

What Parcel Tax Records Tell Us about Homeownership Measurement in Surveys

Shiyu Zhang
Institute for Social Research
University of Michigan

James Wagner
Ford School of Public Policy
University of Michigan

Elisabeth R. Gerber
Institute for Social Research
Ford School of Public Policy
University of Michigan

Jeffrey D. Morenoff
Institute for Social Research
Ford School of Public Policy
Department of Sociology
University of Michigan

This research aims to understand the measurement error in self-reported homeownership data collected by surveys. The analysis focuses on Detroit as a case study. We use legal ownership status in administrative records (the city of Detroit parcel tax records) as the benchmark to validate self-reported ownership status collected from a survey (the Detroit Metro Area Communities Study). We compare data from two question formats, which measure ownership at the household level and at the individual level, respectively. We also study the associations between sociodemographic characteristics and measurement errors in the self-reported ownership. The results suggest that 1) respondents do not always interpret the ownership questions as was intended, 2) the reported ownership status is sensitive to question formats, 3) the risk of measurement error appears to be heterogeneous in the population. The results challenge the assumption that homeownership is a standard fact, the reporting of which is not impacted by how it is measured. The findings are useful for understanding discrepancies across survey results and for advising how to craft homeownership questions in surveys.

Keywords: Survey data, administrative data, survey question wording, measurement error, homeownership

1 Introduction

For many Americans, owning a home is an important component of personal wealth. A great deal of substantive research has been conducted on homeownership (Cortes, Herbert, Wilson, & Clay, 2007; Dietz & Haurin, 2003; Herbert & Belsky, 2008; Wainer & Zabel, 2020, e.g.). In response to the needs of data users, many surveys include a question or questions about homeownership.

Despite the many studies that use homeownership data and the many surveys that produce homeownership data, research on the measurement of homeownership and the resulting data quality is surprisingly scarce. Guidance on effective question wording is also limited. In current practice, a variety of question formats are used to ascertain homeownership across surveys. The lack of attention to measurement issues

would seem to imply that homeownership is simply an objective fact, the reporting of which is not impacted by the form of the question.

However, the difficulty of measuring homeownership may depend on household composition and who is sampled within the household. The question might be straightforward for single-person households, but it can be difficult and prone to error when multiple people live in the same residence and one person responds on behalf of the household (Kilic & Moylan, 2016). The few studies that evaluated measurement errors in homeownership have shown that there can be inconsistencies in people's responses. Self-reported ownership status changes when the data are compared between interview and reinterview (American Housing Survey), between two surveys (American Housing Survey vs. the Decennial Census), and across survey waves (Health and Retirement Study) (Chakrabarty, 1996; Venti, 2011). The previous findings call for more investigation to further understand the problem and guide practice (Chakrabarty, 1996; Venti, 2011).

In this article, we take a new approach and use administra-

Contact information: Shiyu Zhang, Institute for Social Research, University of Michigan, Ann Arbor MI 48104, USA, email: zsy@umich.edu

tive records as the benchmark to validate survey responses. We compare two question formats that were worded with different usage purposes in mind: one captures homeownership at the household level and the other at the individual level. The results provide strong evidence challenging the assumption that homeownership is a fact that can be reliably reported by respondents regardless of the question wording. Instead, the results show that homeownership can be ambiguous and self-reported status is susceptible to changes in the provided response categories. Based on the findings, we provide practical recommendations on how to craft homeownership questions to minimize unreliable responses.

1.1 Literature Review on Homeownership Measurements

When respondents cannot easily map their situations to survey questions and categories, measurement errors tend to occur (Tourangeau, 2003). There are a few reasons why measurement errors may exist in the self-reported homeownership data.

First, ownership can be measured at either the household level or the individual level (Bennett, 2013), but the distinction between the two levels is not always clear in survey questions. Household-level questions differentiate whether the household as a whole owns or rents the residence. These data can be useful for, for example, understanding the relationship between homeownership and health for reasons like better living conditions, more stability, and greater community integration (Finnigan, 2014; Lindblad & Quercia, 2015; Manturuk, 2012). In contrast, individual-level questions capture the legal ownership of specific household members. These data can be useful for, for example, understanding within-household differentials (e.g., gender differentials) in wealth (Bennett, 2013; Deere & Doss, 2006; Kilic & Moylan, 2016; Shiffer-Sebba & Behrman, 2020).

Most surveys include household-level questions on homeownership. But sometimes the wording is ambiguous and risks respondents mixing up the levels. For example, Behavioral Risk Factor Surveillance System (2019) and California Health Interview Survey (2020) word the question as “do you own or rent your home” (see Table 1). Because the subject of the question “you” is not clearly defined, the question could collect either the household-level or the individual-level responses from some respondents (e.g., adult children renting a room from the parents’ house).

Second, specific to measuring household-level ownership, the data may be subject to measurement errors when respondents do not have sufficient knowledge to answer the question or when the response categories come across as confusing (Groves et al., 2009). Depending on the household composition, the selected respondent may not know the arrangement of the place where they live in detail. This becomes a problem especially when the question distinguishes several

types of ownership in the response categories (e.g., owning with a mortgage, owning with a land contract, and owning free and clear). For example, as we will discuss using our data, the ownership question seems difficult when the respondent is from a household where people are related but do not form a nuclear family (e.g., grandparents-grandchildren).

