3.8**	**3.8**	**2.8**	5.3	**3.6**	**3.8**	**3.0**	**3.1**
Replaced Roof	5.7	**3.3**	**3.6**	**3.8**	**3.1**	5.7	**3.9**	**4.0**	**4.1**	**3.6**
Built a Deck/Patio	5.6	**3.0**	**3.1**	**2.9**	**2.7**	4.6	**2.6**	**2.8**	**3.0**	**2.8**
Painted Exterior	5.7	**2.9**	**1.9**	**3.1**	**4.3**	5.1	**2.6**	**2.0**	**3.0**	**3.1**
Replaced Windows	5.6	**2.0**	**2.4**	**2.9**	**3.4**	5.0	**2.2**	**2.0**	**2.7**	**2.9**
Remodeled Kitchen	5.8	**2.4**	**2.4**	**2.9**	**2.9**	5.1	**2.2**	**2.3**	**2.8**	**2.7**
Remodeled Bath	6.0	**2.4**	**3.0**	**3.5**	**3.6**	4.4	**2.3**	**2.8**	**2.6**	**2.7**
Replaced Floor	4.9	**2.0**	**3.0**	**2.9**	**2.5**	4.6	**1.9**	**2.5**	**2.2**	**2.1**
Plumbing Work	5.2	**1.6**	**1.9**	**2.9**	**2.9**	4.5	**1.7**	**1.8**	**2.9**	**2.5**
Landscape Work	4.4	**1.5**	**1.5**	**2.8**	**2.2**	3.9	**1.3**	**1.4**	**2.4**	**2.3**
Painted Interior	4.2	**2.4**	**1.1**	**2.3**	**2.2**	4.1	**2.4**	**1.1**	**2.5**	**2.3**

Bolded is Significantly Different from True Spend at 95% Confidence Level

Table D: Robustness Check: Income by Number of Projects

APPENDIX TABLE D: AVERAGE NUMBER OF PROJECTS IN TOTAL VERSUS THE ASSIGNMENT METHODOLOGIES - INCOME

Project	True Total # Projects Lower	Least Fill Lower	Low Incidence Lower	Random Assign Lower	Wtd Rand Assign Lower	True Total #4 Projects Higher	Least Fill Higher	Low Incidence Higher	Random Assign Higher	Wtd Rand Assign Higher
Added a Bathroom	5.8	5.5	5.8	4.8	4.9	6.5	6.4	6.5	**4.2**	**4.6**
Replaced Siding	4.9	4.4	4.5	**3.9**	**3.7**	6.5	**5.4**	**5.2**	**4.1**	**5.0**
Added a Room	5.7	**4.5**	**4.4**	**3.9**	**4.1**	7.0	**4.6**	**5.1**	**5.0**	**4.7**
Rewired Electric	5.4	**3.3**	**2.9**	**3.1**	**3.4**	5.9	**4.1**	**2.8**	**3.5**	**3.6**
Remodeled Room	5.2	**3.9**	**4.1**	**3.4**	**2.8**	5.8	**3.5**	**3.7**	**3.8**	**3.4**
Replaced Roof	5.9	**3.6**	**3.9**	**4.4**	**3.5**	5.8	**3.9**	**3.9**	**3.9**	**3.4**
Built a Deck/Patio	4.4	**2.5**	**2.6**	**2.8**	**2.4**	5.4	**3.0**	**3.2**	**3.4**	**3.3**
Painted Exterior	5.0	**2.4**	**1.8**	**2.9**	**3.2**	5.5	**2.8**	**2.2**	**3.1**	**4.0**
Replaced Windows	5.0	**2.1**	**2.1**	**2.5**	**3.1**	5.4	**2.1**	**2.0**	**3.0**	**3.0**
Remodeled Kitchen	5.0	**2.2**	**2.4**	**2.5**	**2.8**	5.7	**2.4**	**2.3**	**3.2**	**2.8**
Remodeled Bath	4.8	**2.4**	**2.8**	**3.0**	**2.5**	5.2	**2.2**	**2.9**	**2.8**	**3.2**
Replaced Floor	4.5	**1.9**	**2.4**	**2.4**	**2.2**	5.0	**2.0**	**2.9**	**2.9**	**2.3**
Plumbing Work	4.6	**1.5**	**1.6**	**3.1**	**2.4**	5.0	**1.8**	**2.0**	**2.6**	**2.5**
Landscape Work	4.1	**1.3**	**1.4**	**2.3**	**2.3**	4.2	**1.4**	**1.5**	**2.8**	**2.3**
Painted Interior	4.0	**2.3**	**1.1**	**2.2**	**2.2**	4.3	**2.5**	**1.2**	**2.6**	**2.3**

