



Research Institute
Social Cohesion

RDC

RDC-RISC Papers

Research Data Center of the Research Institute Social Cohesion

RDC-RISC Paper 2: Survey Weights for the German Social Cohesion Panel 2021-2023 - Wave 1-3

Julian B. Axenfeld, Hans Walter Steinhauer



Research Data Center of the Research Institute Social Cohesion

The Research Data Center of the Research Institute Social Cohesion (RDC-RISC) supports the scientific community by establishing a portal to available data relevant for empirical analysis on issues of social cohesion. This data portal includes data collected within the Research Institute Social Cohesion (RISC) as well as external data. The RDC-RISC also fulfills central tasks of documenting, archiving and providing the quantitative data collected at the RISC for secondary scientific use.

The aim of the RDC-RISC Papers is to document specific aspects of the RISC's data infrastructures.

Recommended Citation

Axenfeld, J. B., & Steinhauer, H. W. (2026). Survey Weights for the German Social Cohesion Panel 2021-2023 - Wave 1-3. *RDC-RISC Paper 2*. Bremen: RDC-RISC/SOCIUM. doi:10.60532/rdc-risc_paper_2

- **Authors:** Julian B. Axenfeld, Hans Walter Steinhauer

- **Publisher:** RDC-RISC
SOCIUM, University of Bremen
P.O. Box 330 440
28334 Bremen
Germany

- **DOI:** 10.60532/rdc-risc_paper_2

- **Website:** www.fgz-risc-data.de



The text of this publication is published under the Creative Commons license CC BY-SA 4.0 Attribution-ShareAlike 4.0 International. The exact wording of the license CC BY-SA 4.0 can be found here:

<https://creativecommons.org/licenses/by-sa/4.0/>

Survey Weights for the German Social Cohesion Panel 2021-2023 – Wave 1-3

Julian B. Axenfeld**

Hans Walter Steinhauer*

Summary

*This report describes the construction of survey weights for the Scientific Use File (SUF) of the third wave of the German Social Cohesion Panel (SCP). It covers the recruitment wave – Wave 1 Part 1 (W1P1) – and the first three follow-up waves: Wave 1 Part 2 (W1P2), Wave 2 (W2), and Wave 3 (W3). The weighting procedure includes the construction of design weights, the estimation of cross-sectional nonresponse weights for the recruitment wave, and the estimation of longitudinal inverse staying probabilities and cross-sectional weights for the first three follow-up waves at both person and household level. These weights are included in the *phrf* and *hhrf* datasets of the SCP 2021–23 W1–3 SUF. The report first describes the adjustment of design weights, followed by the construction of nonresponse weights for each wave. It then explains how the nonresponse models were estimated and summarizes their empirical results. A further chapter examines how the distribution of cross-sectional weights evolves across the stages of the weighting process in each wave and compares respondents with high nonresponse weights to the full realized sample. The report closes with a discussion of practical implications and limitations.*

1. Design weights

The German Social Cohesion Panel (see Groh-Samberg et al., 2026 for the data, and Gerlitz et al., 2024, for an overview) is an annual mixed-mode (online and paper) panel survey based on a random sample of 37,874 anchor persons (APs) aged 18 years or older who resided in private households in Germany at the time of recruitment (wave 1 part 1; hereafter W1P1). Individuals living in Eastern Germany and Berlin were oversampled by factors of 2.5 and 1.4, respectively.

Initial design weights for the anchor person sample were provided by the survey institute infas (see Steinwede et al., 2024, for details on the sampling design and the calculation of the design weights). These weights are available in the *phrf* dataset as the variable *DESIGN_AP*.

*German Institute for Economic Research (DIW Berlin)

**Research Institute Social Cohesion (RISC)

To incorporate the household context into the sample, participating APs were asked to list all other household members aged 18 years or older (up to a maximum of 12). These household members (HMs) were subsequently invited to participate in the survey. This recruitment procedure results in unequal inclusion probabilities across households of different sizes because every adult household member could, in principle, have been selected into the initial random sample. Consequently, households with more adult members had a higher probability of being included in the sample.

To obtain design weights reflecting the inverse selection probabilities in the combined sample of APs and HMs (variable *DESIGN* in the *phrf* dataset), the initial *DESIGN_AP* weights were adjusted as follows:

$$DESIGN = \frac{1}{1 - (1 - DESIGN_AP^{-1})^{\min(SIZE_{W1P1}, 13)}} \quad (1)$$

Here, $SIZE_{W1P1}$ denotes the number of individuals aged 18 or older living in the household at W1P1, which corresponds closely to the variable *HGHHAGE03* in the *hgen* dataset¹. As a result, members of larger households receive smaller design weights than members of smaller households because their probability of selection is higher. Table 1 illustrates the effect of this adjustment on the design weights of survey participants.