Third, of the household-level and individual-level measures of homeownership, the bigger challenge lies in the latter. Surveys typically intend to capture the formal ownership of individuals. However, respondents are likely to respond based on how they perceive the situation. For instance, as noted in a cognitive interview report for the Census 2006 Test, a wife factored in considerations like who is the “breadwinner” and who has a higher income when reporting the owner of the house (Childs, Carter III, Norris, Hanaoka, & Schwede, 2006). In addition, household members may feel they are the owner because they have the so-called economic ownership of their residence (i.e., deciding how the money is used if the property is sold) or can make decisions about the property (e.g., rent/sell) (Kilic & Moylan, 2016; also see in similar distinction in saving and debt in Kan & Laurie, 2014). These factors complicate the situation for how surveys can effectively capture ownership at the individual level.

Furthermore, when measuring individual-level ownership, survey mode influences how accurate and detailed the data can be. The American Housing Survey (AHS) is one example of a survey that measures housing arrangements within households. AHS compiles household rosters, establishes relationships between household members, and then explicitly asks in whose name is this home. This question format, however, tends to be restricted to interviewer-administered surveys. Interviewers assist and motivate respondents because enumerating and asking questions about every person in the household is intrusive and burdensome. Self-administered web/mail surveys are much less likely to succeed with a roster-based approach (except for high-profile and mandatory surveys like the Decennial Census and ACS). Also, unless housing is a central topic, most surveys do not invest such extensive effort into measuring homeownership. Thus, there is a need to explore the validity of using simplified question items to collect individual-level ownership data in self-administered surveys.

Table 1 includes examples of homeownership questions from several important surveys. In the right column, we remark on the level of the question. These example questions vary in how they are worded and what response categories are provided. The noticeable variation calls for investigations in whether the quality of survey responses are susceptible to changes in how the question is asked.

1.2 Overview of the Current Study

To investigate the quality of survey responses, we use legal ownership from administrative data on parcels as the

Table 1

Examples of questions measuring homeownership in different surveys (in the most recent questionnaire)

Survey	Question item(s)
<i>Household-level</i>	
The 2020 U.S. Census questionnaire American Community Survey (2021)	Is this house, apartment, or mobile home - Owned by you or someone in this household with a mortgage or loan? Include home equity loans. - Owned by you or someone in this household free and clear (without a mortgage or loan)? - Rented? - Occupied without payment of rent?
National Household Education Survey (2016)	Is this house - Owned or being bought by someone in this household - Rented by someone in this household - Occupied by some other arrangement
General Social Survey (2018)	Do you/Does your family own your home/apartment, pay rent, or what? - Own or is buying - Don't know - Pays rent - Refused - Other
Panel Study of Income (2019)	Do you [or someone else in your family living there] own the apartment/mobile home/home, do you pay rent, or what? - Owns or is buying (fully or jointly) - Pays rent - Neither owns nor rents (PSID includes follow-up questions to further distinguish situations such as a mortgage, a land contract, a home equity loans, foreclosure, public housing etc.)
<i>Can be ambiguous for some respondents because the subject "you" is undefined</i>	
American National Election Study (2020)	Do you pay rent for your home, make monthly mortgage payments for your home, own your home outright with no payment due, or have some other living arrangement? - Pay rent - Some other arrangements - Pay mortgage - Don't know - Own home with no payments due - Refused
Behavioral Risk Factor Surveillance System (2019)	Do you own or rent your home? - Own - Don't know/not sure - Rent - Refused - Other arrangement
California Health Interview Survey (2020)	Do you own or rent your home? - Own - Rent - Other

Continues on next page

Continued from last page

Survey	Question item(s)
<i>Can be ambiguous for some respondents because the subject "you" is undefined</i>	
National Postsecondary Student Aid Study (2015-16)	Do you own a home or pay a mortgage on a home? - Yes - No
<i>Household-level question followed by individual-level probing</i>	
American Housing Survey (2019)	Is this [house type] - Owned or being bought by someone in your household? - Rented? - Occupied without payment of rent? In whose name is this home? - Open-ended. Enter up to five names

benchmark to validate the self-reported homeownership data. Parcels are quantities of land identified for taxation purposes. (Condominium units correspond to separate parcels.) Legal ownership can be established from the parcel data because legal owners of properties are listed as the taxpayers.

Survey researchers have used administrative data as a benchmark to validate various survey variables including earnings, pensions, taxes, assets, labor force status, insurance coverage, education level, marital status and housing values (Abowd & Stinson, 2013; Bound, Brown, & Mathiowetz, 2001; Davern et al., 2008; Groen, 2012; Kapteyn & Ypma, 2007; Kreuter, Müller, & Trappmann, 2010; Olson, 2006; Voorheis, 2020). Depending on the types of records, administrative records are not without errors. Records that depend on self reports (e.g., income tax) are subject to more measurement errors than those that do not (e.g., divorce status). Also, the purpose for which the administrative data are created may not align with the research applications (Groen, 2012; Lyberg & Kasprzyk, 1991). Nevertheless, cross validating data from the administrative and survey sources has been demonstrated to be useful in understanding errors in survey responses. Yet, to the best of our knowledge, no studies have used administrative records to discern measurement errors in homeownership indicators.

Using parcel data, we evaluate two question formats for measuring homeownership. One format assesses household-level ownership, with the question worded similarly to the one in the Behavioral Risk Factor Surveillance System (2019). The other format is a two-part question set. The first part is modified from the Decennial Census and ACS, asking whether the respondent's household owns the residence. The second part follows up on individual ownership by differentiating between the respondent themselves and the other members of the household. The results generate some practical insights on how to word the question and response categories.

Next, because people of different characteristics might be

more or less likely to misreport and misinterpret the questions, we use sociodemographic features to predict the probability of misreporting. The results provide some evidence that the risk of measurement error might not be homogeneous in the population.