Bolded is Significantly Different from True Spend at 95% Confidence Level

Table E: Robustness Check: Income by Total Dollar Amount Spent

APPENDIX TABLE E: AVERAGE NUMBER OF PROJECTS IN TOTAL VERSUS THE ASSIGNMENT METHODOLOGIES - GENDER

Project	True Total # Projects Male	Least Fill Male	Low Incidence Male	Random Assign Male	Wtd Rand Assign Male	True Total $ Projects Female	Least Fill Female	Low Incidence Female	Random Assign Female	Wtd Rand Assign Female
Added a Bathroom	$17,384	$15,142	$14,499	$12,116	$13,056	$19,949	$22,119	$23,536	$11,208	$18,283
Replaced Siding	$11,147	$10,415	$9,953	$12,643	$9,017	$18,506	$13,191	$15,161	$18,948	$18,868
Added a Room	$6,879	**$1,954**	**$1,905**	**$2,654**	$4,152	$10,605	**$1,338**	**$1,297**	$8,195	$7,344
Rewired Electric	$8,430	$6,917	$5,916	**$5,497**	$8,218	$9,052	$6,770	**$4,628**	$7,222	$7,127
Remodeled Room	$20,218	$16,134	*$14,003*	$16,780	$20,200	$19,534	$17,672	*$13,137*	**$9,726**	**$11,581**
Replaced Roof	$11,249	**$2,595**	**$3,137**	**$3,618**	**$3,484**	$7,295	**$3,385**	**$3,359**	$8,183	**$4,100**
Built a Deck/Patio	$11,482	**$5,693**	**$5,416**	*$8,501*	**$7,433**	$9,704	**$5,591**	**$5,301**	$9,343	**$5,539**
Painted Exterior	$8,123	**$2,079**	**$1,574**	**$2,029**	**$3,296**	$6,271	**$2,403**	**$2,081**	**$3,085**	**$4,249**
Replaced Windows	$7,072	**$2,710**	**$2,544**	**$3,366**	**$2,789**	$5,878	**$3,030**	**$2,746**	**$3,647**	**$2,760**
Remodeled Kitchen	$9,895	**$3,620**	**$2,872**	**$5,476**	**$4,692**	$9,575	**$3,886**	**$3,903**	**$4,460**	**$5,096**
Remodeled Bath	$4,748	**$852**	**$1,652**	**$2,595**	*$2,819*	$5,844	**$1,502**	**$1,838**	**$2,255**	$4,325
Replaced Floor	$11,349	**$3,775**	**$3,734**	**$3,787**	**$3,906**	$6,788	**$3,625**	**$3,667**	$6,133	**$3,443**
Plumbing Work	$6,607	**$188**	**$222**	**$2,771**	**$1,158**	**$4,896**	**$282**	**$278**	**$2,488**	**$1,556**
Landscape Work	$3,591	**$759**	**$750**	**$2,039**	**$1,176**	**$4,725**	**$912**	**$833**	**$1,623**	**$1,857**
Painted Interior	$4,129	**$566**	**$344**	**$2,627**	**$1,749**	**$5,661**	**$499**	**$323**	**$2,447**	**$2,466**