Table 1: Design weights of W1P1 survey participants

	Minimum	1%	25%	Median	75%	99%	Maximum
<i>DESIGN_AP</i>	905.885	905.885	905.885	2297.778	2297.778	2297.778	2297.778
<i>DESIGN</i>	70.146	226.847	459.956	1149.139	1149.139	2297.778	2297.778

Note: APs' design weights as initially provided by infas (*DESIGN_AP*, n=13,053) and design weights adjusted for household size (*DESIGN*, n=17,027).

As the design weight variables reflect the initial sampling design, they remain constant across waves.

2. Construction of nonresponse weights

Whereas design weights account for unequal selection probabilities arising from the sampling design, nonresponse weights additionally adjust for selective survey nonresponse. Their

¹ There are some differences between the variable used here and the variable *HGHHAGE03*. Most importantly, the weighting variable includes imputations by last-observation-carried-forward if household size data is missing or incomplete in a wave, while *HGHHAGE03* uses only information from the present wave.

purpose is to reduce nonresponse bias and improve inference to the survey's target population — in the case of the SCP, adults living in private households in Germany.

The SCP weighting procedure follows the general approach established in the Socio-Economic Panel (SOEP), one of the world's longest-running household panel studies. As in the SOEP (see Kroh et al., 2015, for a description of the SOEP weighting scheme), the final nonresponse weights are obtained by combining inverse selection probabilities (design weights), predicted response probabilities from nonresponse models, and a raking adjustment based on the marginal distributions of key sociodemographic characteristics of the target population. For each survey wave, first, the SCP provides both cross-sectional household-level weights (*HHRF*). Second, the SCP supplies person-level weights (*PHRF*) for inferences to private households and the adult population living in private households in Germany, respectively. In addition, person-level nonresponse weights for the initial AP sample (*PHRF_AP*) are provided to support inference based solely on the originally drawn probability sample of APs. This sample does not suffer from the additional stages of nonresponse at play for additional HMs laid out below. For the follow-up waves, longitudinal inverse staying probabilities are provided at both the household level (*HBLEIB*) and the person level (*PBLEIB*) in addition to the cross-sectional weights.

2.1 Wave 1 Part 1 nonresponse weights

Modeling response probabilities

In the SCP recruitment wave (W1P1), nonresponse could occur at three stages:

1. The initially selected anchor person (AP) may or may not participate in the survey. Participation is defined as answering at least one survey question, which is also the criterion for being included as an observation in the person (*p*) or household (*h*) data. If an AP failed to respond in W1P1, their whole household did not participate in the SCP. In W1P1, approximately 34% of invited APs participated in the survey. The estimated probability of APs participation is denoted by $\hat{p}_1$.
2. Participating APs may or may not name all eligible household members for invitation. Household members not listed by the AP could not be invited to the survey. This is not necessarily a binary decision, as APs could list some but not all eligible household members. APs reported the total number of household members and their birth years separately in the beginning of the questionnaire (variables *HHHAGE01* and *HPNAME01–HPNAME12* in the *h* dataset; see also variable *HGHHAGE03* in the *hgen* dataset), which made it possible to determine the number of eligible household

members. Overall, 45% of all adult household members in participating households were successfully listed for invitation and could be invited to the survey. The estimated probability that an adult household member is listed for invitation is denoted by $\hat{p}_2$.

3. Among invited HMs, 67% participated in the survey. The estimated probability of HM participation given invitation is denoted by $\hat{p}_3$.

Each stage was modeled separately in the SCP using generalized linear models. The composition of combined response probabilities $\hat{p}_{W1P1}$ for persons living in sampled households is therefore different for APs and HMs:

$$\hat{p}_{W1P1} = \begin{cases} \hat{p}_1 & \text{for APs} \\ \hat{p}_1 \times \hat{p}_2 \times \hat{p}_3 & \text{for HMs} \end{cases} \quad (2)$$

Details on the exact construction of weights are provided in the following subsections.

Trimming

Variation in selection probabilities due to the survey design, combined with the three stages of nonresponse in the SCP, can produce considerable variation in nonresponse weights as well. This variation in survey weights, however, creates problems for estimation: as a general rule, more variable weights, while potentially reducing bias, also lower the efficiency of an estimate, leading to larger standard errors. In extreme cases, a few observations with extreme outlier weights can gain disproportionate leverage over the results. To guard against this, extreme weights in the SCP were trimmed. Trimming exchanges a small potential increase in bias for a gain in efficiency, aiming for a lower overall mean squared error (Haziza & Beaumont, 2017). As a rule of thumb, final weights were not to exceed ten times the median weight within the stratum with the highest design weight (2,297.778; one-person households in Western Germany). Because raking was applied after trimming (see below) and reintroduces some variation, this target could not always be met in practice.