Furthermore, this research shows that parcel tax records from administrative sources can be a good supplement to survey data. Survey data are still irreplaceable for the purpose of differentiating owner- versus renter-occupied households. However, once respondents report that their household owns the residence, parcel data can enrich the survey data by providing information about differentials within households.

2 Methods

This research focuses on Detroit as a case study. The survey data come from representative samples of Detroit respondents in 2019 and 2020; the corresponding parcel tax records are from the parcel database of the city of Detroit in 2020.

2.1 Survey Data from DMACS

The Detroit Metro Area Communities Study (DMACS, 2020) is an ongoing panel study of representative samples of Detroit residents since 2016. The panelists are recruited with address-based probability samples. Several modes have been used in the past recruitments, including mail-invited web interview (the main mode), canvasser-invited web interview, interviewer-administered telephone interview, and mail-invited pen-and-paper interview. DMACS provides weights that calibrate the sample to the Detroit population. But the current analysis is unweighted because we aim to identify types of measurement errors instead of making population inferences. We analyze DMACS (2020) data from wave 5 (Spring, 2019), wave 6 (Fall, 2019), wave 8 (Spring, 2020), and wave 11 (Fall, 2020).

Questions about homeownership were worded differently across DMACS survey waves. In waves 5, 6, and 8, the

homeownership question was worded as follows (i.e., *question format I*):

Do you own your home, rent it, or have some other arrangement?

- Own (or in the process of being bought)
- Rent
- Have a land contract (aka rent-to-own, lease to purchase, contract for deed)
- Some other arrangement

In wave 11, the question was changed to a two-part question set (i.e., *question format II*). The first part used the standard homeownership question in the Decennial Census and the American Community Survey with one additional response option “Other”:

Is your current residence...

- Owned by you or someone in this household with a mortgage or loan (which could be a home equity loan)
- Owned by you or someone in this household free and clear (without a mortgage or loan)
- Occupied without payment of rent
- Rented
- Other

If the respondents report ownership, the second part further asked:

Do you own the home where you are living or does someone else in your household own it?

- I own it
- Someone else in this household owns it
- I and someone else own it together

The current analyses use two analytical samples, *sample I* and *sample II*, corresponding to the respondents that reported ownership status through question format I and II, respectively.

- The analytical *sample I* consists of 1,947 panelists who provided their names, had a valid address within the boundary of Detroit, and reported their homeownership status in wave 5, 6 or 8.

- The analytical *sample II* consists of 1,005 panelists who provided their names, had a valid address within the boundary of Detroit, and reported their homeownership status in wave 11. There is substantial overlap between sample I and II. Out of the 1,005 wave-11 respondents, 999 of them participated at least once in wave 5, 6, or 8.

2.2 Parcel Records

The parcel tax records are from the City of Detroit Open Data Portal (version released on July 30, 2020). The dataset

contains 382,097 records of all Detroit parcels. On each parcel, available information includes taxpayers’ names, property purpose (e.g., residential or commercial), taxable value, and more. We used the addresses and taxpayer names associated with parcels.

For each DMACS panelist address, we tried to find its parcel record based on the street number and name. Not all DMACS addresses were found in the parcel database: 13% ($n = 248$) of the DMACS addresses in sample I and 12% ($n = 122$) of the DMACS addresses in sample II had no corresponding records in the parcel database. This is because postal addresses (i.e., DMACS panelist addresses) are not always the same as the addresses in the parcel database. To be specific, most addresses correspond to unique parcels, including addresses in condominium buildings. Exceptions occur mostly when multi-unit structures are rental or co-op communities. For example, a series of rental units spanning multiple buildings might be listed under a single address in the parcel system. Such rental and co-op structures are often owned by establishments. We compared demographic characteristics of respondents whose addresses were found in the parcel database with those whose addresses were not. For both sample I and II, the two groups of respondents did not significantly differ in gender, age, education, or race and ethnicity. However, the panelists whose addresses were not found in the parcel database had lower income than those whose addresses were found (60–62% vs. 71% annual household income lower than \$35,000). This difference in income is not surprising because the parcel-/postal-address mismatch problem discussed above is most likely to occur among rental units. Indeed, most of these panelists whose addresses were not found self-reported as renters, at a much higher proportion (74% and 77% in sample I and II) than the overall sample (42% and 39%).

Excluding the panelists whose addresses were not found in the parcel database, the following analysis focuses on the 1698 panelists in sample I and 882 panelists in sample II whose addresses were successfully found in the parcel records.

2.3 Validation Analysis – Comparing Panelists’ Names with Taxpayers’ Names

Panelists’ self-reported homeownership is validated by comparing panelists’ names with taxpayers’ names listed in the parcel tax records. The panelist names collected by DMACS are likely to be accurate because the DMACS questionnaires stated that the names will be used to process incentive payment in the form of a check or a prepaid debit card.

We separate the comparisons by panelists’ first and last names because matching first name and last name indicates different ownership and living situation.

- If panelists’ *first and last names* are both matched to the

taxpayers' names listed in the parcel database (i.e., matching full name), the administrative records suggest legal homeownership.

- If only the *first name* is matched to the parcel database, the panelists are still likely to be the legal homeowner if they also self-reported as homeowners. DMACS might have recorded different last names than property registration because of people changing their names due to marriage or other reasons.

- If only the *last name* is matched to the parcel database, the panelists are probably not the legal homeowner. Instead, the legal owners might be someone closely related to the panelists, as indicated by the same last names.