Bolded is Significantly Different from True Spend at 95% Confidence Level
***Italics* is Significantly Different from True Spend at 90% Confidence Level**

Table F: Robustness Check: Education by Total Dollar Amount Spent

APPENDIX TABLE F: AVERAGE NUMBER OF PROJECTS IN TOTAL VERSUS THE ASSIGNMENT METHODOLOGIES - EDUCATION

Project	True Total # Projects No Degree	Least Fill No Degree	Low Incidence No Degree	Random Assign No Degree	Wtd Rand Assign No Degree	True Total No. Projects Degree	Least Fill Degree	Low Incidence Degree	Random Assign Degree	Wtd Rand Assign Degree
Added a Bathroom	$20,320	$21,253	$20,230	$15,827	*$11,683*	$17,503	$16,454	$17,628	*$9,931*	$18,127
Replaced Siding	$12,236	$8,992	$8,707	$7,933	$9,081	$18,379	$16,342	$18,532	*$28,444*	$21,872
Added a Room	$6,024	**$1,305**	**$1,117**	**$2,140**	**$1,787**	$10,943	**$2,209**	**$2,266**	$7,540	$9,161
Rewired Electric	$9,067	**$5,448**	**$3,680**	**$4,417**	$6,077	$8,493	$8,059	$6,469	$8,283	$9,376
Remodeled Room	$25,180	**$11,614**	**$10,741**	**$11,862**	**$10,333**	$13,583	**$24,593**	$17,636	$14,204	*$21,542*
Replaced Roof	$9,176	**$2,005**	**$2,815**	**$4,211**	**$3,164**	$9,326	**$3,855**	**$3,686**	$8,234	**$4,389**
Built a Deck/Patio	$8,230	**$5,074**	**$5,062**	$8,630	**$6,327**	$12,952	**$6,300**	**$5,677**	$9,268	**$6,485**
Painted Exterior	$5,911	**$1,854**	**$1,276**	**$1,862**	**$2,572**	$8,386	**$2,747**	**$2,599**	**$3,490**	$5,292
Replace Windows	$5,792	**$2,015**	**$1,711**	**$1,988**	**$2,312**	$7,151	**$3,707**	**$3,404**	*$4,848*	**$3,199**
Remodel Kitchen	$9,369	**$2,844**	**$2,689**	**$4,758**	**$5,081**	$10,205	**$4,654**	**$4,564**	**$5,163**	**$4,789**
Remodeled Bath	$6,772	**$976**	**$1,431**	**$2,396**	$4,918	$4,134	**$1,454**	**$2,034**	**$2,437**	**$2,341**
Replaced Floor	$6,765	**$2,632**	**$2,631**	**$3,605**	**$3,742**	$11,177	**$4,790**	**$4,816**	**$6,775**	**$3,609**
Plumbing Work	$5,783	**$261**	**$288**	**$2,353**	**$1,541**	$5,778	**$218**	**$221**	**$3,004**	**$1,164**
Landscape Work	$4,874	**$755**	**$690**	**$1,593**	**$1,090**	$3,695	**$962**	**$904**	**$2,004**	**$2,019**
Painted Interior	$4,336	**$596**	**$319**	**$2,379**	**$2,081**	$5,440	**$478**	**$343**	**$2,653**	**$2,332**

Bolded is Significantly Different from True Spend at 95% Confidence Level
***Italics* is Significantly Different from True Spend at 90% Confidence Level**

