Understanding Society
Working Paper Series
No. 2020 - 03
March 2020

Understanding the role of Temporary Sample Members for Understanding Society

Laura Fumagalli (University of Essex)

Understanding Society

Economic and Social Research Council

Non-technical summary

Understanding Society is a household sample, designed to represent the population living in households at wave 1. Members of the original sample are followed indefinitely. Over time, new people join *Understanding Society* households, by birth, partnership or cohabitation. Depending on their relationship with the original sample some of these joiners are followed if they subsequently leave the original household, but many stop being part of the Study at this stage. The latter group are known as Temporary Sample Members (TSMs) and are the focus of this report.

- While part of an original *Understanding Society* household, TSMs provide important contextual information, for example different components of household income.
- TSMS are drawn disproportionately from interesting (mobile, hard-to-reach) groups. For example, people who return home after a period of absence, which means they were not in the household when it was first sampled – for example students, people returning from living abroad, or from prison or other long-term institutions, and short-term house-sharers (during or shortly after university).
- TSMS can help maintain cross-sectional representativeness of a study. For example, immigrants who arrive in the country after the date of the original baseline sample being selected will not be captured. However, new immigrants, who may be an increasingly important part of the changing population, and of considerable policy interest, do enter the study (until a new boost sample) through cohabitation. In *Understanding Society*, significant numbers of European migrants are TSMs.
- When TSMs leave a household, they might provide important contextual information on the *Understanding Society* household. For example, data from nonresident parents may be useful for understanding the lives of children who remain in the sample, and former cohabitants may continue to share financial resources.
- The paper shows that expanding the following rules would increase sample size, and in particular, it might increase the sample size of otherwise hard to reach sub-groups and therefore help to support sub-group analysis.
- Data quality from TSM interviews is similar to original sample members. However, the rates of refusals and proxy interviews are higher among TSMs.
- Further research is underway to consider if, in any circumstances, we might continue to follow TSMs once they leave an original *Understanding Society* household.

Understanding the role of Temporary Sample Members for Understanding Society

Laura Fumagalli (University of Essex)

Abstract: This paper assesses the value to *Understanding Society* of joiners, and whether there might be benefit to the Study of following more of them indefinitely, through a change in the following rules that would transform (some of) the current TSMs in OSMs. This analysis of *Understanding Society* also provides guidance to other panel surveys of the potential benefits and costs of extending the number of joiners followed. Our two research questions are: whether TSMs (who do not join the study as new-borns) are drawn disproportionately from policy-relevant sub-population which surveys generally struggle to capture in adequate numbers; and whether such TSMs provide good quality data. Of course the quality of data provided by these TSMs when cohabiting with an OSM may differ from the data they would provide when not cohabiting with an OSM, should the following rules be expanded to collect this.

Keywords: population, sample size, household composition.

Acknowledgements: I thank Michaela Benzeval, Chris Bollinger, Magda Borkowska, Jon Burton, Tom Crossley, Paul Fisher, Peter Lynn, and Alina Pelikh for useful comments and suggestions, and George Poquillon for skilful research assistance.

Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service. This methodological briefing has been funded by the Economic and Social Research Council UK Population Lab Innovation Development Grant (ES/S016651/1).

Data note: University of Essex. Institute for Social and Economic Research, NatCen Social Research, Kantar Public. (2018). *Understanding Society*: Waves 1-8, 2009-2017 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 11th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-12

Corresponding author: Laura Fumagalli, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, lfumag@essex.ac.uk

Introduction

Household panels such as *Understanding Society* are surveys where a sample of a population of interest is repeatedly interviewed over time. Household panels are an important resource for social sciences: they allow the analysis of the changes in the life of the population under study over an extended period of years. As such, household panels provide a powerful tool to identify the causes and the effects, including long-term effects, of societal changes.

Household panels begin by selecting a sample of households from of a given target population in a way that allows each households' probability of inclusion in the sample to be calculated. This allows population inferences to be drawn from the sample. The members of the households that are initially sampled (henceforth: "starters") are interviewed repeatedly over time. Starters are followed indefinitely, in principle until death or emigration, including when they leave to form new households.

Households are dynamic and household composition changes over time. New household members (henceforth: 'joiners') join starter households through birth/adoption or cohabitation. All major household panels collect data on joiners as long as they live with a starter. For example, income data are generally collected from all family members (starters and joiners alike). This permits the computation of household income: a measure of the resources available to the household widely used in research on poverty and inequality, and in analyses of the impact of households' socio-economic status on child outcomes.

What happens when joiners cease to live with a starter? Household panels differ in the choices they make regarding which joiners are followed indefinitely (for a full review of these differences, see: Schonlau et al., 2010). The rules defining the circumstances under which some types of joiners are followed indefinitely, and others are not, are called "following rules".

There are two main reasons why not all joiners are followed indefinitely. The first is simply that total fieldwork costs increase with the number of people followed. This reason applies to all types of joiners. The second reason is that the sample inclusion probabilities of some types of joiners can be difficult to calculate. For example, the selection probability of a woman joining the sample as a partner of a starter depends both on her and her partner's characteristics and past history, and some of this information may be unknown. When a sample inclusion probability cannot be calculated for an individual, standard statistical methods for population inferences cannot make use of that individual's responses. This second reason only applies to respondents joining through cohabitation with a starter, as there exist well-worked out ways of assigning joiners born from starters a selection probability based on the known selection probabilities of their parents (Schonlau et al., 2013). For these reasons, most panel surveys follow joiners born from starters, while joiners who entered the survey through cohabitation with a starter (e.g., partners and house sharers) are followed less frequently.

Despite the difficulties, expanding following rules to follow joiners who enter the survey through cohabitation beyond that period of cohabitation represents an important opportunity for surveys. There are several reasons why this is the case. First, following joiners entering through cohabitation with a starter can help to maintain the cross-sectional representativeness of a longitudinal study (see also: Fitzgerald et al., 1998; Schonlau et al., 2013). One population of great interest is the *current* population of a jurisdiction (e.g. the current population living in the UK), and this changes over time through migration or demographic transitions occurring after the first wave of a longitudinal study. Following starters alone can maintain longitudinal representativeness of the target population at the time of the first wave, a population that becomes increasingly dated over time. Following births from starters can capture the changes due to demographic transitions, but it cannot capture changes due to migration. Following other types of joiners can help to capture changes due to migration. New immigrants, who were not in the country at the time the sample was first collected, cannot enter the sample as starters, but they can still join the survey as partners or house sharers.

Second, following joiners who entered through cohabitation with a starter beyond the period of cohabitation can facilitate the analysis of peer and contextual effects on starters (see, for example: Fadlon and Nielsen, 2019). Peer and contextual effects may extend beyond the household. Respondents may be influenced by (or share resources with) not just those with whom they currently cohabit, but also by former cohabitants. Following individuals who join the sample as cohabitants of starters can help uncover these effects.

Third, following joiners who entered through cohabitation with a starter indefinitely can boost sample size. In particular, widening the following rules in this way can increase the coverage of policy-relevant populations for which the sample size of starters alone is too small to support analysis. For example, continuing to follow joiners who entered the study through cohabitation with a starter can provide valuable information on groups - such as students or young professionals - that are more likely to be mobile, and thus are less likely to be captured by the initial sample design. Where joiners enter households as respondents' partners, following these joiners when they leave can also help study the consequences of partnership dissolution (Schonlau et al., 2013). With time, more and more couples in the panel will include a joiner. Therefore, if only starters are followed, the share of partners (and their children) followed after a separation will decrease over time. This will reduce the quantity and the usefulness of data available to study the consequences of partnership dissolution.

The main alternative way to increase sample sizes of policy-relevant populations is via boost samples. Boost sample have the important advantage of providing known selection probabilities. Indeed, Understanding Society uses boost samples (the Ethnic Minorities and the Immigrants and Ethnic Minorities Boost samples) to increase the sample of ethnic minorities and recent immigrants. However, boost sample can be expensive. Moreover, they can be of limited use in targeting rare, mobile, or difficult to identify populations. For example, designing a boost sample of splitting

partners would be difficult. In contrast, joiners enter the study as a part of the natural social processes of household change, and thus are relatively easy and cheaply incorporated into the survey.

Past research on the potential benefits generated by following different groups of joiners has focused on the impact of different following rules on total sample size (Schonlau et al., 2010). However, the value of extending the types of joiners that are followed indefinitely hinges on i) whether these joiners contribute sample size on specific subgroups of interest, and ii) whether they provide data of reasonable quality. These two issues have not been previously studied, and this paper addresses that evidence gap. This paper fills that gap by analysing data from the *Understanding Society* panel: a major panel survey started in 2009 with a probability sample of approximately 40,000 households drawn from the population living in the United Kingdom and carried out annually after that. In *Understanding Society* starters and most of their children are called Original Sample Members (OSM). OSMs are followed indefinitely unless they die, leave the country or repeatedly refuse to continue to take part in the survey. People joining the sample through cohabitation and most of their children are called Temporary Sample Members (TSM). TSMs are currently followed only while they live with an OSM. Male TSMs who have a child with a female OSM become Permanent Sample Members (PSM). PSMs are also followed indefinitely (full details of the following rules used by *Understanding Society* can be found in the Appendix).

This paper assesses the value to *Understanding Society* of joiners, and whether there might be benefit to the Study of following more of them indefinitely, through a change in the following rules that would transform (some of) the current TSMs in OSMs. This analysis of *Understanding Society* also provides guidance to other panel surveys of the potential benefits and costs of extending the number of joiners followed. Our two research questions are: whether TSMs (who do not join the study as new-borns) are drawn disproportionately from policy-relevant sub-population which surveys generally struggle to capture in adequate numbers; and whether such TSMs provide good quality data. Of course, the quality of data provided by these TSMs when cohabiting with an OSM may differ from the data they would provide when not cohabiting with an OSM, should the following rules be expanded to collect this. Nevertheless, assessing the quality of the data TSMs currently provide gives a useful starting point for considering this question.[1]

The rest of the paper proceeds as follows. Section 2 describes the data and the sample used. Section 3 presents the results. In particular, Section 3.1 sets the stage by documenting how the number of TSMs evolves over time under the existing following rules. Section 3.2 describes who TSMs are, and studies whether TSMs provide data on small, policy-relevant sub-populations. Section 3.3 presents a case study of one interesting group: respondents involved in partnership dissolution. Section 3.4

[1] The data that TSMs provide when they no longer live with OSMs might be better or worse than the data that they currently provide while living with an OSM. On the one hand, the absence of OSMs may increase the probability of non-contact or refusal for TSMs, thus increasing unit non-response among TSMs. On the other hand, it will not be possible for an OSMs to give a proxy interview for TSMs after they cohabit, and this may result in an increase in the probability TSMs give a full interview.

explores the quality of the data provided by TSMs. Finally, Section 4 concludes and offers some recommendations for the future development of *Understanding Society*.

Methods and Sample

The analysis that follows uses *Understanding Society* data (University of Essex. Institute for Social and Economic Research, NatCen Social Research, Kantar Public. 2018). All figures are provided unweighted, as our aim is to describe the realized sample, and to compare OSMs and TSMs in the sample, rather than to draw inferences about the population. All pieces of analysis share the following sample inclusions and restrictions: i) The General Population Sample, Ethnic Minorities Boost (EMB) and the Immigrants and Ethnic Minorities Boost (IEMB) are included when available; ii) the BHPS sample component is excluded; iii) individuals who entered the survey as new-borns are excluded, irrespective of their sample status.

New-borns are excluded for two reasons.[2] First, OSMs and TSMs new-borns, despite their different sample status, are brought into the sample by the same types of demographic changes in the population. Therefore, OSM and TSM new-borns are likely to have very similar characteristics. Newborn TSMs are unlikely to cover specific policy-relevant sub-populations not covered by newborn OSMs. Second, babies and young children do not provide an interview themselves. Therefore, comparing the quality of the data from TSM and OSM new-borns is not possible.

The BHPS sample component is excluded for the following reason. When they joined *Understanding Society*, the BHPS respondents had been in the panel for long time (some of them for as long as 18 years). Therefore, BHPS respondents in *Understanding Society* are potentially a selected subsample of respondents. Moreover, the following rules for BHPS and *Understanding Society* are slightly different and this complicates the inferences we might draw from the combined sample.

The different pieces of analysis use different sets of waves from *Understanding Society.* The description of TSM numbers uses Waves One through Eight (that is, all the waves available at the time of writing.) The analysis of the TSM to PSM transitions and the analysis of the coverage of individuals going through partnership dissolution use waves two to eight, as, by definition, no TSM to PSM transition nor partnership dissolution can be observed at Wave One. The analysis of the characteristics of TSMs and the analysis of the quality of the data TSMs provide use a single wave. This helps minimize the danger that the results are driven by differential attrition probabilities: if all waves were retained, individuals less likely to drop out from the sample would be over-represented. We chose to use Wave Five data for this purpose.

[2] New-borns are those recorded as new-entrants at age zero or one.