Trimming was done both at the household and individual level. In the following, let t_1 denote the according adjustment operations applied to the weights on the household level HHRF, t_2 the adjustment operation on the person level to generate the *PHRF* weights, and t_3 the adjustment operation on the person level for the *PHRF*_AP weights.

Raking

After trimming, weights were also adjusted such that they fit known marginal distributions on certain variables. This helps compensating for additional selective nonresponse as well as sampling and coverage errors not captured through the nonresponse models.

On the household level, these variables were:

- (a) The number of persons living in the household. Households of five and more persons were grouped into a single category.
- (b) Population size of the municipality. This variable had four categories: less than 20,000 inhabitants, 20,000-99,999 inhabitants, 100,000-499,999 inhabitants, and 500,000 or more inhabitants.
- (c) Federal state. Several smaller states were grouped into joint categories with other states. Grouped states were: (1) Berlin, Brandenburg, and Mecklenburg-Vorpommern; (2) Hamburg, Schleswig-Holstein, Bremen, and Lower Saxony; (3) Hesse, Rhineland-Palatinate, and Saarland; and (4) Sachsen-Anhalt and Thuringia. All other states made up their own categories, respectively.

On the individual level, raking variables were:

- (a) Geographic region. Categories were Eastern Germany including Berlin, and Western Germany.
- (b) German citizenship, differentiating persons with and without German citizenship.
- (c) Age groups and gender. Age categories were 18-29, 30-39, 40-49, 50-59, 60-69, and 70 or more years old, each for male and non-male respondents.
- (d) Highest school-leaving certificate. Categories were low education (incl. no school degree, currently in school, Hauptschule degree or equivalent), medium education (Mittlere Reife/polytechnical secondary school degree after 10 years in the GDR or equivalent), and high education (Fachhochschulreife/advanced technical certificate or Allgemeine Hochschulreife/A-level certificate).

Marginal distributions were obtained from Microcensus estimates by the German Federal Statistical Office for the year 2021. A small number of missing values in the SCP raking variables were imputed deterministically.

Assembling Wave 1 Part 1 nonresponse weights

In the following, let r_1 denote the according adjustment operation applied to the weights on the household level. Let r_2 and r_3 stand for the adjustment operations on the individual level for the *PHRF* and *PHRF_AP* variables, respectively.

Household-level inclusion depended on being selected into the sample and on the AP's probability to respond. Therefore, household-level nonresponse weights $HHRF_{W1P1}$ are composed of the design weight multiplied by the inverse of the estimated AP response probability, with subsequent household-level trimming and raking:

$$HHRF_{W1P1} = DESIGN \times \hat{p}_1^{-1} \times t_1 \times r_1 \quad (3)$$

The calculation of person-level nonresponse weights $PHRF_{W1P1}$ starts from the $HHRF_{W1P1}$ household weight. For HMs, who have two additional stages of nonresponse, $HHRF_{W1P1}$ was multiplied by the inverse product of the estimated probability of being named for invitation and the estimated probability to respond if invited. Thereafter, individual-level trimming and raking were applied. This means individual-level weights are comprised as follows:

$$PHRF_{W1P1} = \begin{cases} HHRF_{W1P1} & \times t_2 \times r_2 & \text{for APs} \\ HHRF_{W1P1} \times (\hat{p}_2 \times \hat{p}_3)^{-1} & \times t_2 \times r_2 & \text{for HMs} \end{cases} \quad (4)$$

Anchor persons' nonresponse weights $PHRF_{AP_{W1P1}}$ are based on the AP design weight (*DESIGN_AP*) multiplied by the inverse of the estimated AP response probability, with subsequent person-level trimming and raking:

$$PHRF_{AP_{W1P1}} = \begin{cases} DESIGN_AP \times \hat{p}_1^{-1} \times t_3 \times r_3 & \text{for APs} \\ 0 & \text{for HMs} \end{cases} \quad (5)$$

2.2 Wave 1 Part 2 nonresponse weights

To generate weights for wave 1 part 2 (W1P2), the probability $\hat{p}_4$ of responding to W1P2 was estimated among all participants of the recruitment wave (W1P1). In the following, let $\hat{p}_4^{AP}$ denote the estimated response probability of a household's AP. As described for W1P1, extreme weights were again trimmed (as denoted by t_4 for the household level *HHRF* weights, t_5 for the person-level *PHRF* weights, and t_6 for the person-level *PHRF_AP* weights); and trimmed weights were raked thereafter (as denoted by r_4 for *HHRF*, r_5 for *PHRF*, and r_6 for *PHRF_AP*).