Alternatively, it could be because the panelists reported a nickname to DMACS (e.g., Rob instead of Robert). However, because we consider the names that DMACS collected to be highly accurate (for the reason of processing incentives), we treat different first names in a literal form as indicating different people. This approach is expected to produce a more conservative estimate.

- If neither last name nor first name is matched to the parcel database, the panelists are probably not the legal homeowner. But we do not have the information to speculate the relationship between the panelists and the legal owners of the residences.

In the Results section, the validation results are reported separately for panelists' who self-reported different ownership status (e.g., owning, renting, and others).

One other issue about taxpayers' names in the parcel database is that taxpayers can be establishments or other legal entities rather than persons. We identified these entities through looking for keywords (e.g., llc, inc, corp, bank, invest, housing, etc.) in the listed taxpayer names.

2.4 Consistency Analysis – Comparing Responses through Two Question Formats

Recall that there is substantial overlap between analytical sample I and sample II because many panelists who responded to wave 11 also responded at least once in wave 5, 6, and/or 8. Using the overlapped sub-sample, we were able to compare responses obtained through the two question formats. This comparison informs how consistent and robust the ownership data were with regard to the different measurement formats.

This within-person comparison is only valid for panelists who stayed in the same address because we can reasonably assume that their living arrangements stayed the same and thus expect consistent responses. In contrast, any inconsistencies reported by movers could be due to actual changes in living status. Therefore, we excluded a small number of panelists who changed their addresses between waves, including 16 homeowners and 72 people who rented or had other

arrangements, which leaves 905 panelists in the overlapped sub-sample.

2.5 Multivariate Analysis

The analyses above identify potential measurement errors in survey responses. One question that follows is whether the risk of measurement errors is heterogeneous in the population. If the errors are heterogeneous, namely, certain people are more or less likely to misreport, then the measurement errors could introduce systematic biases into survey estimates.

We use sociodemographic characteristics to model the probability of observing discrepant and inconsistent ownership data. The list of predictors includes age (continuous variable; range=19–93; mean=50), gender (binary variable: 1=male; 0=female), race and ethnicity (categorical variables: non-Hispanic white, non-Hispanic black, non-Hispanic other, Hispanic), education level (categorical variable: less than high school, high school, some college, college and above), income (categorical variable: < \$10,000, \$10,000–29,999, \$30,000–49,999, \$50,000–99,000, \$100,000 or more), marital status (binary variable: 1= married, 0= otherwise), and living alone (binary variable: 1= alone; 0=not alone).

We define two outcome variables to indicate potential measurement errors. The first outcome variable is based on the comparisons between parcel tax records and ownership data collected by question *format II*. We focus only on panelists who self-reported as owners (“by you or someone in the household”) because these responses can be the most meaningfully validated by parcel records. A binary variable is constructed, taking the value of 1 if the self-reported status is consistent with the parcel records and 0 otherwise. By “consistent,” we refer to situations where 1) panelists reported to be personal owners and their first names indeed appeared in the parcel tax records or 2) panelists reported that someone else in the households owns the residence and their first names did not appear in the records. (Recall that if people report residence ownership and their first names appear in the record, we consider it as evidence supporting their legal ownership.)

The second outcome variable is based on comparisons between the ownership data collected from the two question formats. A binary variable is constructed, taking the value of 1 if responses from the two question formats are reconcilable and 0 otherwise. By “reconcilable,” we refer to situations where the panelists 1) reported to be owners in both formats, 2) reported to be renters in both formats, 3) reported having “some other arrangement” in question *format I* and “free”/“other” in question *format II*. The small number of panelists who reported having a land contract in *format I* are excluded from this analysis because it is not clear how this response should be mapped to *format II*.

3 Results

3.1 Validation Based on Question Format I

Panelists in the analytical *sample I* were asked the question “Do you own your home, rent it, or have some other arrangement?” (question *format I*). Four response options were provided: own, rent, have a land contract, and some other arrangement. The validation results are separately reported for self-reported owners, renters, and the rest.

Own. A total of 883 panelists (52% of the analytical sample I) reported ownership. These self-reported owners should have their names listed as taxpayers.

Our analysis first checked whether the taxpayers are person’s names (as opposed to establishments or other entities). Person names were listed for 97% of the self-reported owners; the remaining 3% have other entities as their taxpayer. These entities include organizations such as the land banks and the city planning and development department, suggesting that some of the self-reported owners might be in the process of obtaining homeownership with some special arrangements.

Next, for the 97% self-reported owners whose taxpayers are not entities, we cross-compared panelists’ first and last names with the taxpayer names. The matching results are reported in the left panel in Table 2. If we consider matching full name (first and last name) and matching first name both as evidence of legal ownership, then the results confirmed 64% (=60%+4%) of the self-reported homeowner status. For the remaining 36% of the self-reported owners, their first names did not show up in the property records. Such a result is expected because the question item was ambiguous – If the “you” and “your” in the question stem (i.e., “do you own your home. . .”) were interpreted as plural, then the respondents might check “own” if someone else in the household legally owns the residence.

Rent. Another 626 panelists (37% of the analytical sample I) reported rentership. These self-reported renters should *not* have their names listed as taxpayers. The taxpayers of their residences can be either entities or some other people.

Our analysis first checked whether the taxpayers are persons or entities. As expected, a large portion of self-reported renters (45%) lived in properties owned by entities.