Results

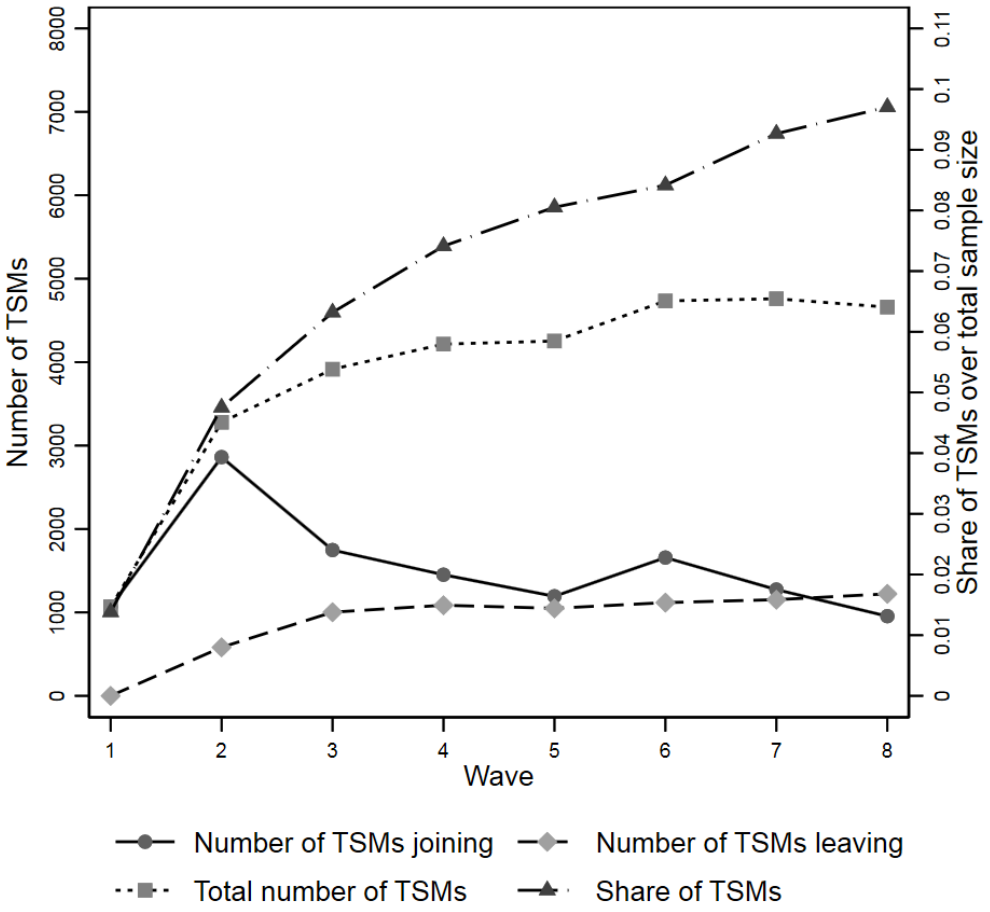
The evolution of TSM numbers

Figure 1 looks at the evolution of TSM absolute numbers (left axis) and the share of TSMs over the total sample size (right axis). All numbers in Figure 1 are based on all individuals in responding households, irrespectively of whether they provide a full interview. The dashed black line shows the number of TSMs joining the sample by wave. TSMs are considered to have joined the sample at a given wave if they are first observed in that wave. The dotted black line shows the number of TSMs leaving the sample by wave. A TSM is considered to have left the sample at a given wave if they are last observed in the wave previous to that one. TSMs temporarily leaving the sample at wave *t,* and then returning to the sample later on are not considered leavers. The solid black line shows the total number of TSMs by wave. This is computed by counting the number of TSMs in a given wave at any point in time. This count excludes TSMs who leave the sample temporarily. Therefore, the total number of TSMs by wave, indicated by the solid black line, may be slightly lower than the total implied by the dotted and the dashed black lines. Finally, the grey solid line shows the share of TSMs over the total sample size.

Figure 1 shows that the number of TSMs joining the sample is low at Wave One. TSMs joining at Wave One are a specific category, that is, non-ethnic minority members of ethnic minority households in the Ethnic Minority Boost Sample (EMBS) who were designated as TSMs by design. The other types of TSMs (e.g., partners and house sharers) enter the sample at Wave Two. The number of TSMs then decreases wave-on-wave, has an increase at Wave Six, then decreases wave-on-wave again. The small peak at Wave Six is due to the fact that the IEMB was added and non-IEM members of IEMB households were given the TSM status. The wave-on-wave decline in the number of joining TSMs observed in waves two to five and seven to eight is likely to be due to a slight decline in the overall sample size (due to attrition), with the percentage of TSMs joining the sample being roughly the same across waves. Finally, the dotted black line shows that the number of TSMs leaving the sample increases very slowly over time.

Due to the movements in and out of the sample discussed above, the total number of TSMs, shown by the solid black line, increases over time until almost 5000 units at wave six, and then remains stable at waves seven and eight. The total number of TSMs increases at a decreasing rate. This suggests that increasing the types of TSMs that are followed indefinitely is unlikely to lead to an explosion of sample size and costs (Schonlau et al., 2010 reach a similar conclusion for the German panel SOEP). Finally, the grey solid line suggests that the share of TSMs over the total sample size increases (at a slightly decreasing rate) over time. However, with the current following rules, this share remains rather limited, and is always below 10%.

Figure 1: The evolution of TSM numbers

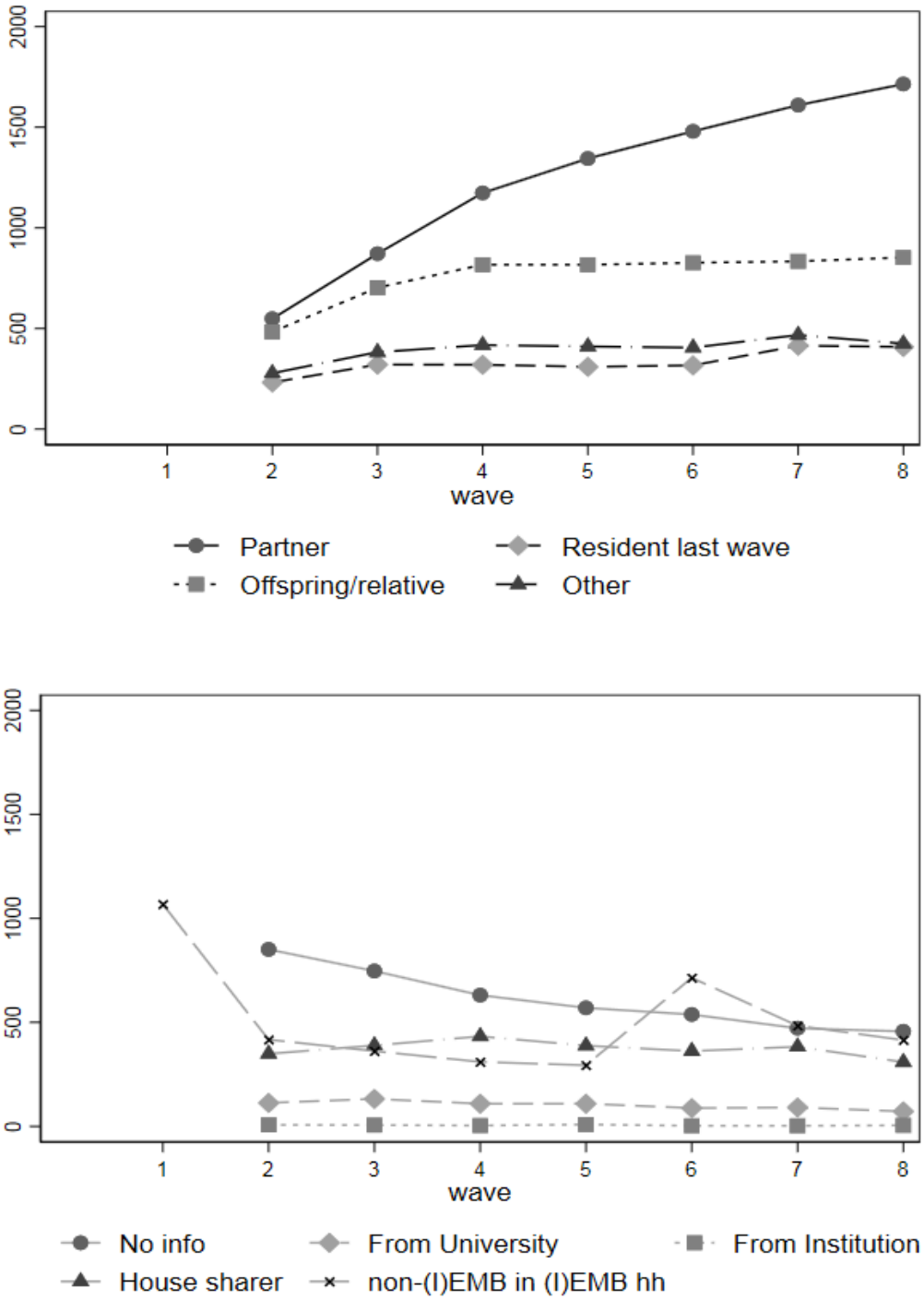


Note: *Understanding Society*, Waves 1 to 8, all individuals in responding households. People who entered as new-borns and from the BHPS sample components are excluded. The wave when TSMs join the sample is the wave in which TSMs are first observed; the wave when TSMs leave the sample is the wave after the one in which TSMs are last observed. The total number of TSMs also takes into account the fact that some TSMs leave the sample temporarily. Therefore, the number indicated by the solid black line may be slightly lower than the one implied by the dotted and the dashed black lines. The number of TSMs leaving the sample at wave eight is likely to be slightly overestimated as wave nine has not yet be released and thus it is impossible identify TSMs who leave the sample only temporarily.

Figure 2 shows trends in the total number of TSMs by reason of becoming a TSM. The TSM types displayed on the left panel increase in number over time. Partners (black solid line) are the TSMs whose number increases more rapidly. Partners are likely to stick with their OSM family members, and thus their entry rate is higher than their exit rate (figures on entry and exit rates by TSM type are available upon request). Most of the TSM types in the right panel (namely, house sharer, people joining from university or from an institution) exhibit stable numbers over time: the entry rate of these TSM types is almost perfectly compensated by their exit rate. The number of non-immigrant/non-ethnic minority TSMs in immigrant/ethnic minority households peaks at wave one

and six (when the Ethnic Minority Boost Sample and the Immigrant and Ethnic Minority Boost Sample were introduced) and then decreases over time, as exit rates exceed entry rates.

Figure 2: The evolution of TSM numbers (by reasons of becoming TSM)

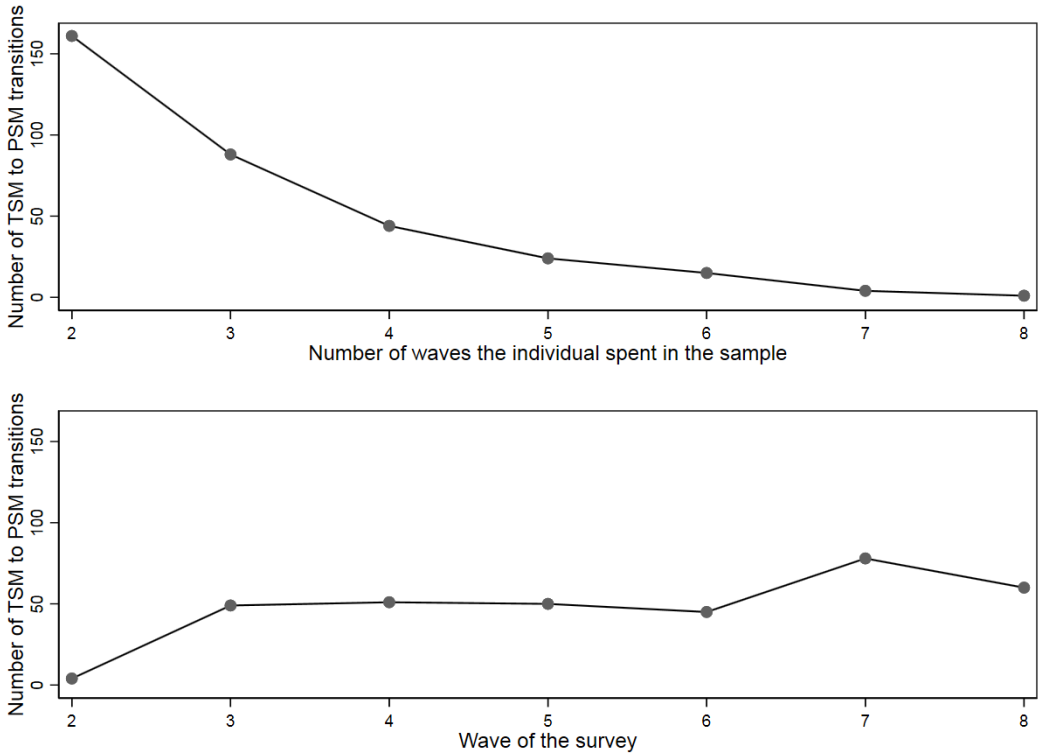


Note: *Understanding Society*, Waves 1 to 8, all individuals in responding households. People who entered as new-borns and from the BHPS sample component are excluded.