Household-level cross-sectional nonresponse weights ($HHRF_{W1P2}$) were created by multiplying the previous weight with the inverse of the AP's estimated response probability, followed by trimming and raking:

$$HHRF_{W1P2} = HHRF_{W1P1} \times \hat{p}_4^{AP-1} \times t_4 \times r_4 \quad (6)$$

Similarly, person-level $PHRF_{W1P2}$ and $PHRF_{W1P2}$ weights also rely on their respective previous weight multiplied by the inverse of the estimated response probability, followed by trimming and raking:

$$PHRF_{W1P2} = PHRF_{W1P1} \times \hat{p}_4^{-1} \times t_5 \times r_5 \quad (7)$$

$$PHRF_AP_{W1P2} = \begin{cases} PHRF_AP_{W1P1} \times \hat{p}_4^{-1} \times t_6 \times r_6 & \text{for APs} \\ 0 & \text{for HMs} \end{cases} \quad (8)$$

Note that while the AP role in a household can switch from one person to another across waves, the AP nonresponse weight $PHRF_AP$ refers to panelists' original role at recruitment. This is because $PHRF_AP$ represents the original AP sample drawn at the population registers. Consequently, APs who originally started as HMs received a zero $PHRF_AP$ weight, while HMs who started as APs but meanwhile passed their AP role to a fellow household member (for instance, because of temporary nonresponse) nevertheless received a $PHRF_AP$ weight if they still responded to the survey.

There are some cases in the SCP in which the AP failed to respond to a wave, while one or more HMs participated in the wave. Thus, in these households, while partial person-level data is available, there is no household-level data for this wave because only APs could have answered the household questionnaire module. In this case, participating HMs still received person-level weights, but their household was treated as not responding, therefore not receiving a household-level $HHRF$ weight.

Inverse staying probabilities $HBLEIB_{W1P2}$ and $PBLEIB_{W1P2}$ were constructed from the inverse of the estimated W1P2 response probability:

$$HBLEIB_{W1P2} = \hat{p}_4^{*-1} \quad (9)$$

$$PBLEIB_{W1P2} = \hat{p}_4^{-1} \quad (10)$$

2.3 Nonresponse weights from Wave 2 onwards

The generation of weights for the subsequent waves 2 (W2) and 3 (W3) was more complex compared to W1P2. This is because from W2 onwards, new household members (NHM) were recruited into the panel (which was not the case in W1P2). In principle, this covers (a) persons that moved into a participating household, (b) if an individual SCP panelist (AP or HM) moved out of their original household into a new one, the persons that live in that new household, (c)

young adults living in a participating household who recently turned 18, and (d) other persons who already lived in a participating household before but were initially not named as HMs by their AP. The existence of group (d), however, renders the previously estimated probability of HMs to be named by the AP (and therefore invited to the survey) $\hat{p}_2$ invalid for W2, as the number of HMs named by the AP could in fact have increased since then.

Therefore, the generation of cross-sectional nonresponse weights for Wave 2 and 3 mimicked the generation of initial cross-sectional nonresponse weights for W1P1: the weights reflect a combination of the probability of a household to be in the panel, the proportion of persons in the household that were invited to the survey, and the probability of invited persons to respond. The following description refers to W2, and the equivalent procedure for W3.

The generation of *HHRF* and *PHRF* weights involved the estimation of the following:

- (1) Whether a once participating household was still part of the panel (in the sense that at least one person could be invited to the surveys). If the original AP had dropped out of the panel, another HM was assigned the role of AP, so this essentially corresponds to the question of whether there was still an AP left in the household who could answer the household questionnaire module. The estimated probability of this is denoted by $\hat{p}_5$.
- (2) The proportion of adult persons in the household other than the AP that were part of the panel and could therefore be invited to participate in this wave. This accounts for (a) underreporting of household members by the AP and (b) initial nonresponse and panel attrition by household members once invited to the survey. This proportion, conceived as the probability of a non-AP person in a household to be part of the panel, is denoted by $\hat{p}_6$.
- (3) Whether invited persons participated in this wave. Assuming that the mechanism of continued participation for panelists (APs and HMs) and the mechanism of first-time participation of newly invited NHMs may be substantially different (and considering that fundamentally different predictive data is available for both groups), this must be differentiated between two separate models:
 - APs' and HMs' response probabilities (estimates in the following denoted by $\hat{p}_7$, whereas $\hat{p}_7^{AP}$ represents the estimated response probabilities of the AP of a household), and
 - NHMs' response probabilities (estimates in the following denoted by $\hat{p}_8$).