Next, for the remaining self-reported renters whose taxpayers are not establishments, we again compared the panelists’ first and last names with taxpayer names. The matching results are reported in the middle panel of Table 2. As it should be, most renters’ names are not matched to taxpayers in the parcel records (86%). However, still seven renters have both their first and last names in parcel tax records; it is not clear why they did not report as the owner.

3) Land contract or other. The remaining 184 panelists reported having a land contract (26) or some other arrangements (158). There is no clear expectation on how these

people’s names should appear on the parcel records.

We first checked whether the taxpayers are persons or entities. About 13% of these respondents have entities listed as their residences’ taxpayers. One interesting finding is that, among these entities, we once again see organizations like land banks. Recall that this kind of entity is also listed as taxpayer for a small number of self-reported owners. It could be that ownership is blurry for people who are purchasing their residence through some special project, and thus people in these situations interpreted the question and chose the response options differently.

Next, for the remaining respondents whose taxpayers are not entities, we cross-compared their first and last names with taxpayer names. About 13% of these respondents have both their first and last names in taxpayer records. Another 37% of these respondents have their last names matched with the names of the taxpayers, suggesting that they may be related to the legal owners. Recall that a substantial proportion of the self-reported owners only have their last names match with the taxpayers. These last-name-matching respondents who reported having other arrangements may have similar living situations to those who reported owning. The different responses might depend on how the respondents perceive and feel about the situation.

3.2 Validation Based on Question Format II

Panelists in the analytical *sample II* reported their homeownership through two questions (question *format II*). The question set first asks about homeownership by “you or someone else in this household”. If respondents reported ownership, then a second question follows to further differentiate between “you” and “someone else in this household”. From these two questions, we separated four situations: the current residence 1) owned by “I” or “I and someone else”, 2) owned by “someone else” in the household, 3) rented, 4) occupied for free or under other arrangement. The validation results are separately reported for respondents in each situation.

Owned by “I” or “I and someone else”. A total of 427 panelists (48% of the analytical *sample II*) reported that “I” (337) or “I and someone else” (90) own the current residence. As these respondents self reported as personal owners of their residences, their names should be listed as one of taxpayers.

Parallel to the analysis above, we first checked whether the taxpayers are entities or persons. The vast majority of taxpayers are person names; however, still seven of them have entities like land banks as their listed taxpayers.

Next, for the majority of personal owners whose taxpayers are not entities, we cross-compared their first and last names with taxpayer names. The matching results are reported in Table 3. If we consider matching full name (first and last name) and matching first name both as evidence of legal

Table 2
Validation results of question format I. For the panelists whose taxpayers are not establishments, the numbers of them (and percentage) whose first and last names are in the parcel records

		Own				Rent				Land contract or Other			
		Match		Not		Last Name Match?				Match		Not	
		N	%	N	%	N	%	N	%	N	%	N	%
First Name	Match	511	60	34	4	7	2	2	1	20	13	3	2
Match?	Not	175	20	138	16	41	12	295	86	59	37	76	48

ownership, then the results confirmed 71% (=67%+4%) of the self-reported homeowner status. For the remaining 29% of panelists who identified as personal owners, their names did not show up in the property records. This result is surprising because the question items explicitly differentiate between ownership by “I” and by “someone else in this household”. The fact that still 29% of the self-identified owners were not listed in parcel tax records suggests the respondents might not interpret the question as was intended.

Owned by “someone else”. Another 110 panelists (12% of the analytical *sample II*) reported that “someone else” in the household owned their current residence. Different from the personal owners in the above category, these respondents should not have their names listed as taxpayers.

Such expectation is met for the most part: The vast majority of respondents (95%) have person names listed in records, and only six of them have entities as taxpayers. For the respondents whose taxpayers are not entities, most of their names cannot be matched to the property records, including 49% whose first names are not matched and 47% whose first and last names are not matched.

Rent. Another 300 panelists (34% of analytical sample II) reported rentership. Since question *format I* and *format II* mostly differed in how homeownership was measured, renters should provide the same response regardless of the question formats. Indeed, we observe almost identical response patterns: Based on format II, 46% self-reported renters lived in properties owned by establishments, compared to 45% based on format I. The validation results in percentages reported in the “Rent” column in Table 3 are also almost identical to the percentages in Table 2.

Occupied without payment of rent and Other. A small number of panelists checked the response options “occupied without payment of rent” and “other”. The result patterns do not substantially differ between these two response options, and thus they are reported together.

Among the 45 respondents who stayed for free or had other arrangements, only three had entities as taxpayers. Next, the results of the comparison between panelist names and taxpayer names are reported in the last column of Table 3. It is noticeable that 29% of these respondents had their

full names listed in the parcel tax records; it is not clear why they did not identify themselves as the owner.

3.3 Consistency—Different Responses via the Two Question Formats

Based on the 905 panelists who responded to both question formats while staying in the same addresses, we compared the homeownership data collected through these two formats. Table 4 reports the frequency table cross-tabulating the two sets of responses. The responses were consistent and reliable for most panelists: 95% of the panelists who reported ownership and 87% of those who reported rentership through format I reported the same status through format II. However, the dispersion in the frequency table is not negligible.

One possible reason for the dispersion is measurement error, and another reason is changes in living arrangement within the same addresses. To evaluate the plausibility of changing arrangement, we performed sensitivity analysis by comparing format I data from wave 8 only (April-May 2020) and format II data from wave 11 (October 2020). Given the short time interval, it is much less likely that people genuinely changed their living arrangement. The result patterns stay the same as reported in Table 4.