The top panel of Figure 2 shows that the partners are the fastest growing category among TSMs. This is in spite of the fact that TSM partners are also the most likely category to transition to PSM status. TSMs can become PSMs by fathering a child.[3] The top panel of Figure 3 shows transitions from TSM to PSM status by waves the individual has been in the sample. Out of the total number of transitions (337), almost a half occurs in the second wave those TSMs are in the sample, i.e., in the first wave they can make that transition. The number of TSM to PSM transitions declines with permanence in the sample, and, by the eighth wave in the sample, there is virtually no TSM to PSM transition.

The bottom panel of Figure 3 shows transitions from TSM to PSM status by wave in which they occur. Transitions start occurring at wave three. These transitions are driven by individuals joining the sample as TSMs (mainly as partners) at wave two and transitioning to PSM in their second wave in the sample. The number of TSM to PSM transitions remains roughly stable and it is mainly driven by individuals transitioning to PSM the second wave they are in the sample. The slight decline in sample size, and thus in the number of TSMs joining the study (see Figure 1) is compensated by the small number of transitions occurring after the second wave individuals are in the sample. There is a small increase in the number of TSM to PSM transitions at wave seven, likely due to the higher number of TSMs entering the sample at wave six (see Figure 1) becoming PSMs the second wave they are in the sample.

Figure 3. TSM to PSM transitions



[3] Children of a TSM woman and OSM man do not become OSM, and thus it is not possible for women to become PSM through motherhood.

Note: *Understanding Society*, Waves 1 to 8, all individuals in respondent households. People who entered as new-borns and from the BHPS sample component are excluded. Total number of TSM to PSM transitions: 337.

Who are the TSMs?

This subsection describes TSMs and compares them to OSMs. This helps identify whether TSMs can capture some small, policy-relevant populations that cannot be adequately covered using data on OSMs only (our first research question). We do not analyse PSMs, as they are a very specific category with characteristics that are halfway between the characteristics of OSMs and the characteristics of TSMs.

We select characteristics that mark interesting policy-relevant populations, namely: age, gender, marital status, birthplace, education and ethnicity. For each of these characteristics, Table 1 reports mean, standard deviation and sample size for TSMs (column 1), OSMs (column 2) and OSMs living with a TSM (column 3). OSMs living with TSMs are an interesting subsample, as they are likely to be more similar to TSMs than the full group of OSMs. Table 1 also compares the characteristics of TSM and OSMs (column 4) and the characteristics of TSMs and OSMs living with a TSM (column 5). Consider first the comparison of TSMs and OSMs. TSMs are younger than OSMs. TSMs are 34 years old on average, while OSMs are 41 years old (compare columns (2) and (5)). TSMs are 8 pp more likely than OSMs to be employed. Compared to OSMs, TSMs are 26 percentage points more likely to be single and 25 pp less likely to be in a couple. There are not striking differences in where TSMs and OSMs live. However, TSMs are more prevalent than OSMs in London (20 vs 15 pp) and in Yorkshire and the Humber (10 vs 9 pp); TSMs are less prevalent than OSMs in the South West (7 vs 8 pp), in Scotland (5 vs 7 pp) and in Northern Ireland (2 vs 4 pp). TSMs have on average higher education than OSMs. TSMs are more likely than OSMs to have a University degree (29 vs 24 pp), an A-level qualification (34 vs 33 pp), and less likely to have a GCSE qualification (27 vs 30 pp) or no qualification at all (10 vs 13 pp). Finally, TSMs are almost one pp more likely to be born in Europe and over one pp less likely to be born in Asia. Consequently, TSMs are more likely than OSMs to be White and less likely to be Asian.

Turning to a comparison of TSMs with OSMs who live with TSMs, most of the differences shown above reduce or even disappear. For example, there is still suggestive evidence that TSMs have higher education than the OSMs they live with, but the differences between the two groups seldom reach conventional levels of statistical significance. However, in the case of birth/ethnicity, the differences between TSMs and OSMs become even more pronounced when we only consider OSMs living with TSMs. TSMs are 5 pp more likely to be born in UK and 3 pp less likely to be born in Asia/other countries than their cohabiting OSMs. As a consequence, TSMs are more likely to be White and less likely to be Asians, Black or other ethnicity.

Finally, in a couple of meaningful cases, the sign of the difference between TSMs and OSMs even changes sign when we only look at OSMs cohabiting with TSMs. This is the case of the variables for employment status and for living in London. When the entire group of OSMs is considered, TSMs are 8 pp points more likely to be employed; when we restrict the sample of OSMs to cohabitants of TSMs, our analysis suggests TSMs are 2 pp less likely to be employed than their cohabiting OSMs. Moreover, column (10) suggests TSMs are 5 pp more likely to live in London than OSMs, while column (13) suggests TSMs are 3 pp less likely to live in London than OSMs who live with a TSM.[4]

Figures 4 to 7 present some selected characteristics by reasons for becoming TSM. For comparison, the dashed black line shows corresponding values calculated in the sample of all OSMs; the dotted black line shows corresponding values calculated in the sample of OSMs living with TSMs; the dashed grey line shows corresponding values calculated in the sample of all TSMs.

Figure 4 shows that almost all groups of TSMs are younger and less likely to be female than OSMs (see left and middle panel). The higher employment rate of TSMs (compared to the full sample of OSMs) is driven by partners and house sharers (see the right panel of Figure 4). Figure 5 looks at marital status. The left panel suggests the high prevalence of single people among TSMs is driven by all categories (including people becoming TSMs as partners), but particularly the younger groups: people joining the household from institutions and university, offsprings and relatives, house sharers. Both TSM house sharers and TSM partners are less likely than OSMs to be the surviving partner of a past relationship (see right panel, second row); however, while TSM house sharers are also less likely than OSMs to be separated, TSM partners are more likely than OSMs to be separated. Figure 6 looks at education: it shows that the higher level of education estimated for TSMs is driven by TSMs returning from University, and, although less so, by house sharers. These TSMs are more likely than OSMs to have a degree and an A-level qualification, and less likely to have no qualification. Finally, figure 7 looks at birthplace. Partners -and partially house sharers- are more likely to be born in Europe (middle panel) and less likely to born in Asia (right panel).

In sum, Table 1 and Figures 4 to 7 suggest that these TSMs provide sample on a number of policy-relevant groups that may be difficult to capture using only data from OSMs. The first group is composed of people returning home after a period of absence. In particular, data on TSMs can help capture those who return home after university (the “boomerang generation”). Data from the ONS show the number of adults living with their parents is increasing. Slow economic growth, and rising housing prices and university fees mean that the number of young people who return to the parental home after university is increasing. There is still very little research on the consequences of this phenomenon, but recent studies (e.g., Tosi and Grundy, 2018) suggest potentially negative consequences at least for parents. The second group is house sharers: predominantly people who decide to share an accommodation with unrelated individuals during or soon after university. Household sharing is also on the rise (Green and McCarthy, 2015), with consequences that are still unknown. The third policy relevant group is recent European migrants. Table 1 and Figure 4 show UK

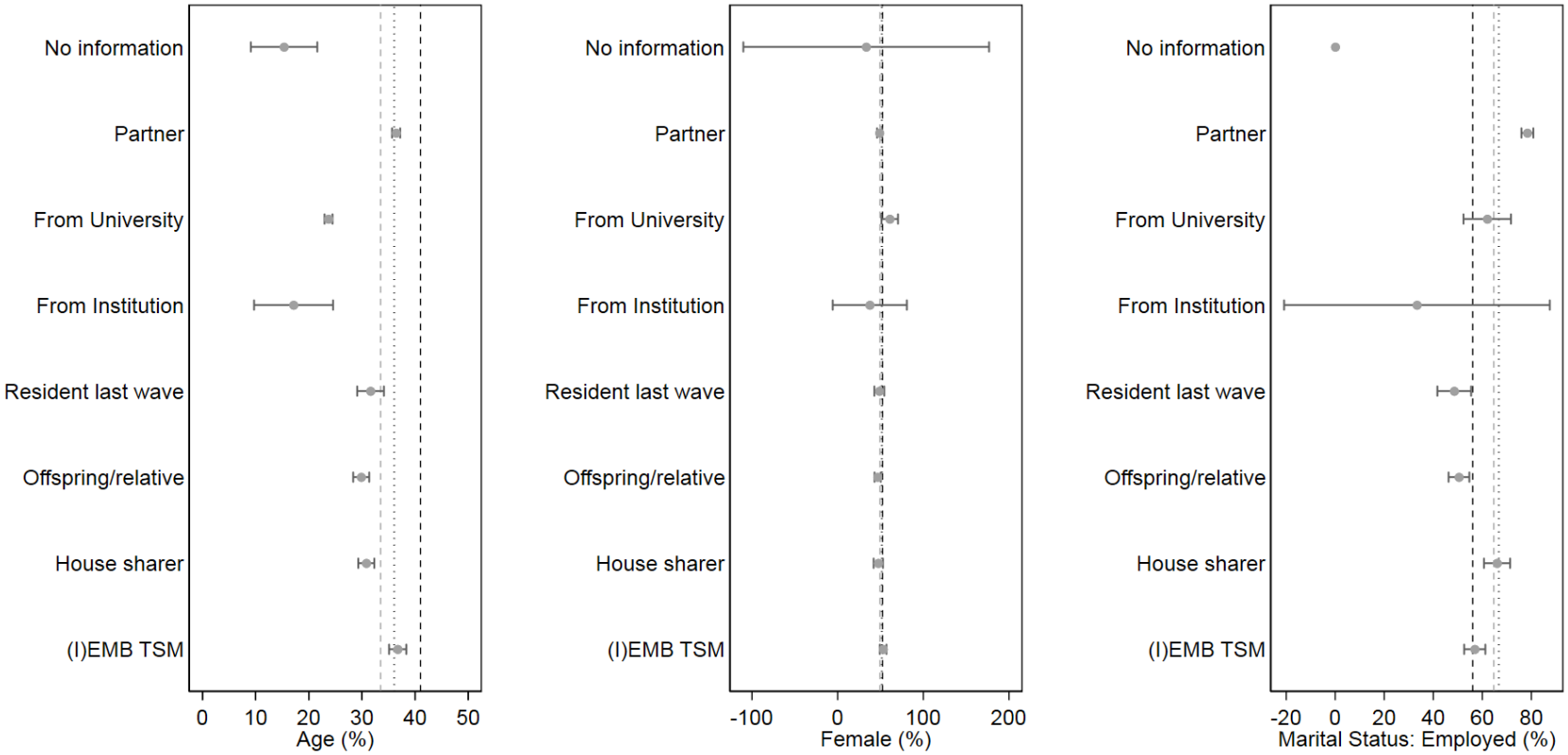
natives and people born in countries of traditional migration to the UK (such as South-Asians) are under-represented among TSMs. In contrast, people born in Europe are over-represented, especially among partners. Immigrants from Europe are a very relevant category, particularly in the aftermath of Brexit.