The starting point of the *HHRF* and *PHRF* weight for W2 were the W1P1 *HHRF* weights and *PHRF* weights of the household's AP, respectively. First, a start weight $HHRF_{\text{start}}$ was derived from $HHRF_{W1P1}$, with an adjustment for changes in household size (i.e., the number of adult persons living in the household):

$$HHRF_{\text{start}} = 1 - (1 - a)^{\min(\text{SIZE}_{W2}, 13)} \quad (11)$$

with

$$a = 1 - \sqrt{\frac{1}{HHRF_{W1P1}}}^{\min(\text{SIZE}_{W1P1}, 13)} \quad (12)$$

The procedure for *PHRF* was the same.

To generate $HHRF_{W2}$, the start weight was multiplied by the estimated probability of the household to remain in the panel and with the estimated probability of the household's AP to respond in W2 given panel inclusion. Thereafter, extreme weights were trimmed (t_7), and household-level raking (r_7) was applied to the trimmed weights. The raking procedure was the same as for W1P1 and W1P2, with margins updated to the survey year of 2022 (instead of 2021 for wave 1). In consequence, the generation of $HHRF_{W2}$ can be described by the operation

$$HHRF_{W2} = HHRF_{\text{start}} \times (\hat{p}_5 \times \hat{p}_7^{\text{AP}})^{-1} \times t_7 \times r_7 \quad (14)$$

As in W1P1, the construction of person-level weights $PHRF_{W2}$ depend on respondent type (i.e., whether a respondent has the role of AP, HM, or NHM). All $PHRF_{W2}$ weights rely on the adjusted $PHRF_{W1P1}$ weight multiplied by a combined W2 response probability, with subsequent trimming of extreme weights (t_8) and person-level raking (r_8). For APs, the combined response probability relies on the estimated probability of the household to remain in the panel, and their probability of response provided they are in the panel at W2. For HMs and NHMs, it additionally relies on the proportion of household members available for invitation. This results in the following operation to create $PHRF_{W2}$:

$$PHRF_{W2} = \begin{cases} PHRF_{\text{start}} \times (\hat{p}_5 \times \hat{p}_7)^{-1} \times t_8 \times r_8 & \text{for APs} \\ PHRF_{\text{start}} \times (\hat{p}_5 \times \hat{p}_6 \times \hat{p}_7)^{-1} \times t_8 \times r_8 & \text{for HMs} \\ PHRF_{\text{start}} \times (\hat{p}_5 \times \hat{p}_6 \times \hat{p}_8)^{-1} \times t_8 \times r_8 & \text{for NHMs} \end{cases} \quad (15)$$

The generation procedure for *PHRF_AP* is much simpler. It required only an estimation of the response probability to W2 among all initially drawn APs (denoted by $\hat{p}_9$). The $PHRF_{AP_{W2}}$ weight was then calculated by multiplying the W1P1 *PHRF_AP* weight with the inverse of the

$\hat{p}_9$ response probability, followed by trimming of extreme weights (t_9) and person-level raking (r_9):

$$\text{PHRF_AP}_{W2} = \text{PHRF_AP}_{W1P1} \times \hat{p}_9^{-1} \times t_9 \times r_9 \quad (16)$$

For *HBLEIB*, the probability of the AP to respond to W2 was estimated among all households with AP participation at W1P2 ($\hat{p}_{10}$). This means that for households in which the AP role switched from one person to another, this is the probability of the current AP to respond given that the previous AP responded in the wave before. For *PBLEIB*, the probability to respond was estimated among all persons who participated in W1P2 ($\hat{p}_{11}$). The longitudinal weights for W2 can therefore be calculated through:

$$\text{HBLEIB}_{W2} = \hat{p}_{10}^{-1} \quad (17)$$

$$\text{PBLEIB}_{W2} = \hat{p}_{11}^{-1} \quad (18)$$

3. Nonresponse model selection and estimation

3.1 Data preparation

Predictor data for nonresponse models were drawn from different data sources:

- Sampling frame data (primarily gender, age, citizenship status)
- Microgeographic data (hereafter geodata) on various (address- to district-) levels of aggregation provided by the companies microm and infas360
- Paradata from the current wave, such as the timing of invitation (excluding variables that are effects of (non)response behavior, like whether reminders were sent)
- For HMs and NHMs: APs' survey data, particularly HMs' gender, age, relationship to the AP, and other data from the household questionnaire module
- For HMs and NHMs: APs' paradata (especially early versus late responders)
- For follow-up waves: individual and household-level survey data from previous waves
- For follow-up waves: paradata from previous waves

Data from these different sources had to be merged and prepared for use in the nonresponse models.