Table G: Robustness Check: Age by Total Dollar Amount Spent

APPENDIX TABLE G: AVERAGE NUMBER OF PROJECTS IN TOTAL VERSUS THE ASSIGNMENT METHODOLOGIES - AGE

Project	True Total # Projects Younger	Least Fill Younger	Low Incidence Younger	Random Assign Younger	Wtd Rand Assign Younger	True Total # Projects Older	Least Fill Older	Low Incidence Older	Random Assign Older	Wtd Rand Assign Older
Added a Bathroom	$17,887	$13,972	$13,831	**$6,307**	$11,367	$19,721	$22,676	$23,796	$17,704	$18,500
Replaced Siding	$11,735	**$6,733**	$9,524	$16,596	$15,498	$15,498	$13,541	$12,946	$16,883	$12,057
Added a Room	$8,300	**$1,686**	**$1,580**	**$2,956**	$5,725	$9,112	**$1,831**	**$1,791**	**$3,335**	**$2,272**
Rewired Electric	$7,495	$5,901	**$3,401**	*$4,832*	$5,596	$10,082	$7,903	*$6,481*	$7,223	$8,777
Remodeled Room	$17,898	$19,591	$14,469	$12,768	$15,131	$21,388	$15,924	**$13,243**	**$13,665**	$15,805
Replaced Roof	$11,271	$2,881	$2,779	$7,396	$3,330	$7,156	$3,072	$3,576	$4,974	$4,215
Built a Deck/Patio	$12,873	**$4,821**	**$4,560**	*$7,237*	**$5,186**	$9,763	**$5,894**	**$5,528**	$10,096	$7,104
Painted Exterior	$7,590	**$2,078**	**$1,393**	**$2,821**	*$4,043*	$6,467	**$2,369**	**$2,080**	**$2,558**	**$3,916**
Replaced Windows	$7,994	**$2,620**	**$2,341**	**$2,699**	**$2,384**	$5,642	**$3,003**	**$2,797**	*$3,966*	**$2,984**
Remodeled Kitchen	$10,587	**$2,929**	**$2,751**	**$2,944**	**$3,735**	$8,905	**$4,204**	**$3,807**	*$5,862*	**$5,558**
Remodeled Bath	$5,943	**$857**	**$1,673**	**$2,456**	**$2,929**	$5,182	**$1,353**	**$1,845**	**$2,380**	$3,990
Replaced Floor	$8,371	**$3,412**	**$3,871**	$6,563	**$2,813**	$9,308	**$3,566**	**$3,344**	**$4,006**	**$4,070**
Plumbing Work	**$7,287**	**$184**	**$180**	**$2,136**	**$1,035**	**$5,086**	**$239**	**$266**	**$2,827**	**$1,262**
Landscape Work	**$3,390**	**$1,146**	**$944**	**$2,079**	**$2,042**	**$4,298**	**$780**	**$772**	**$1,734**	**$1,410**
Painted Interior	**$5,859**	**$463**	**$239**	**$2,258**	**$1,910**	**$4,386**	**$598**	**$405**	**$2,923**	**$2,359**

Bolded is Significantly Different from True Spend at 95% Confidence Level
***Italics* is Significantly Different from True Spend at 90% Confidence Level**