Table 1 TSMs vs OSMs: main characteristics

	(1) TSM			(2) OSM			(3) OSM living with TSM			(4) TSM-OSM			(5) TSM-OSM living with TSM		
	N	Mean	sd	N	Mean	sd	N	Mean	sd	Point Estimate	se	p-value	Point Estimate	se	p-value
Age	3215	33.514	17.909	42388	40.970	22.409	4070	36.081	18.692	-7.457	0.334	0.000	-2.567	0.431	0.000
Female	3216	49.378	50.004	42388	52.116	49.956	4070	51.597	49.981	-2.738	0.915	0.003	-2.219	1.180	0.060
Employed	2764	64.689	47.802	34784	56.247	49.609	3473	66.743	47.120	8.442	0.947	0.000	-2.055	1.211	0.090
Marital Status: Single	2766	56.761	49.550	34812	30.880	46.201	3478	44.710	49.726	25.881	0.974	0.000	12.051	1.264	0.000
Marital Status: In a couple	2766	28.091	44.953	34812	52.614	49.932	3478	39.620	48.918	-24.523	0.896	0.000	-11.529	1.191	0.000
Marital Status: Separated	2766	11.822	32.293	34812	10.390	30.514	3478	11.932	32.421	1.432	0.635	0.024	-0.110	0.824	0.894
Marital Status: Surviving Partner	2766	3.326	17.935	34812	6.116	23.962	3478	3.738	18.971	-2.790	0.364	0.000	-0.412	0.469	0.380
GOR: North-East	3208	3.335	17.959	42367	3.909	19.380	4060	2.906	16.801	-0.573	0.331	0.083	0.429	0.412	0.298
GOR: North-West	3208	10.006	30.013	42367	10.600	30.784	4060	10.665	30.871	-0.594	0.551	0.281	-0.659	0.718	0.359
GOR: Yorkshire and the Humber	3208	9.726	29.635	42367	8.688	28.167	4060	9.310	29.061	1.037	0.541	0.055	0.415	0.694	0.550
GOR: East Midlands	3208	7.824	26.859	42367	7.980	27.099	4060	7.488	26.322	-0.156	0.492	0.751	0.337	0.629	0.593
GOR: West Midlands	3208	9.695	29.593	42367	8.990	28.605	4060	8.818	28.359	0.704	0.541	0.193	0.877	0.686	0.201
GOR: East of London	3208	8.853	28.411	42367	9.134	28.810	4060	8.424	27.778	-0.282	0.521	0.589	0.429	0.665	0.518
GOR: London	3208	19.545	39.661	42367	14.851	35.561	4060	22.069	41.476	4.694	0.721	0.000	-2.524	0.956	0.008
GOR: South East	3208	12.313	32.864	42367	12.496	33.067	4060	12.660	33.257	-0.183	0.602	0.762	-0.347	0.780	0.656
GOR: South West	3208	7.014	25.542	42367	8.112	27.303	4060	6.749	25.090	-1.099	0.470	0.019	0.265	0.599	0.658
GOR: Wales	3208	4.988	21.772	42367	4.397	20.504	4060	4.458	20.641	0.590	0.397	0.137	0.529	0.503	0.292
GOR: Scotland	3208	5.019	21.836	42367	6.614	24.852	4060	4.458	20.641	-1.595	0.404	0.000	0.561	0.504	0.266
GOR: Northern Ireland	3208	1.683	12.867	42367	4.227	20.121	4060	1.995	13.985	-2.544	0.247	0.000	-0.312	0.316	0.324
Birthplace: UK	2397	83.605	37.031	35381	84.283	36.397	3540	79.153	40.628	-0.678	0.781	0.385	4.452	1.019	0.000
Birthplace: Ireland	2397	0.417	6.447	35381	0.545	7.366	3540	0.480	6.914	-0.128	0.137	0.350	-0.063	0.176	0.720
Birthplace: Europe	2397	2.295	14.976	35381	1.441	11.919	3540	1.215	10.956	0.853	0.312	0.006	1.080	0.357	0.003
Birthplace: Asia	2397	5.298	22.405	35381	6.096	23.927	3540	8.333	27.642	-0.798	0.475	0.093	-3.035	0.652	0.000
Birthplace: Africa	2397	2.169	14.571	35381	1.761	13.152	3540	2.062	14.213	0.409	0.306	0.182	0.107	0.382	0.779
Birthplace: Other	2397	6.216	24.150	35381	5.873	23.513	3540	8.757	28.271	0.343	0.509	0.500	-2.541	0.685	0.000
Education: Degree	2041	28.760	45.276	32607	24.173	42.814	3228	27.014	44.410	4.588	1.030	0.000	1.747	1.271	0.169
Education: A level	2041	34.444	47.530	32607	32.545	46.855	3228	33.736	47.288	1.899	1.084	0.080	0.708	1.341	0.598
Education: GCSE	2041	26.507	44.148	32607	29.905	45.785	3228	29.337	45.538	-3.398	1.010	0.001	-2.830	1.264	0.025
Education: No Qualification	2041	10.289	30.389	32607	13.377	34.042	3228	9.913	29.889	-3.088	0.699	0.000	0.376	0.854	0.660
Ethnicity: White	2046	82.014	38.417	33910	80.431	39.674	3437	68.199	46.577	1.583	0.876	0.071	13.815	1.163	0.000
Ethnicity: Asian	2046	10.117	30.163	33910	11.737	32.186	3437	17.719	38.188	-1.620	0.689	0.019	-7.602	0.932	0.000
Ethnicity: Black	2046	4.203	20.071	33910	4.951	21.694	3437	8.088	27.270	-0.748	0.459	0.103	-3.885	0.643	0.000
Ethnicity: Other	2046	3.666	18.796	33910	2.881	16.728	3437	5.994	23.740	0.785	0.425	0.065	-2.328	0.580	0.000

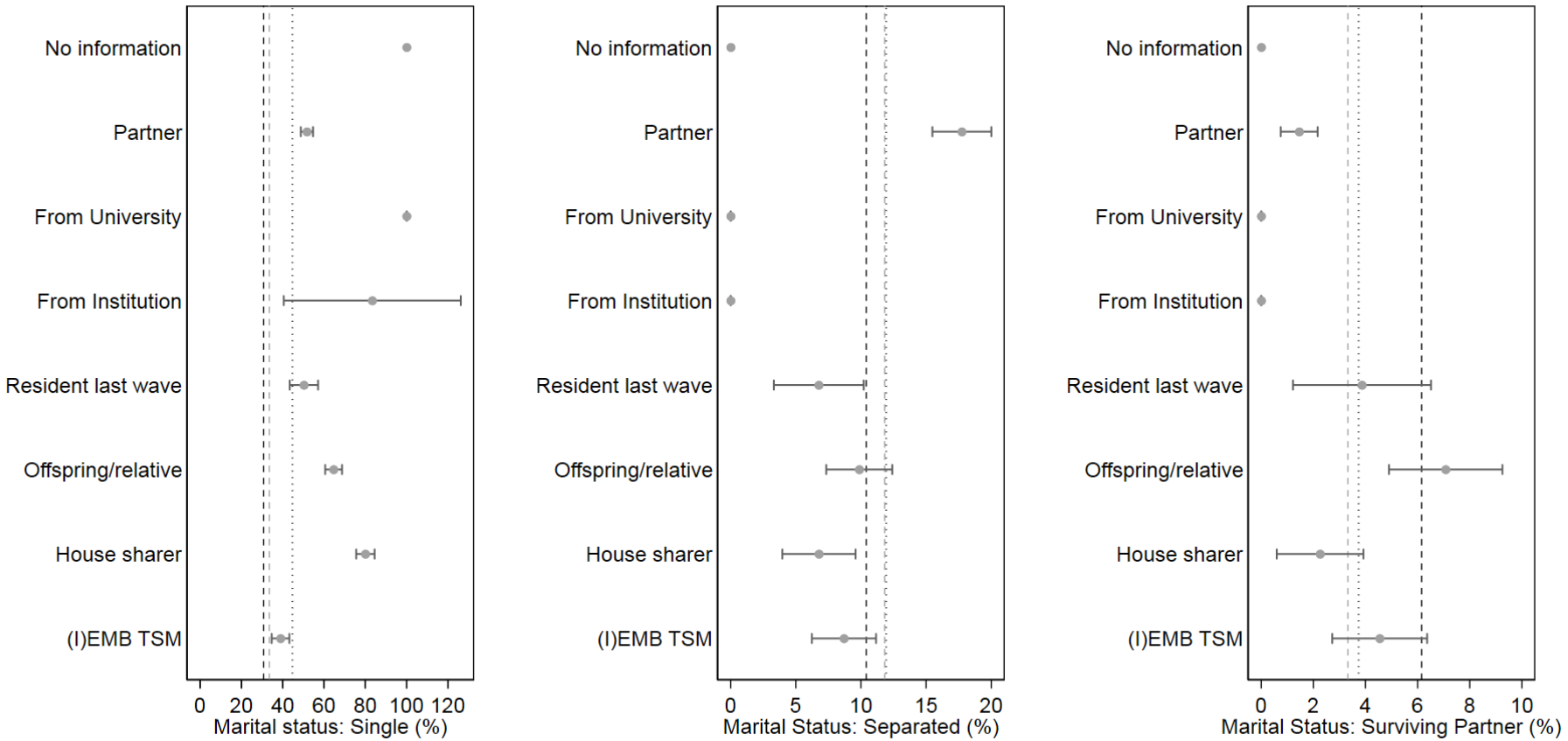
Note: *Understanding Society*, Wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines indicate the value in the TSM sample. The variable "Birthplace: Europe" excludes Ireland, the variable "Birthplace: Asia" includes both China and South Asia. The variable "Education: A level'" includes higher education qualifications which cannot be considered university degrees. The variable "Education: GCSE" includes any other qualification that cannot be considered an A-level or a higher education qualification.

Figure 4: TSM Demographics, by reasons of becoming TSM



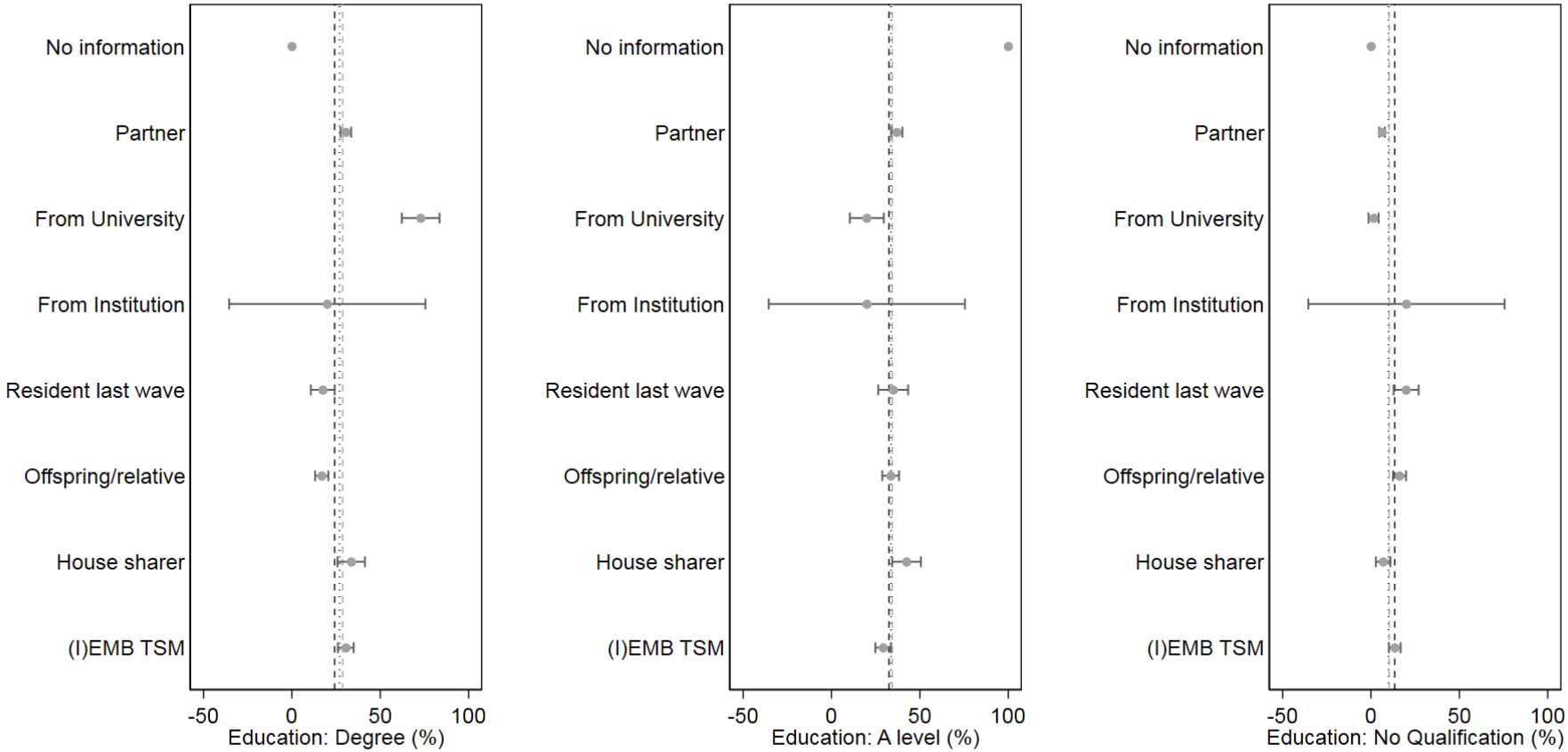
Note: *Understanding Society*, Wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines indicate the value in the TSM sample.

Figure 5: TSM Marital status, by reasons of becoming TSM



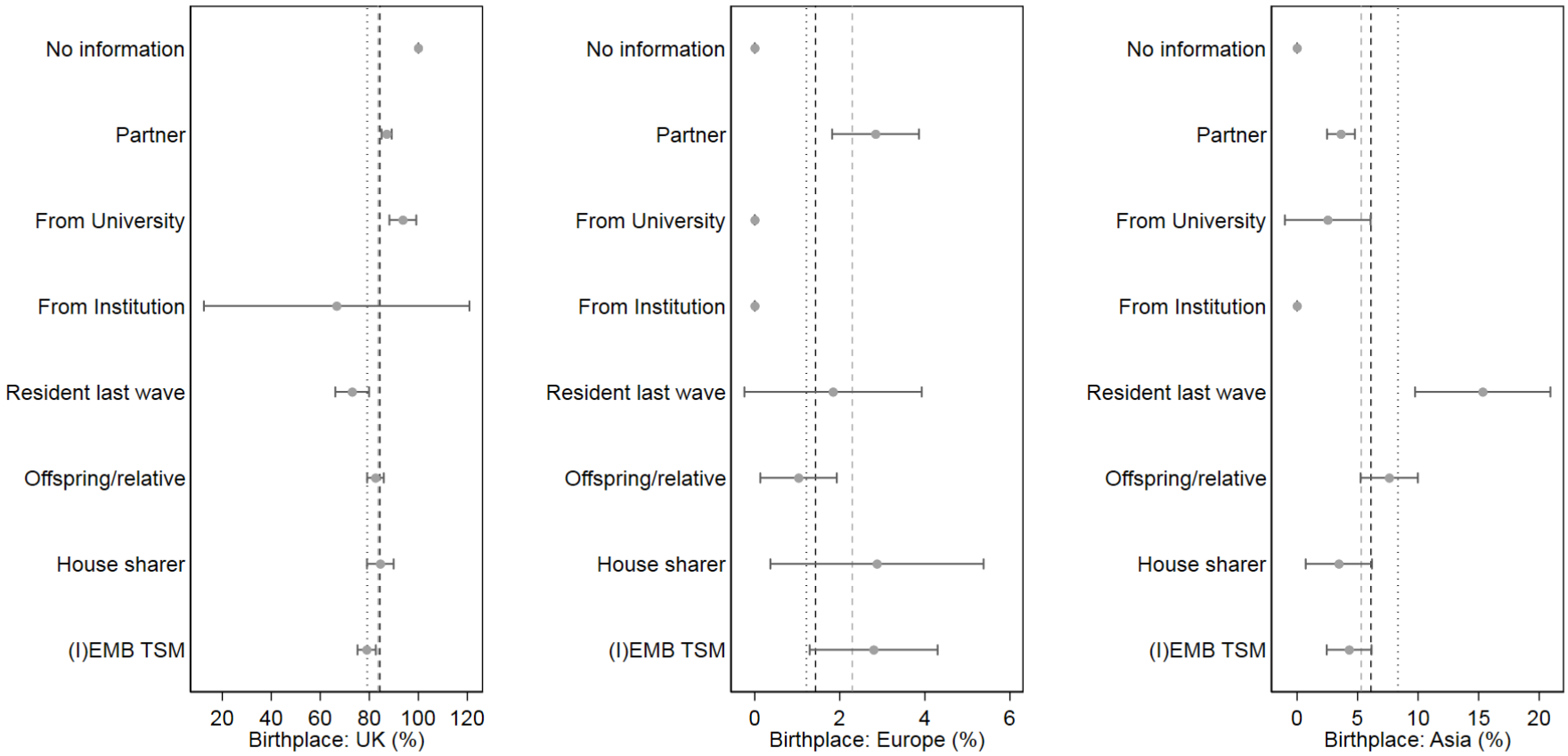
Note: *Understanding Society*, Wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines indicate the value in the TSM sample.