First, where available, data from multiple sources were combined to avoid missing data. If the different sources gave conflicting information, responses by the respondents themselves were used. For example, gender was normally inferred from survey responses in the *PGENDER* variable. If the *PGENDER* variable was missing, alternative data from the sampling frame (for APs), or from the household composition data provided by the AP, were used instead.

Remaining missing information was coded either as missing value or as a separate category. Missing values were subsequently treated by multiple imputation (Rubin, 1987) in the estimation stage (see below).

Second, rare categories in predictor variables can distort the estimation of response probabilities, as random variation within a few observations may push an estimated response propensity to an artificially extreme value (Seaman & White, 2013). In an extreme case, if only responses (or only nonresponses) are observed empirically within a category, the estimated probability can equal one (or zero). To avoid this, predictor variables with rare categories were collapsed into broader groups (see also Kroh et al., 2017; Schmidt & Woll, 2017), and continuous variables were either categorized into broader groups or truncated at the 1st and 99th percentiles.

3.2 Model selection

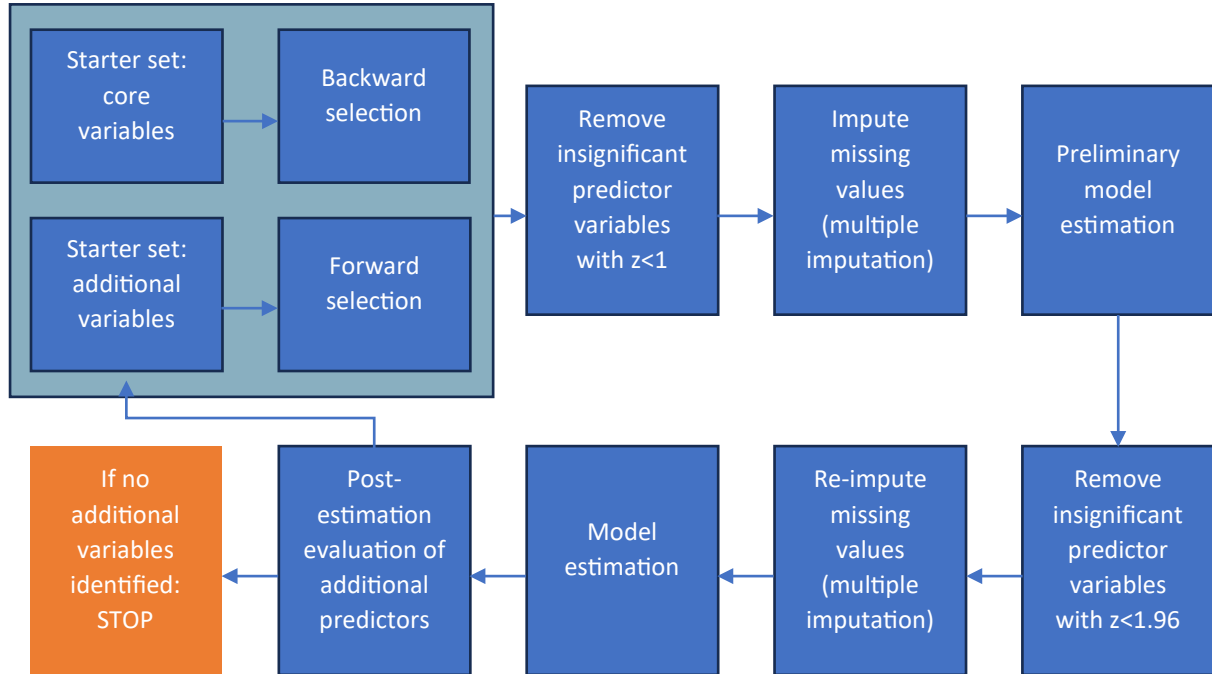
Stepwise selection² was applied to the prepared candidate predictors to reduce the risk of overfitting. Predictor selection was performed using a semi-automated multi-step, iterative procedure in which relevant variables were added and irrelevant variables removed. Figure 1 provides an overview of the predictor selection and model-building process.