This inconsistency echoes the reliability issue of homeownership measurement identified by previous research (Chakrabarty, 1996). Comparing to a previously reported 6% difference between the American Housing Survey data and the Decennial Census data under the same question item (Chakrabarty, 1996), here, the inconsistency between the two question formats is much more pronounced. The result suggests that people’s interpretation and report are susceptible to differences in question formats.

In addition, the less common categories such as land contract, free of rent, and other arrangement did not seem to have a consensual definition and were left for the respondents to interpret. For example, the second-last row of Table 4 shows how cases who reported “some other arrangement” in format I significantly changed their responses in format II. Only eight out of 69 cases chose the “other” option again, suggesting that the data on the “other” category may not be reliable or meaningful.

Table 3

Validation results of question format II. For the panelists whose taxpayers are not establishments, the numbers of them (and percentage) whose first and last names are in the parcel records

		Owned by "I" or "someone else and I"				Owned by "someone else"				Rent				Free or Other			
		Match		Not		Match		Not		Match		Not		Match		Not	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
First Name Match?	Match	284	67	16	4	4	4	1	1	3	2	1	1	12	29	1	2
	Not	68	16	52	12	51	49	49	47	17	11	140	87	15	36	14	33

Table 4

Frequency table that cross-tabulates responses through question format I and II

		Question format II					
		Owned by			Rent	Free	Other
Question format I		I or I and someone else	someone else				
	Own		403	45	6	11	7
	Rent		10	25	301	9	2
	Other arrangement		10	29	17	5	8
	Land contract		9	3	2	0	3

3.4 Multivariate Analysis

To understand whether the risk of measurement error is heterogeneous in the population, we used sociodemographic variables to predict the probability of observing discrepant and inconsistent responses. We constructed two outcome variables to indicate potential measurement errors, explained in detail in the Methods section. The first outcome is based on comparisons between parcel tax records and ownership data collected from question *format II*. The binary variable indicates whether the self-reported ownership status could be confirmed by the parcel records. The second outcome is based on comparisons between ownership data collected from the two question formats. The binary variable indicates whether the responses from the two formats were consistent. The results of the two logistic regression models are included in Table 5.

For outcome 1, none of the explanatory variables significantly predicts whether the self-reported ownership status could be confirmed by parcel tax records. Thus, even though the parcel validation suggests that some respondents might not have interpreted the questions as was intended, there is no evidence that the (mis-)interpretation was systematic with regard to the sociodemographic features.

For outcome 2, four significant and two marginally significant predictors emerge. Panelists who were older, were married, and lived alone were significantly more likely to give consistent responses. Compared to the lowest income

bracket, those with household income between \$50,000-99,999 were significantly more likely to respond consistently. In addition, non-Hispanic white panelists tended to give more consistent responses (marginally significant) than Hispanic and non-Hispanic Black panelists. One possible interpretation is that the living arrangement tends to be simpler for people who live alone and stabler for people who are older, have medium-high income, and are married. The simplicity and stability might facilitate consistent survey responses.

4 Discussion

We focus on Detroit as a case study and use parcel tax records as the benchmark to validate self-reported homeownership in survey responses. The response data were collected through two question formats. One was intended to capture household-level ownership; the other was intended to capture individual ownership. We validate self-reported homeownership by checking whether self-reported owners are listed in the parcel tax records for their reported residential addresses. Based on the validation results, we were able to infer how people interpreted the questions and to advise how homeownership can be differently measured to fit research purposes. The key findings are summarized below.

First, the self-reported homeownership status collected through question format I and format II are similar. Based on question format I, a substantial portion (36%) of self-

Table 5

Coefficients of logistic regression models predicting the probability of observing non-discrepant and consistent responses on homeownership

	Outcome 1 – No discrepancy between and parcel tax records (N=495) ^a				Outcome 2 – Consistent responses across question formats (N=845) ^b			
	Coefficient	Odds ratio	SE	p-value	Coefficient	Odds ratio	SE	p-value
Intercept	1.304		0.815	0.110	1.108		0.658	0.092
age	0.010	1.010	0.008	0.200	0.017*	1.017	0.007	0.014
male	0.059	1.061	0.240	0.806	-0.004	0.996	0.237	0.986
<i>race (non-Hispanic white=reference group)</i>								
black	-0.279	0.757	0.313	0.372	-0.827	0.438	0.428	0.054
Hispanic	0.324	1.383	0.576	0.573	-1.064	0.345	0.544	0.051
other	0.154	1.167	0.474	0.744	-0.488	0.614	0.525	0.353
<i>education (less than high school = reference group)</i>								
high school	-0.879	0.415	0.696	0.207	-0.040	0.961	0.446	0.929
some college	-0.505	0.604	0.673	0.453	0.157	1.170	0.435	0.718
college and above	-0.301	0.740	0.695	0.664	0.024	1.025	0.485	0.960
<i>income (<\$ 10,000 = reference group)</i>								
\$10,000–29,999	0.184	1.202	0.342	0.590	-0.011	0.989	0.257	0.967
\$30,000–49,999	0.344	1.411	0.355	0.333	0.215	1.240	0.305	0.481
\$50,000–99,999	0.369	1.446	0.374	0.325	0.944*	2.570	0.411	0.021
> \$100,000	-0.108	0.898	0.460	0.815	0.704	2.022	0.570	0.217
marry	-0.412	0.662	0.256	0.108	0.652*	1.920	0.283	0.021
alone	-0.242	0.785	0.256	0.345	0.451*	1.569	0.229	0.049

^a The model was fitted to a sample consisting of the cases under Table 3 columns “owned by I or someone else and I” and “owned by someone else”, except that some cases with missing values on the predictors were excluded.