Table H: Robustness Check: Income by Total Dollar Amount Spent

APPENDIX TABLE H: AVERAGE NUMBER OF PROJECTS IN TOTAL VERSUS THE ASSIGNMENT METHODOLOGIES - INCOME

Project	True Total # Projects Lower	Least Fill Lower	Low Incidence Lower	Random Assign Lower	Wtd Rand Assign Lower	True Total #4 Projects Higher	Least Fill Higher	Low Incidence Higher	Random Assign Higher	Wtd Rand Assign Higher
Added a Bathroom	$17,597	*$8,146*	$10,312	**$3,707**	$10,947	$20,982	$21,613	$21,070	$17,362	$19,949
Replaced Siding	$9,469	$10,416	$9,976	$9,661	$7,773	$18,887	$13,905	$16,424	$21,974	$19,391
Added a Room	$6,890	**$1,491**	**$1,355**	$5,114	$5,101	$9,339	**$2,685**	**$2,439**	**$4,673**	**$4,897**
Rewired Electric	$8,381	**$4,277**	**$2,936**	**$4,297**	**$4,163**	$9,730	$8,886	$7,324	$7,985	$12,088
Remodeled Room	$21,718	*$13,749*	**$10,252**	**$10,145**	**$8,815**	$18,589	$21,932	$18,264	$17,880	$22,263
Replaced Roof	$6,649	**$2,053**	**$2,648**	**$3,843**	**$3,326**	$11,603	**$3,415**	**$3,452**	$7,496	**$3,595**
Built a Deck/Patio	$9,213	**$4,322**	**$3,949**	$9,370	**$5,723**	$12,810	**$7,146**	**$6,989**	$9,083	**$6,680**
Painted Exterior	$7,755	**$1,742**	**$1,383**	**$1,947**	**$2,689**	$7,050	**$3,134**	**$2,552**	**$3,725**	$5,463
Replaced Windows	$6,743	**$1,560**	**$1,331**	**$1,668**	**$1,463**	$6,495	*$4,184*	**$4,012**	$4,904	**$3,983**
Remodeled Kitchen	$8,231	**$3,721**	**$3,578**	**$3,639**	**$3,497**	$11,006	**$4,197**	**$4,008**	$7,023	**$6,284**
Remodeled Bath	$4,616	**$1,073**	**$1,248**	**$1,581**	**$1,626**	$6,176	**$1,478**	**$2,005**	**$3,054**	$6,341
Replaced Floor	$6,778	**$2,151**	**$2,200**	**$2,332**	**$3,021**	$9,636	**$4,377**	**$4,701**	**$5,549**	**$4,163**
Plumbing Work	$5,335	**$193**	**$218**	**$2,657**	**$1,096**	$6,788	**$251**	**$257**	**$1,977**	**$1,446**
Landscape Work	$4,524	**$584**	**$605**	**$1,264**	**$1,093**	$4,458	**$1,144**	**$1,090**	**$2,453**	**$2,272**
Painted Interior	$4,945	**$375**	**$258**	**$1,284**	**$2,398**	$4,869	**$690**	**$444**	$4,127	**$2,272**

Bolded is Significantly Different from True Spend at 95% Confidence Level
***Italics* is Significantly Different from True Spend at 90% Confidence Level**

Table I: Sample Size by Methodology with Incidence Weighted Scheme Added

			Lowest	Random	Incidence Weighted
		Least-Filled	Incidence	Assignment	Random Assignment
	Incidence	Sample Size	Sample Size	Sample Size	Sample Size
Added a Bathroom	5.8%	156	160	155	160
Replaced Siding	6.7%	163	164	170	176
Added a Room	8.7%	156	163	214	155
Rewired Electric	11.0%	157	175	222	204
Remodeled Room	11.2%	163	172	222	185
Replaced Roof	11.5%	167	178	283	239
Built a Deck/Patio	14.9%	177	232	337	275
Painted Exterior	16.3%	167	227	364	285
Replaced Windows	16.7%	178	177	344	283
Remodeled Kitchen	17.1%	183	158	282	253
Remodeled Bath	18.8%	163	172	388	332
Replaced Floor	22.0%	184	155	351	337
Plumbing Work	30.0%	155	193	572	403
Landscape Work	39.0%	185	239	910	578
Painted Interior	45.1%	326	208	956	611
		2680	**2773**	**5770**	**4476**

Table J: Error in Total Amount Spent by Methodology with Incidence Weighted Scheme