- (1) First, a small initial set of core predictors (primarily sociodemographic variables assumed to be associated with nonresponse) was used to form a baseline nonresponse model. From this model, irrelevant predictors were removed using backward selection, with the cross-validation error of the predicted outcome serving as the selection criterion. Subsequently, additional variables from a broader pool of candidate predictors were added using forward selection, applying the same criterion.
- (2) Because this procedure may retain some variables that did not meaningfully contribute to predicting the outcome in the final model, selected predictors that were deemed particularly implausible were excluded after model selection. This applied if z-values were below 1 (i.e., if the coefficient $\pm$ one standard error covered zero). Note that this

² Regularization methods, such as lasso and elastic-net regression, were considered as an alternative to stepwise variable selection because they offer several advantages, particularly their ability to identify globally rather than locally optimal predictor sets (Hastie et al., 2015). However, most observations had missing values in at least one candidate predictor. Standard implementations of lasso and elastic-net regression require complete data, and multiple imputation of the full set of candidate predictors at once was not feasible. With stepwise selection, by contrast, observations need only be excluded if they have missing values in one of the predictors included in the current model. Therefore, stepwise selection was preferred.

is only a pre-selection; a more conservative procedure to exclude insignificant predictors was applied in step (5).

Figure 1: Predictor selection procedure for nonresponse models.



- (3) Using this restricted dataset³, missing data were multiply imputed (Rubin, 1987) using the mice and miceadds packages in R (R Core Team, 2022; van Buuren & Groothuis-Oudshoorn, 2011; Robitzsch & Grund, 2023). Categorical variables were imputed via classification trees (Burgette & Reiter, 2010), while continuous variables were imputed via partial-least-squares predictive mean matching (Grund et al., 2024). The imputation model covered all analysis variables. In addition, for variables with a large proportion of missing data, a small number of auxiliary variables that were highly correlated to the concerned variable were sometimes added to the imputation model in a case-by-case decision.
- (4) After imputation, the nonresponse models was estimated separately in each imputed dataset (using logistic regression for dichotomous outcome variables and fractional logistic regression for outcome variables that reflect probabilities), and the resulting estimates were combined using Rubin's rules (Rubin, 1987).

³ Large numbers of predictor variables in an imputation model can affect the performance of multiple imputation negatively (see, for example, van Buuren, 2018). This is why imputation was applied on variables selected into the model after step (2), and not on all candidate predictor variables before step (1).

- (5) For parsimony, predictors that were insignificant on the 5% level ($z < 1.96$) were subsequently removed from the model. The data were then re-imputed based on the new predictor set and the model re-estimated. Based on this final model, predicted response propensities were derived for constructing the nonresponse weights by averaging the linear predictions of response of response (log-odds values) for each individual across imputations and thereafter transforming them to probabilities.
- (6) The performance of the model selection procedure was then evaluated by regressing the response indicator on the predicted response propensity $\hat{p}_1$ and, sequentially, on each remaining variable eligible as a predictor using logistic regression. This step served to identify additional variables that may improve the prediction of nonresponse. Variables that led to a reduction in the Bayesian information criterion (BIC) compared to a model including only $\hat{p}_1$ were fed back into step (1) as additional predictors. This yielded an iterative selection procedure that continued until step (6) ceased to identify any new variables for inclusion.

3.3 Estimated models: Wave 1 Part 1

Anchor person response probability ($\hat{p}_1$)