^b The model was fitted to all cases in Table 4, except that some cases with missing values on the predictors were excluded.

* $p < 0.05$

reported owners did not have their names in parcel tax records. This result is not surprising because the question stem “do you own your home” did not clarify what “you” meant. If the respondents interpreted “you” as plural, then it is legitimate to report ownership if anyone in the household owned the residence. However, we are surprised that question format II resulted in similar response patterns, even though it explicitly differentiated between “I” and “someone else”. Still, for the respondents who identified themselves as personal owners (i.e., selected “I own it” or “I and someone else own it together”), 29% did not have their names in parcel tax records. The findings from format II indicate the complexity in measuring homeownership and living arrangement. People are likely to respond based on their perceived ownership, which can differ from the legal status indicated by names on the deeds (Bennett, 2013; Deere & Doss, 2006; Kan & Laurie, 2014; Kilic & Moylan, 2016).

Second and relatedly, although question format I and II both generated some survey responses that seem inconsistent with parcel tax records, the two-part format II was not completely unproductive. Format II prompted some self-reported

owners to clarify that someone else in their household own the residence. When respondents identified someone else as the owner, these responses are highly consistent with parcel records. Taken together, among the format II respondents who reported household ownership but did not have their own names on the deeds, roughly half of them claimed that they were personal owners (i.e., $\frac{68+52}{68+52+51+49} = 54.5\%$) and the other half clarified that someone else owned the residence (i.e., $\frac{51+49}{68+52+51+49} = 45.5\%$). These results suggest that some respondents might interpret format II as intended, but this intention was not clear to all respondents. Thus, if surveys need to capture individual-level ownership, it is worth explicitly defining what ownership is on the survey.

Third, the results suggest that adding additional categories (e.g., land contract, stay for free, and other arrangements) doesn’t necessarily make the question clearer. Because of the diversity in people’s financial and living situations, the distinctions are sometimes blurry between owning, renting, staying for free, and some other arrangements. For example, when we look into the 23 open-ended clarifying responses for the “Other” category, seven respondents wrote that they

were living with parents, children, or grandparents. This situation is equally valid to be reported as ownership by someone else in the household. Similarly, being “a paid member of [housing] cooperatives” may be alternatively interpreted as renting. Because response categories are supposed to be mutually exclusive, adding additional options changes the meanings of the standard options of “rent” and “own”. Such changes in meanings can be undesirable, especially if they are interpreted heterogeneously across people.

Fourth, when comparing the responses to the two question formats by panelists who stayed in the same addresses, the inconsistencies in responses were not negligible, as shown by the dispersion in Table 4. The inconsistencies show, on the one hand, that people’s responses are susceptible to small changes in response categories. On the other hand, responses might be subjected to measurement error like satisficing because some inconsistencies are hard to interpret – reporting “other arrangement” at one time and ownership by “I” at another time.

Fifth, the multivariate analyses provide some evidence that the error in measuring homeownership might be heterogenous in the population. The homeownership status was more consistently reported by people with stabler (older, married, and have medium-high income) and simpler (living alone) living arrangements. If ownership is more accurately measured in certain groups than in others, the heterogeneous error can potentially introduce biases in survey estimates. This finding again highlights the importance of carefully defining the survey question and providing unambiguous response categories.

There are at least three caveats in interpreting the current results. First, there are time differences between the survey responses and the parcel tax records. We try minimizing the impact of time by focusing on survey responses in 2019 and 2020 to be consistent with the parcel records in 2020. Still, the results cannot eliminate the possibility that some of the mismatches resulted from actual changes in living arrangements. Thus, as a sensitivity analysis, we repeated the validation with only survey wave 8 (April–May 2020) and wave 11 (October 2020), which are very close in time to the parcel records (July 2020). The result patterns and conclusions stay the same.

Second, when validating self-reported ownership with tax records, we consider matching only first name and matching full name both as evidence of confirmed legal ownership, whereas matching only last name as evidence that someone else in the household legally owns the residence. These assumptions may not be met. Matching only first or last name can be coincidental, and the respondents may have no relationship with the legal owner. Nonetheless, we believe that a small number of coincidental matches are unlikely to change the substantive conclusions.

Third, our validation benchmark focuses on legal owner-

ship and we assume that legal ownership is indicated by taxpayer names on parcel records. However, there are situations where the “true” ownership is not consistent with the names in the deed. For example, other legal claims like prenuptial agreements can make ownership arrangements different from the records. Also, the question item did not explicitly ask for legal ownership. If someone (e.g., an adult child) was paying for the home that is legally under a family member’s name (e.g., a parent), the person had substantively meaningful reasons to report themselves as an owner. These situations can challenge the validity of our benchmark data and they again highlight the importance of clearly defining ownership in the measurement.

4.1 Recommendations for Practice

Based on our results, we have three recommendations for measuring and analyzing homeownership. First, the questions should be “fit for use.” That is, when designing the homeownership question, it is important to determine the anticipated usage of the data. If the intention is to identify household-level ownership, then the question wording is relatively forgiving. Even if the wording is slightly ambiguous (as in *format I*), the general division between owner and non-owner should be reasonably robust. However, if the intention is to capture individual ownership, then researchers may need to clearly define ownership in the survey question. For example, is the ownership defined by perception, by property deed, by ability to access and dispose, or by asset division in case of household dissolution? The two-part question (*format II*) in the current study did not seem to be sufficient in capturing the complexity in household arrangements. A simple change may be to ask whose names are on the deed in the follow-up question, rather than generically using the word “own”.