Figure 6: TSM: Education, by reasons of becoming TSM



Note: *Understanding Society*, Wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines indicate the value in the TSM sample. The variable "Education: A level'" includes higher education qualifications which cannot be considered university degrees. The variable "Education: GCSE" includes any other qualification that cannot be considered an A-level or a higher education qualification.

Figure 7: TSM Country of birth, by reasons of becoming TSM



Note: *Understanding Society*, wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines indicate the value in the TSM sample. The variable "Birthplace: Europe" excludes Ireland, the variable "Birthplace: Asia" includes both China and South Asia.

Case study: Respondents experiencing partnership dissolution.

This subsection provides a case study of a particularly interesting group: respondents experiencing partnership dissolution. We document how many partnership dissolutions are observed in the current sample, and many partners can be studied with the current following rules and this might change if the following rules were broadened. This exercise is relevant for at least two reasons. First, it quantifies the sample size available for researchers who want to investigate the consequences of partnership dissolution. This is a policy relevant question, and also a research question, longitudinal in nature, that it is best-answered using panel data. Second, by definition, the sample size available to study partnership dissolution is affected by the following rules: partnership dissolution often coincides with cohabitation dissolution, and TSMs partners are not followed when they stop co-habiting with OSMs.

For each wave, the left panel of Figure 8 shows: the number of partnerships formed (dashed line), the number of partnership dissolutions (dotted line), the total number of partnerships (solid black line), and the share of partnership dissolutions over all partnerships (solid grey line). Note that we can generally only observe those partnerships dissolutions after which at least one former partner remains in the panel. Partnerships dissolutions ending with both former partners leaving the sample cannot be observed, as no partner is left in the survey to report that the partnership is not anymore in place.

The number of partnerships forming and splitting is roughly constant: around 400/500 partnerships form/dissolve in each wave, with formations slightly exceeding dissolutions, especially in earlier waves. The total number of partnerships goes from almost 18000 in wave one to just over 10000 in later waves, with a descending trend due to attrition, and a small peak at wave six due to the inclusion of the IEMB sample component. As the total number of partnerships decreases, while the number of dissolutions remains constant, the share of partnership dissolutions over all partnerships increases from just below two percent at Wave two, to around three percent at Waves seven and eight.

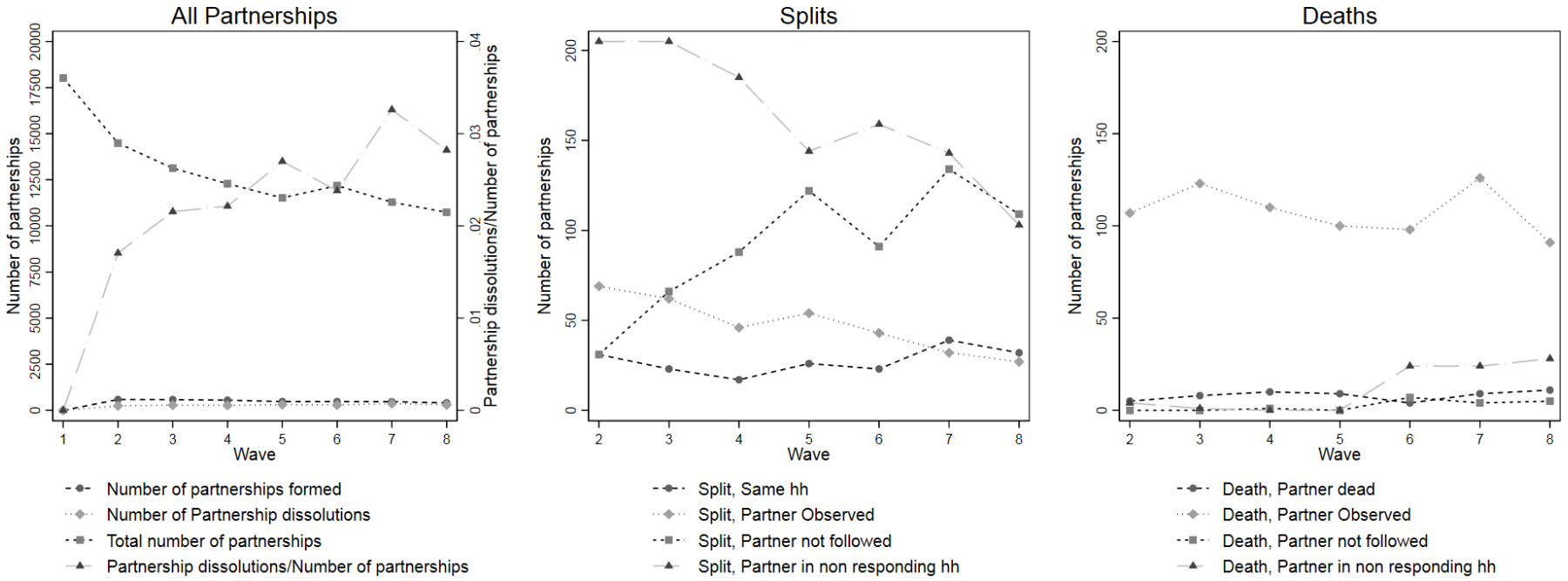
The middle panel of Figure 8 looks at partnership dissolutions when both partners remain alive. We call these cases “splits”. Cases where the two former partners remain in the same household after partnership dissolution are few and constant in number (black dashed line, around 30 to 40 cases per wave). Cases where the survey attempts to follow the former partner leaving the household and this attempt is successful (the former partner is observed) exceed 50 units at wave two, but decrease over time (dotted line). Cases where the survey attempts to follow the former partner but this attempt is not successful (the former partner ends up in a non-responding household) are three/four times more frequent, and they also decrease over time (grey line). Finally, cases where the survey does not attempt to follow the leaving former partner start low (less than 50 cases) at wave two, but then increase sharply over time (black line). As TSMs are not followed after a split,

while OSMs are, the decrease over time in the followed cases and the increase in the non-followed cases reflect the increase in the TSM to OSM ratio shown in Section 3.1.

The right panel of Figure 8 looks at partnership dissolutions due to death. For all categories the trend over time looks generally flat. The only two exceptions to this are i) the peak in number of cases with an observed surviving partner (dotted line) at wave seven, reflecting the increase in sample size due to the inclusion of the IEMB sample component at wave six, and ii) the increase in the number of cases where the surviving partner ends up in a non-responding household, possibly due to the change of the survey agency at wave six. Among partnerships due to death, the cases where the partner is observed after the split are by far the most frequent (above 100 cases in every wave); the other cases are all well below 50 units.

Figure 8 provides several indications regarding whether switching to more inclusive following rules can be beneficial for the study of the effects of partnership dissolution. First, partnership dissolutions are relatively rare and decrease over time. Thus, one may want to increase the available sample size of people going through partnership dissolution by increasing the number of people followed. Second, only in few cases we can successfully follow both surviving partners after the dissolution of their partnership; in most cases we can only follow one. In particular, the number of cases where the survey does not even attempt to follow one surviving partner is not negligible and increases over time both in absolute value and as a share of total dissolutions. This suggests that changing the following rules, such that the survey attempts to follow more people after partnership dissolution, may increase the number of cases where both surviving partners are successfully followed. This may be useful for researchers wanting to study the effects of partnership dissolution on both surviving partners. As TSMs and OSMs have been found to be different, following TSMs may help investigate heterogeneous effects of partnership dissolution. For example, following TSMs may help investigate the effects of partnership dissolution on recent migrants in a former partnership with a native. For these people, partnership dissolution may have substantive legal and psychological consequences, as it may lead to significant reductions in the rights to stay in the country and the size of the social network. Finally, in most of the cases where the partnership ends due to death, the surviving partner is followed. Cases where the surviving partner is not fielded are a minority: OSMs are older than TSMs and are likely to stay longer in the sample, thus OSM-only couples are over-represented among partnerships ending due to death. Therefore, a change in the following rules would have a limited effect on the sample size available for the study of the effects of a partner's dead on the surviving partner.

Figure 8: Partnership dissolution



Note: *Understanding Society*, waves 1 to 8. Information about partnerships is derived from the household grid. Excludes people entered as new-borns and from the BHPS sample component

Do TSMs provide good quality data?

This subsection analyses the quality of the data provided by TSMs. Extending the types of TSMs followed can be a valuable opportunity for a survey. However, the potential benefits will only be realised only if TSMs provide good quality data. We select a number of indicators that can track the entire process of obtaining good quality data. First, we look at measures of effort the survey agency had to exert to establish contact in the respondent's household (Figure 9). We look at three indicators of such effort: the total number of calls made to the household per capita, the total number of issues per capita, and the total call length per capita. These measure are collected at the household level, therefore they can only indirectly shed light on the impact of TSMs on data quality. Second, we look at the individual interview outcome once the household has been contacted, that is, if the respondent gives a full interview, a proxy interview, a refusal or other forms of non-response (Figure 10). Third, we look at measures of data quality when an interview is given (Figures 11, 12 and 13). We look at the prevalence of item non-response in variables on: birthplace, education, ethnicity, and the main UKHLS individual income measures (earnings, investment income, total income from benefits). We also look at indicators of responding style recorded by the interviewer, namely whether the interviewer recorded: “very good” cooperation, that the respondent was “not at all suspicious” and that the respondent's understanding was “excellent” (Figure 14).

Figure 9 shows indicators of the effort the survey agency had to exert to contact the household. The measures are given per-capita, and are broken down by type of TSM. For comparison, we also report the corresponding measures computed for the total sample of TSMs (grey dashed line), the total sample of OSMs (black dashed line) and the sample of OSMs who live with at least one TSM (black dotted line). TSMs are not associated with higher effort exerted by the survey agency to contact the household. In fact, with the exception of partners, TSMs are generally associated with fewer calls and issue per capita, and shorter average call length than TSMs.

Figure 10 shows that all categories of TSMs are less likely to give a full interview and more likely to give a proxy interview or a refusal than OSMs. The high incidence of proxy interviews among TSMs is driven by family members: partners, people previously at university, and offspring and relatives. House sharers drive the high incidence of refusals. This suggests OSMs may be likely to give proxy interviews for their family members, but less so for unrelated house sharers.

Figure 11 show share of missingness on variables measuring: marital status, country of birth, ethnicity and education. Marital status, country of birth, ethnicity can be non-missing across all response outcomes analysed in figure 7. Education can only be non-missing in case of a full or proxy interview (it is always missing in case of refusals and other type of non-response). Rates of missingness for marital status for TSMs are quite low, in line with those for OSMs. This is probably due to the fact that marital status is easily known and can also be reported by somebody else in a proxy or household interview. Apart from partners, most TSM categories have higher non-response

than OSMs in variables measuring other less observable aspects, such as country of birth, ethnicity and education (see top right panel and bottom panels). Particularly high rates of non-response for these variables are found among residents in the previous wave, offspring and relatives, and -above all- house sharers.