								Incidence	Incidence
				Low	Low			Weighted	Weighted
	TRUE	Least Fill	Least Fill	Incidence	Incidence	Random	Random	Random	Random
Project/CELL	Avg Spend	Avg Spend	Spend Diff	Avg Spend	Spend Diff	Avg Spend	Spend Diff	Avg Spend	Spend Diff
Added a Bathroom	$18,527	$18,182	2%	$18,527	0%	$11,791	36%	$14,911	20%
Replaced Siding	$14,770	$11,829	20%	$12,573	15%	$15,981	+8%	$13,625	8%
Added a Room	$19,867	$17,022	14%	**$13,539***	32%	**$12,956***	35%	$14,172	29%
Rewired Electric	$8,390	**$1,730***	79%	**$1,662***	80%	**$4,750***	43%	**$4,531***	46%
Remodeled Room	$9,272	**$3,027***	67%	**$3,260***	65%	$6,003	35%	**$3,901***	58%
Replaced Roof	$10,563	**$5,636***	47%	**$5,352***	49%	$8,954	15%	**$6,242***	41%
Built a Deck/Patio	$8,920	**$3,693***	59%	**$3,695***	59%	**$5,069***	43%	**$3,881***	56%
Painted Exterior	$5,310	**$1,226***	77%	**$1,750***	67%	**$2,406***	55%	$3,350	37%
Replaced Windows	$9,721	**3,775***	61%	**$3,493***	64%	**$4,908***	50%	**$4,657***	52%
Remodeled Kitchen	$8,745	$6,840	22%	**$5,169***	41%	$6,350	27%	$8,193	6%
Remodeled Bath	$6,480	**$2,887***	55%	**$2,652***	59%	**$3,506***	46%	**$2,857***	56%
Replaced Floor	$7,112	**$2,262***	68%	**$1,849***	74%	**$2,643***	63%	**$3,647***	49%
Plumbing Work	$5,746	**$235***	96%	**$251***	96%	**$2,622***	54%	**$1,409***	75%
Landscape Work	$4,232	**$858***	80%	**$801**	81%	**$1,801***	57%	**$1,657***	61%
Painted Interior	$5,025	**$524***	90%	**$329***	93%	**$2,512***	50%	**$2,068***	59%
			56%		**58%**		**44%**		**43%**
	* Significantly Different from True Spend at 95% Confidence Level								

Table K: Error in Total Amount Spent by Method with Post Hoc Weighting on # of Projects

				WTD				Weighted	
	TRUE	Least Filled		Least Filled		Random		Random	
Project/CELL	Avg Spend	Avg Spend	Error	Avg Spend	Error	Avg Spend	Error	Avg Spend	Error
Added a Bathroom	$18,527	$18,182	2%	$24,133	30%	**$11,790***	36%	$15,733	15%
Replaced Siding	$14,770	$11,829	20%	$13,170	11%	$15,981	8%	**$24,336***	65%
Added a Room	$19,867	$17,022	14%	$21,575	9%	**$12,956***	35%	**$9,662***	51%
Rewired Electric	$8,390	**$1,730***	79%	**$2,088***	75%	**$4,750***	43%	$5,920	29%
Remodeled Room	$9,272	**$3,027***	67%	**$3,377***	64%	**$6,003***	35%	$10,564	14%
Replaced Roof	$10,563	**$5,636***	47%	**$5,420***	49%	$8,954	15%	$8,932	15%
Built a Deck/Patio	$8,920	**$3,693***	59%	**$3,756***	58%	**$5,069***	43%	$7,751	13%
Painted Exterior	$5,310	**$1,226***	77%	**$1,137***	79%	**$2,406***	55%	**$3,180***	40%
Replaced Windows	$9,721	**$3,775***	61%	**$4,157***	57%	**$4,908***	50%	$9,524	2%
Remodeled Kitchen	$8,745	$6,840	22%	$7,053	19%	$6,350	27%	$6,481	26%
Remodeled Bath	$6,480	**$2,887***	55%	**$2,644***	59%	**$3,506***	46%	**$4,085***	37%
Replaced Floor	$7,112	**$2,262***	68%	**$2,574***	64%	**$2,643***	63%	**$3,783***	47%
Plumbing Work	$5,746	**$235***	96%	**$244***	96%	**$2,622***	54%	$6,287	9%
Landscape Work	$4,232	**$858***	80%	**$874***	79%	**$1,801***	57%	**$2,748***	35%
Painted Interior	$5,025	$524	90%	$893	82%	$2,512	50%	$6,483	29%
			56%		**55%**		**41%**		**29%**
	*Significantly Different from True Spend at 95% Confidence Level								