Second, unless the survey has a clear interest in capturing special arrangements, we suggest not adding too many additional response options. If any additional option needs to be included, it is important to define it as mutually exclusive from the standard “own” and “rent” categories. Our results also suggest not including the “other” category if possible. People’s interpretations of “other” seemed highly unstable. Many of the “other” responses could validly fit under categories like rented or owned by someone else in the household. Adding the “other” category did not make the data more accurate.

Third, even if homeownership is defined in the survey item, people are still likely to interpret the question and response options differently. Our case study suggests that combining survey data and administrative records is one promising way to understand asset ownership within households. For example, one substantively meaningful question is how equal and unequal ownerships are related to household members’ relationship satisfaction (Deere & Doss, 2006; Kan &

Laurie, 2014). With data combined from surveys and property records, it is possible to differentiate between the effects of perceived versus actual equal/unequal ownerships on household members' psychological wellbeing.

References

- Abowd, J., & Stinson, M. (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *Review of Economics and Statistics*, 95(5), 1451–1467.
- Behavioral Risk Factor Surveillance System. (2019). BRFSS questionnaires. Retrieved from <https://www.cdc.gov/brfss/questionnaires/index.htm>
- Bennett, F. (2013). Researching within-household distribution: Overview, developments, debates, and methodological challenges. *Journal of Marriage and Family*, 75(3), 582–597.
- Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In *Handbook of econometrics* (Vol. 5, pp. 3705–3843). Elsevier.
- California Health Interview Survey. (2020). CHIS questionnaires. Retrieved from <http://healthpolicy.ucla.edu/chis/design/pages/questionnairesenglish.aspx>
- Chakrabarty, R. (1996). Current housing reports: A quality profile. *American Housing Survey*, 121.
- Childs, J. H., Carter III, G. R., Norris, D., Hanaoka, K., & Schwede, L. (2006). Cognitive test of the NRFU round 3: Revised questions. Retrieved from <https://www.census.gov/library/working-papers/2006/adrm/rsm2007-09.html>
- Cortes, A., Herbert, C., Wilson, E., & Clay, E. (2007). Factors affecting Hispanic homeownership: A review of the literature. *Cityscape*, 53–91.
- Davern, M., Call, K., Ziegenfuss, J., Davidson, G., Beebe, T., & Blewett, L. (2008). Validating health insurance coverage survey estimates: A comparison of self-reported coverage and administrative data records. *Public Opinion Quarterly*, 72(2), 241–259.
- Deere, C., & Doss, C. (2006). The gender asset gap: What do we know and why does it matter? *Feminist economics*, 12(1-2), 1–50.
- Dietz, R., & Haurin, D. (2003). The social and private micro-level consequences of homeownership. *Journal of Urban Economics*, 54(3), 401–450.
- DMACS. (2020). Detroit Metro Area Communities Study 2019–2020, wave 5, wave 6, wave 8, and wave 11 data files. Retrieved from <https://detroitssurvey.umich.edu/>
- Finnigan, R. (2014). Racial and ethnic stratification in the relationship between homeownership and self-rated health. *Social Science & Medicine*, 115, 72–81.
- Groen, J. (2012). Sources of error in survey and administrative data: The importance of reporting procedures. *Journal of Official Statistics*, 28(2), 173–198.
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). Questions and answers in surveys. *Survey methodology*, 217–257.
- Herbert, C., & Belsky, E. (2008). The homeownership experience of low-income and minority households: A review and synthesis of the literature. *Cityscape*, 5–59.
- Kan, M., & Laurie, H. (2014). Changing patterns in the allocation of savings, investments and debts within couple relationships. *The Sociological Review*, 62(2), 335–358.
- Kapteyn, A., & Ypma, J. (2007). Measurement error and misclassification: A comparison of survey and administrative data. *Journal of Labor Economics*, 25(3), 513–551.
- Kilic, T., & Moylan, H. (2016). *Methodological experiment on measuring asset ownership from a gender perspective (MEXA)*. Washington, DC: World Bank Group.
- Kreuter, F., Müller, G., & Trappmann, M. (2010). Non-response and measurement error in employment research: Making use of administrative data. *Public Opinion Quarterly*, 74(5), 880–906.
- Lindblad, M., & Quercia, R. (2015). Why is homeownership associated with nonfinancial benefits? A path analysis of competing mechanisms. *Housing Policy Debate*, 25(2), 263–288.
- Lyberg, L., & Kasprzyk, D. (1991). Data collection methods and measurement error: An overview. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys*. New York: Wiley.
- Manturuk, K. (2012). Urban homeownership and mental health: Mediating effect of perceived sense of control. *City & Community*, 11(4), 409–430.
- Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *International Journal of Public Opinion Quarterly*, 70(5), 737–758.
- Shiffer-Sebba, D., & Behrman, J. (2020). Gender and wealth in demographic research: A research brief on a new method and application. *Population Research and Policy Review*, 1–17.
- Tourangeau, R. (2003). Cognitive aspects of survey measurement and mismeasurement. *International Journal of Public Opinion Research*, 15(1), 3–7.
- Venti, S. (2011). Economic measurement in the Health and Retirement Study. *Forum for Health Economics and Policy*, 14(3).
- Voorheis, J. (2020). Evaluating administrative records as a potential sample frame for the national survey of college graduates. Presented at BigSurv20. Retrieved from <https://www.bigsurv20.org/conf20/program/?sess=19#217>

Wainer, A., & Zabel, J. (2020). Homeownership and wealth accumulation for low-income households. *Journal of Housing Economics*, 47, 101624.