Figure 12 shows nonresponse for crucial income variables: gross monthly earnings for employees and gross monthly earnings from self-employment. The left column shows the share of missing data over all enumerated individuals. Therefore, all people who do not give a full interview are counted as having missing data on income.[5] The middle column shows the share of missing data over those who give a full interview. Finally, the right panel shows the share of missing data for the applicable cases. This means the share of gross monthly earnings (earnings from self-employment) is only computed for employees (self-employed).[6]

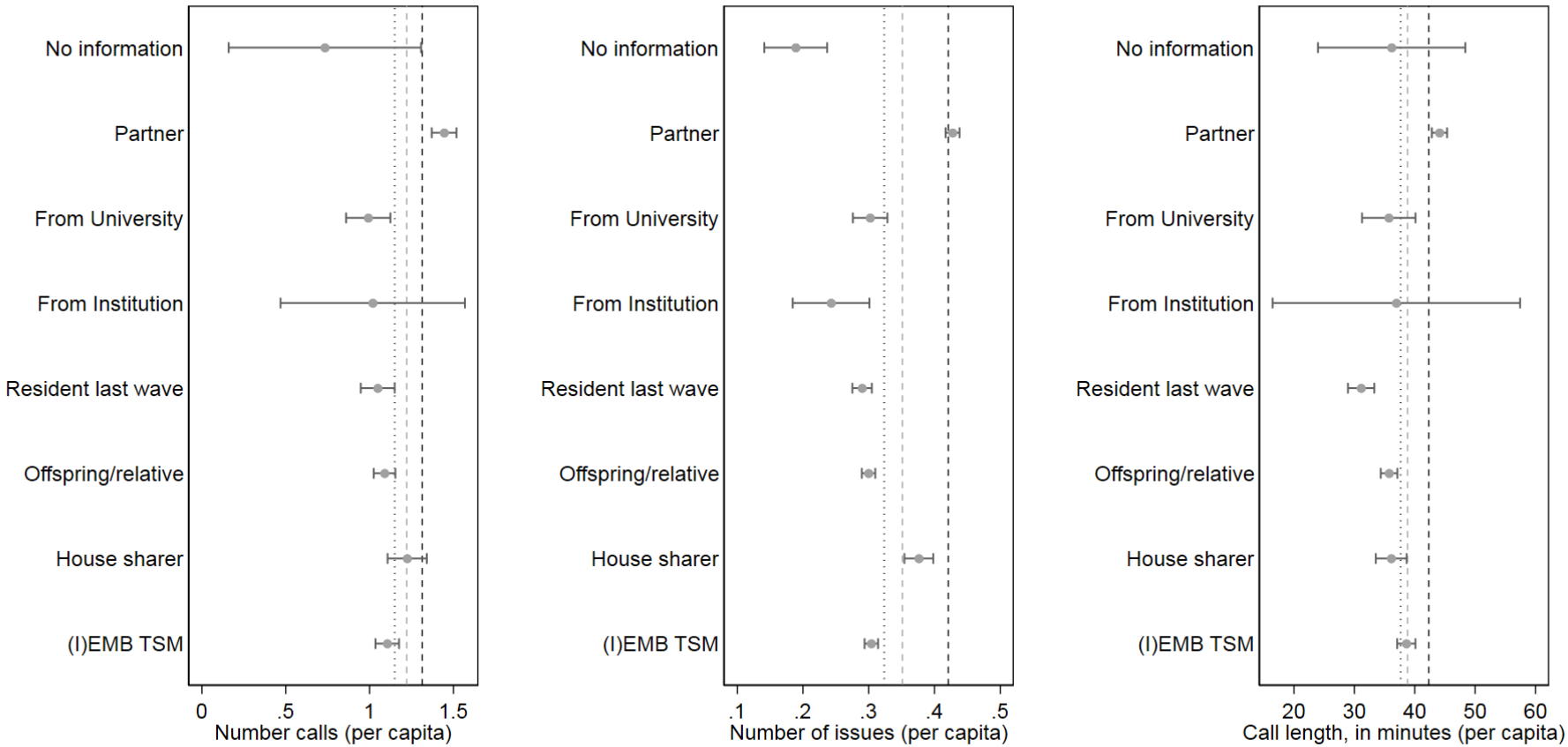
Figure 13 shows nonresponse for incomes from saving and investment and other incomes, including income from benefits and transfers. Again, the left column shows the share of missing data over all the enumerated individuals; the right panel shows the share of missing data for those who give an interview (who, for these income variables, are also the applicable ones).

The left panels of Figures 12 and 13 show that missing income is more prevalent among TSMs when all interview outcomes are considered. The missingness shown in the left panels of Figures 12 and 13 is likely to be driven largely by the interview outcome. However, the middle and right panels of figures 12 an 14 show that, when the analysis is restricted to full-interviews and to applicable cases, TSMs do not necessarily provide worse data than OSMs.

[5] Proxy interviews only ask for information on total earnings and income and record them in bands

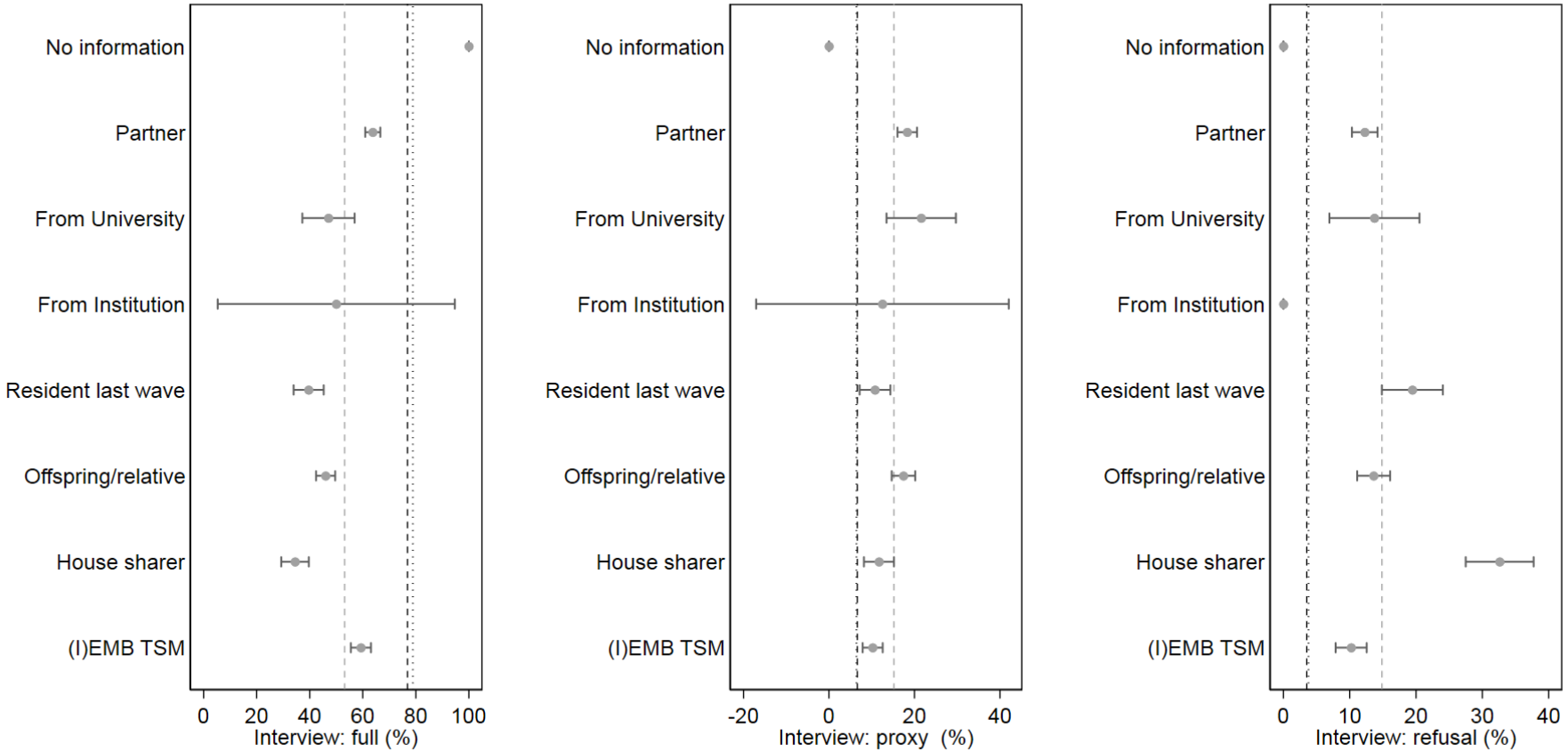
[6] Note that, in the case of investment income and other incomes, the middle and the right panel are the same, as all people giving a full interview are potentially applicable for these two categories of income

Figure 9. TSMs vs OSMs: Resources needed to contact the household



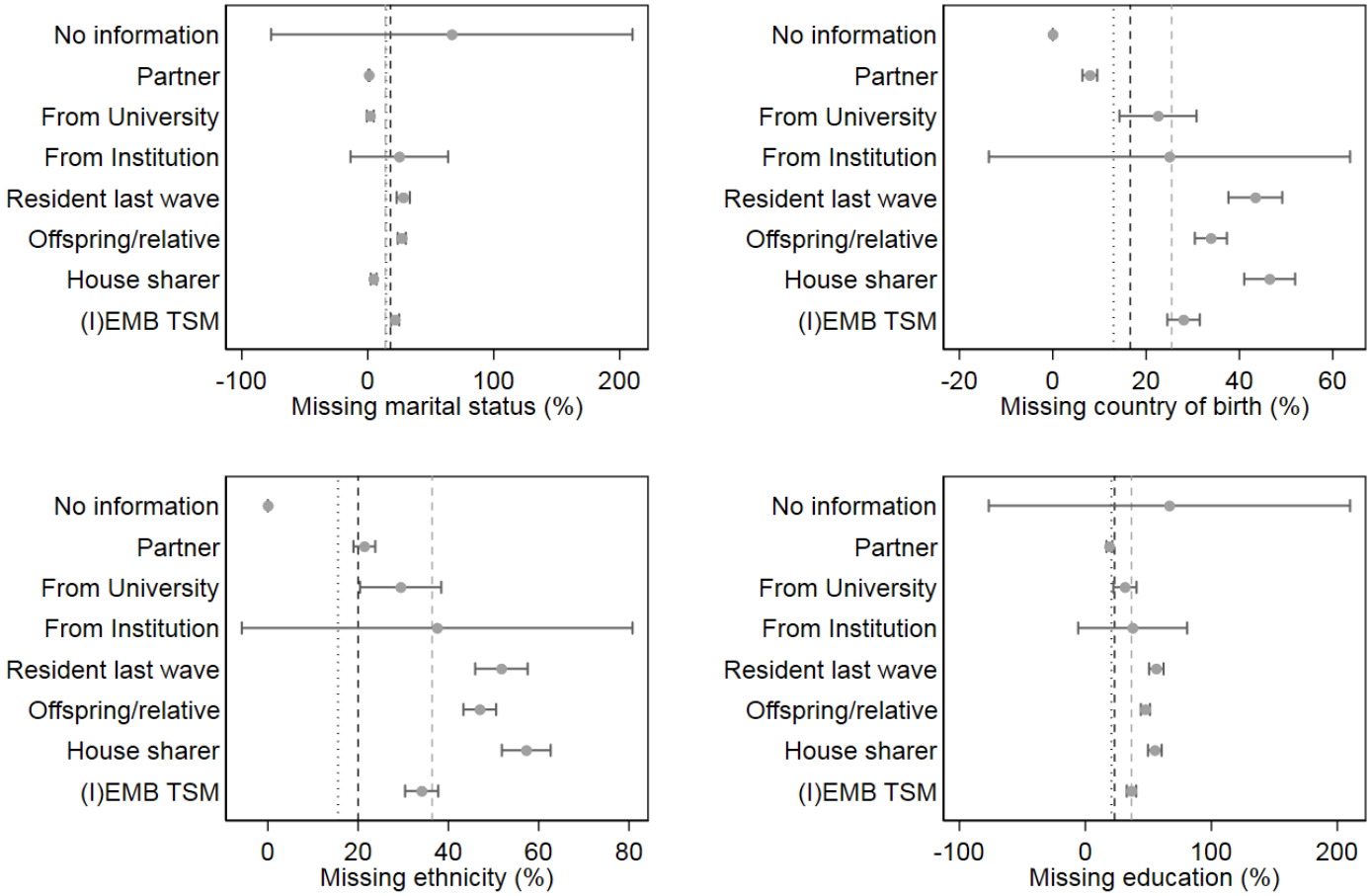
Note: *Understanding Society*, wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines indicate the value in the TSM sample. Measures per capital are obtained by dividing the values at the household level by the number of household members.

Figure 10. TSMs vs OSMs: Interview outcome



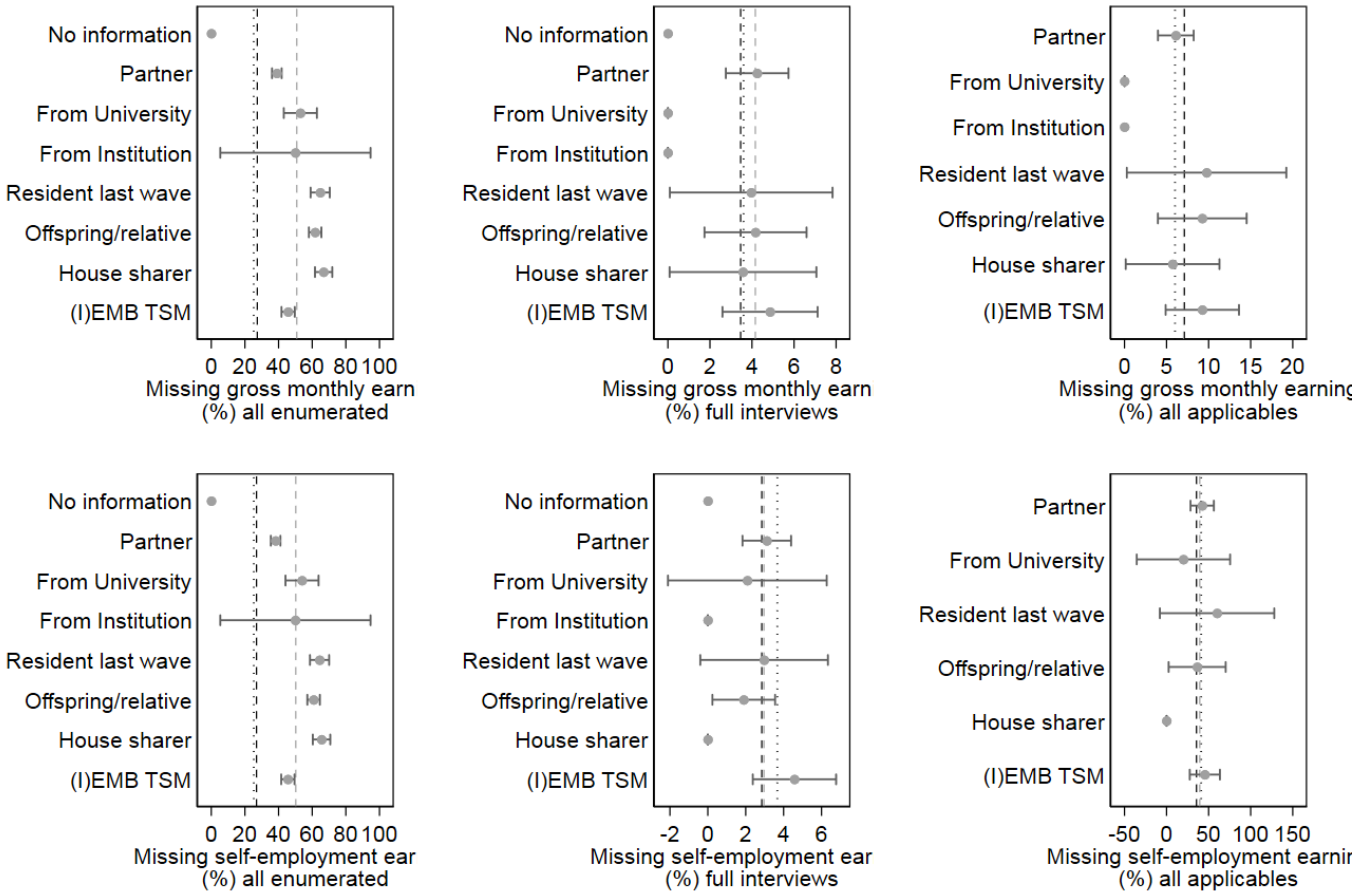
Note: *Understanding Society*, wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines indicate the value in the TSM sample.

Figure 11. TSMs vs OSMs: Item non-response on demographics



Note: *Understanding Society*, wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines indicate the value in the TSM sample.

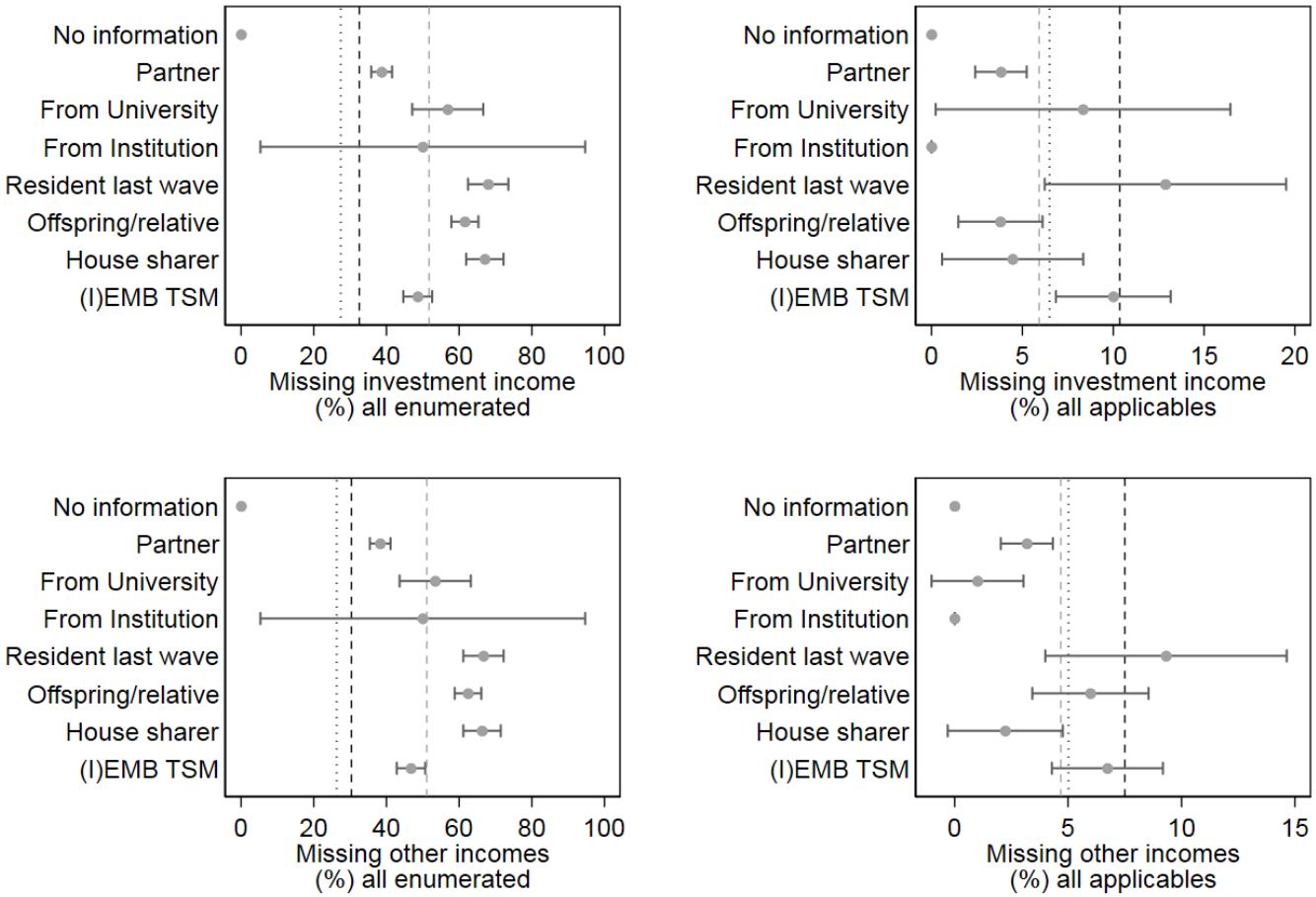
Figure 12. TSMs vs OSMs: Item non-response on income data



Note: *Understanding Society*, wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines

indicate the value in the TSM sample. Applicables for gross monthly earnings (first row, right panel) are employees; applicables for self-employed earnings (second row, right panel) are self-employed.

Figure 13. TSMs vs OSMs: Item non-response on income data



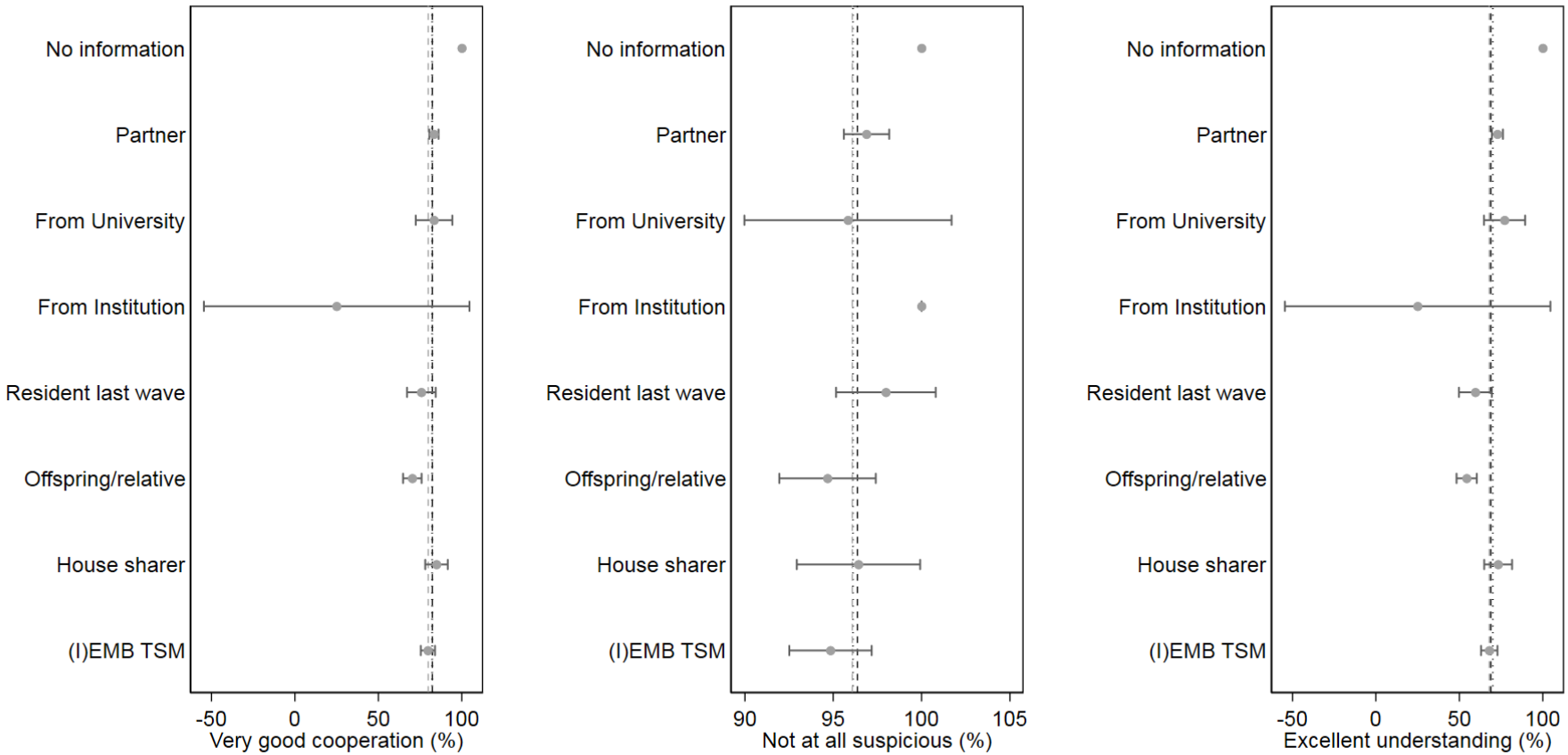
Note: *Understanding Society*, wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines

indicate the value in the TSM sample. Applicables for investment income and other incomes (bottom two rows, right panel) are everybody who gives a full interview.

Finally, Figure 10 shows data on individuals' responding style, that is, if respondents show “very good cooperation” if they look “not at all suspicious” and if they show “excellent understanding”. These data are collected by the interviewers about respondents providing a full interview. Again, when only respondents giving a full interview are considered, interviewers do not perceive TSMs to be worse respondents than TSMs.

In sum, this section suggests TSMs do not disproportionally increase the cost of contacting the household, and, once they give a full interview, they do not provide worse quality data than OSMs. However, TSMs are significantly less likely to give a full interview than OSMs. This is the main cause of missingness for data from TSMs.

Figure 14. TSMs vs OSMs: Responding style from interviewers' perspective (full interviews only)



Note: *Understanding Society*, wave 5, excluding people entered as new-borns and from the BHPS sample component. 95% confidence intervals. Black dashed lines indicate the value in the OSM sample, black dotted lines indicate the value in the OSM sample living with at least one TSM; grey dashed lines indicate the value in the TSM sample.

Conclusions and recommendations

Our analysis shows that TSMs provide useful sample on several policy relevant sub-populations, such as house sharers, new migrants from the European Union, or young people returning to the parental home. We also show that, with the current following rules, the survey loses a significant number of the respondents going through partnership dissolution.

Regarding the quality of the data provided by TSMs, we show that the presence of TSMs is not associated to more effort required to the survey agency to contact the household, and the quality of the data from TSMs giving a full interview is as high as the one provided by OSMs. However, the rates of refusals and proxy interviews are high among TSMs. High refusal rates are driven by house sharers, high rates of proxy interviews are driven by OSMs' family members, such as partners or offspring.

Together these findings suggest that broadening following rules can be a useful complement to boost samples. It is important to note, though, that broadening the following rules cannot entirely eliminate the need to boost longitudinal studies. The inclusion probabilities of additional joiners are difficult or impossible to calculate, and moreover, while additional joiners provide good coverage of some subpopulations of interest, they provide poor coverage of others.

Several specific recommendations for *Understanding Society* follow from our analysis.

First, *Understanding Society* should consider following TSMs when they leave the household. This should improve the cross-sectional representativeness of the study and increase the sample size available for the study of small, policy-relevant populations. Moreover, following TSMs when they leave the household can facilitate research on questions that are longitudinal in nature, such as the consequences of partnership dissolution. In particular, more inclusive following rules would allow the study the consequences of partnership dissolution on TSM partners, which is now impossible. This would increase both sample size and survey representativeness (TSMs are different from OSMs and thus the effects of partnership dissolution may differ between OSMs and TSMs) and will enable the analysis of heterogeneous effects.

Second, efforts should be made to raise the share of TSMs giving a full interview. For the case of house sharers, this should be achieved through a reduction in refusals; for the case of partners and other relatives of OSMs, this should be achieved through a reduction in proxy interviews. As our analysis has uncovered for the main characteristics of TSMs, these findings can be used to design tailored strategies to increase the probability that these groups give a full interview (see: Fumagalli et al., 2013).

Third, to increase the amount of information available on TSMs, *Understanding Society* may want to add retrospective questions for TSMs such that the quality of retrospective information from TSMs matches the quality of retrospective information currently available for OSMs. This retrospective information could also be used to compute inclusion probabilities -and thus weights- for TSMs.

Fourth, as TSMs data capture under-researched populations, we recommend that a broadening of the sample rules be accompanied by new questions to collect information on the most policy-relevant of these populations. For example, questions can be added to understand the reasons of the rise in cohabitations among adults, and the rise of the “sandwich generations”, that is young people returning to the parental home after living for a period on their own.

Fifth, between-interviews data collection might be used to avoid respondents dropping out in case of partnership dissolution. If partnerships dissolutions are detected as soon as they occur, specific strategies can be put in place to keep at least one of the former partners in the survey.

References

Fadlon, I. and T. H. Nielsen (2019). Family health behaviors. *American Economic Review*, forthcoming.

Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998). An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources,* 33(2), 251–299.

Fumagalli, Laura, Heather Laurie, and Peter Lynn. "Experiments with methods to reduce attrition in longitudinal surveys." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176, no. 2 (2013): 499-519

Green, S. and L. McCarthy (2015). Is sharing the solution?: exploring the opportunities and challenges of privately rented shared accommodation for single people in housing need. *People, place and policy* 9(3), 159–178.

Schonlau, M., M. Kroh, N. Watson, et al. (2013). The implementation of cross-sectional weights in household panel surveys. *Statistics Surveys* 7, 37–57.

Schonlau, M., N. Watson, and M. Kroh (2010). Household survey panels: how much do following rules affect sample size? SOEP papers on Multidisciplinary Panel Data Research 347, DIW Berlin, The German Socio-Economic Panel (SOEP)

Tosi, M. and E. Grundy (2018). Returns home by children and changes in parents' well-being in Europe. *Social Science & Medicine* 200, 99–106.

University of Essex. Institute for Social and Economic Research, NatCen Social Research, Kantar Public. (2018). *Understanding Society: Waves 1-8, 2009-2017 and Harmonised BHPS: Waves 1-18, 1991-2009*. [data collection]. *11th Edition.* UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-12

Appendix A. Details of Understanding Society following rules

Understanding Society is a representative sample of households living in the United Kingdom in 2009, when wave one was collected. It is a longitudinal study and thus household members included in the initial sample are followed over time indefinitely. Individuals joining the sample after the first wave are also followed over time, but whether they are followed indefinitely depends on their sample status. Following rules determine how the sample status is assigned to each household member. *Understanding Society* has three possible sample statuses: Original Sample Members, Temporary Sample Members, and Permanent Sample members. OSMs and PSMs are followed indefinitely, TSMs are followed only if they live with an OSM.

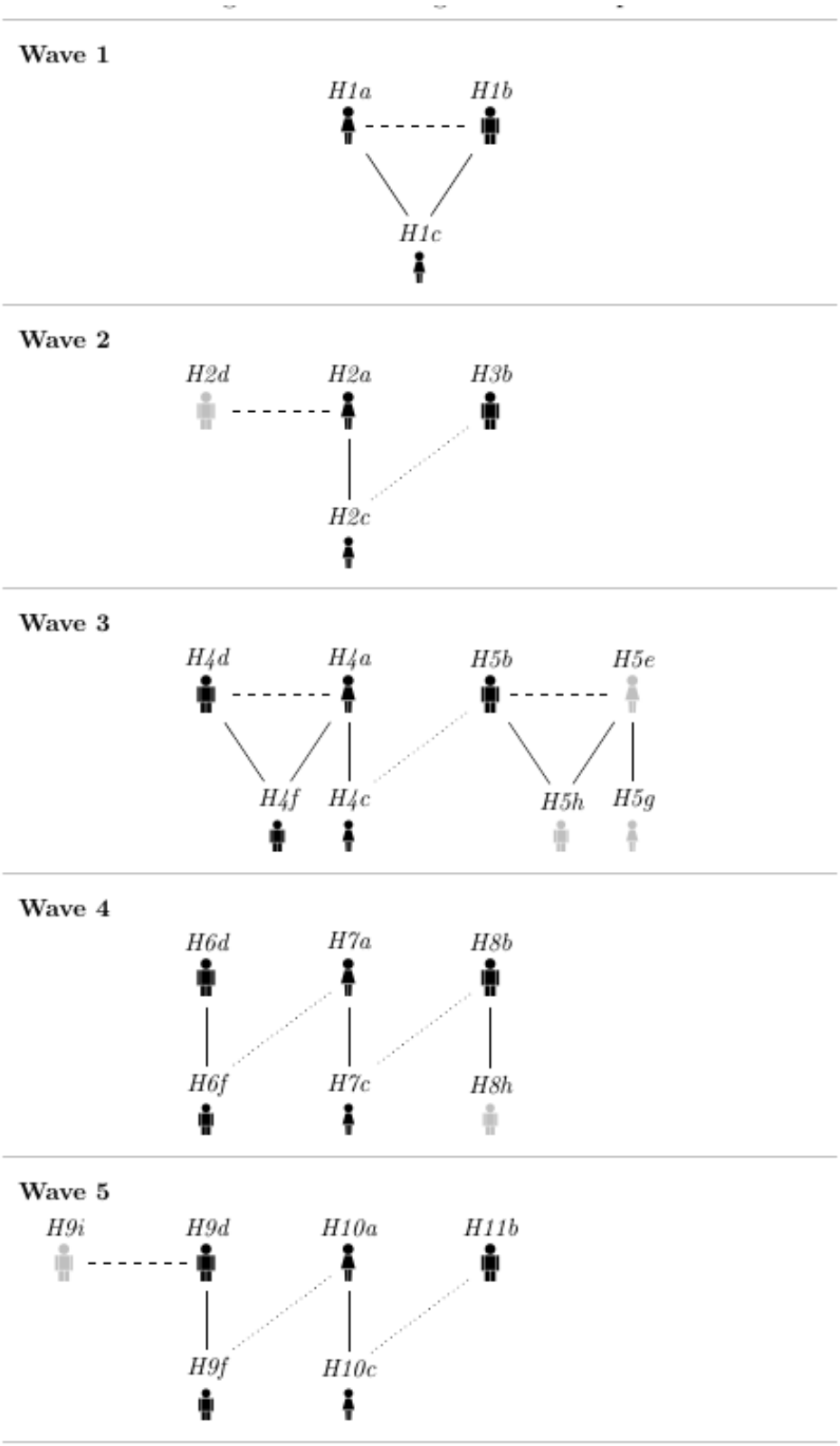
Are considered Original Sample Members (OSMs): i) individuals found at selected households at wave one, individuals temporarily absent but otherwise resident; ii) individuals in households eligible for the inclusion in Ethnic Minority Boost Sample (EMBS) or the Immigrants and Ethnic Minorities Boost Sample (IEMBS), and from a target ethnic minority; iii) children born to a female OSM and living with the OSM mother at the wave following their birth; iv) individuals from the British Household Panel Survey (BHPS) sample component who were enumerated at the first wave of the BHPS sample they belong to (1991 BHPS sample, Scotland, Wales, or Northern Ireland boost samples); v) children born to a BHPS OSM (irrespective of the gender) before the inclusion of BHPS in *Understanding Society*. OSM respondents are followed as long as they live in the UK. If they leave, they are not followed abroad, but they remain eligible for interview in the case they come back to the UK.

Are considered Temporary Sample Members (TSMs): i) Individuals joining the household of an OSM after the sample selection/first interview, such as partners, house sharers or family members joining the household; ii) individuals in households eligible for the inclusion in EMBS, but not from a target ethnic minority; iii) individuals in households eligible for the inclusion in IEMBS, but not from a target ethnic minority or immigration background; iv) children born to male OSM and a female TSM and co-resident with the father (or any other OSM) at the survey wave following the child's birth. TSM are followed as long as they live in the household of an OSM.

Are considered Permanent Sample members (PSMs): TSMs who do no longer live with OSMs, but they are still followed as they can bring to the panel crucial information about the OSMs. Currently, the only PSMs are TSMs who are fathers of an OSM child and are observed to live with this child the wave after their birth.

Figure A1 provides the example of a fictitious household *H1* sampled at wave one. *H1, H2, ..., H11* indicate households, *a*, *b*, ..., *i* indicate people, so that *H1a* is person *a* in household *H1*. Note that the person identifier is longitudinal, in contrast, there is no concept of longitudinal household, and thus household identifiers change every wave. Black symbols are OSMs/PSMs (followed indefinitely); grey symbols are TSMs (followed as long as they live with an OSM). Black dotted lines indicate partnerships with cohabitations among adults; black solid lines indicate parent/offspring relationships with cohabitation; black dotted lines indicate parent/offspring relationships without cohabitation.

Figure A1: Following rules, example



Note: *H1, H2, ..., H11* indicate households, *a, b, ..., i* indicate people, so that *H1a* is person a in household *H1*. Black symbols are OSMs/PSMs; grey symbols are TSMs. Black dotted lines indicate partnerships with cohabitations among adults; black solid lines indicate parent/offspring relationships with cohabitation; black dotted lines indicate parent/offspring relationships without cohabitation.

Consider the household *H1* at wave one, composed by one woman (*a*), one man (*b*) and their daughter (*c*). *H1* was sampled at wave one, and thus all its components are OSMs, as indicated by their black colour in figure A1. At wave two, the couple splits. The woman page 33 rtners and cohabits with her new partner *d* and her own daughter *c* from the previous partnership with *b*. At wave two, *d* is a TSM (grey colour). At wave three, *a* and *d* have a child: *f*, who lives with them and *a*'s daughter *c*. Since *a* is a female OSM, *f* also becomes an OSM, and his father *d* becomes a PSM. At wave three, *b* also re-partners with *e*, who has a child (*g*) from a previous partnership we do not observe. Both *e* and *g* are TSMs. Household *H5* is also composed by *h*: the son of *b* and *e*. Unlike *f*, *h* is not given the OSM status, as he is born from an OSM father and a TSM mother. At wave four, both households split. We keep following *d*, *a, f, c*, as they are all OSMs; moreover, we follow *b* (OSM) and *h* (TSM), who lives with him. Note that we lose *e* and *g*, because they do not live anymore with OSMs. Finally, at wave five, we lose *h*, as he stops living with his OSM father *b*. We also start collecting data on *i*, *d*'s new partner, but we will follow him only as long as he lives with an OSM.

Figure A1 give an idea of the information we can obtain given the current following rules and the information we could have by changing them. For example, the current rules permit to collect extensive information on the children *f* and *c*, but limited information on *h*, and almost no information on *g*. Switching to very inclusive following rules prescribing to follow indefinitely everybody who joined the household at some point will make sure we do not lose *h* and *g* at wave five and four, respectively. One may also want to collect information on what happened to *d, i, e, g* before they join the sample. This information has to be retrieved through retrospective modules whose length and characteristics may also vary by sample status.

With the exception of new migrants, TSMs are sampled from the same population as OSMs: the target population of at the initial wave. The difference between OSMs and TSMs is simply that the selection probabilities of OSMs are determined solely by the sample design, while those of TSMs are determined additionally by frequency of household change and size of households they join. Therefore, OSMs and TSMs are assigned different weights.

Weights (both cross-sectional and longitudinal) of OSMs who were included in the original sample are computed based on individual selection probabilities (plus non-response adjustments). Offsprings of these OSMs or offsprings of a female OSM and a male TSM are OSMs themselves and will be followed indefinitely. However, as they did not exist at the time of the sampling, their individual sampling probability was not defined at the time of the sampling. Therefore, their weights are based on their mother's selection probability. Cross sectional weights for TSMs are based on the selection probabilities and continuing response probabilities of the members of the household they join and they can be zero or non-zero. Longitudinal weights for TSMs are always equal to zero as they were not part of the original sample.

There are no differences in fieldwork practices by sample status, other than following rules. Questionnaire content is not linked to sample status either, apart from two exceptions. First, the retrospective questions on arguments like fertility and employment histories asked to TSMs in the new entrant module are not as detailed as those asked to OSMs. Second, some of the extra questions asked to ethnic minorities in the EMB and IEMB differ by sample status. The reason for this is that in the EMB/IEMB sample status is assigned also on the basis of ethnicity. For example, non-ethnic minority individuals in households eligible for the EMB or the IEMB are considered TSMs even if they are part of the household when the household was originally included in the sample. These TSMs are not asked the extra questions on ethnicity. This exclusion, however, does not derive from the fact that these individuals are TSMs, but from the fact that

these individuals are not ethnic minorities, and thus asking them the extra questions for ethnic minorities would not be meaningful.