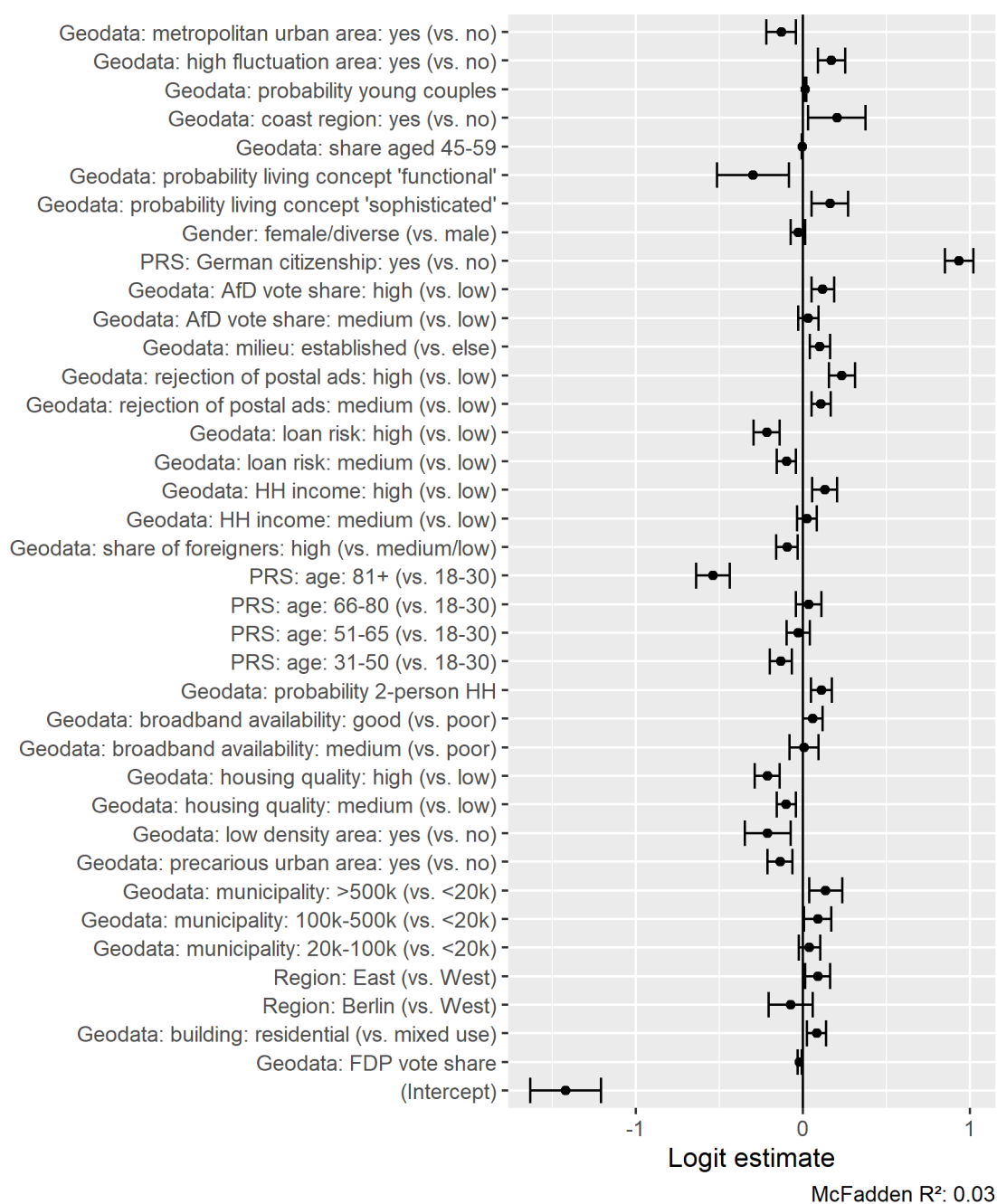
APs' response probabilities in W1P1 conditional on being sampled ($\hat{p}_1$) were estimated using logistic regression. At recruitment, limited information was available for modeling $\hat{p}_1$. This includes:

- information from the sampling frame: age, gender, citizenship status (German or non-German), federal state, and municipality size;
- a variety of geodata at various levels of aggregation provided by the companies microm and infas360.

The coefficients of the final model are displayed in Figure 2. The McFadden R^2 of the logistic regression is 0.03. This very low value might indicate that there is little non-random selection into nonresponse for APs, but may as well originate from the lack of relevant predictive information available at this early stage of nonresponse.

The log-odds of participation were predicted for each anchor person in each imputed dataset. The final $\hat{p}_1$ for each anchor person was obtained by averaging the predicted log-odds across imputations and subsequently transforming the result into a probability.

Figure 2: Logistic regression of APs' response (1) vs. nonresponse (0) on selected predictor variables.



Note: HH = household; PRS = population register sample data. Confidence intervals are shown at the 5% level of significance.

Proportion of household members named by their anchor person ($\hat{p}_2$)

For modelling $\hat{p}_2$, beyond geodata and sampling-frame data, we could additionally rely on the AP's responses to the recruitment survey instrument. Furthermore, we could also rely on paradata from fieldwork, such as survey mode or days between invitation and participation. Variables considered as predictors covered sociodemographic characteristics, household composition as well as attitudes (e.g., trust in science).

Since the proportion could take any value from 0 to 1, fractional logistic regression was used.⁴ A coefficient plot for this model is displayed in Figure 3. This model has a McFadden R^2 of 0.05.

Household member response probability ($\hat{p}_3$)

For modelling $\hat{p}_3$, sources of information were

- AP survey data about themselves and their household
- paradata of AP interviews
- Information about the HMs provided by the AP, namely their age, gender, and the kind of relation between AP and HMs
- The same geodata as before

Logistic regression was used to predict the named HMs' participation probabilities (see Figure 4 for a coefficient plot). The model has a McFadden R^2 of 0.04.

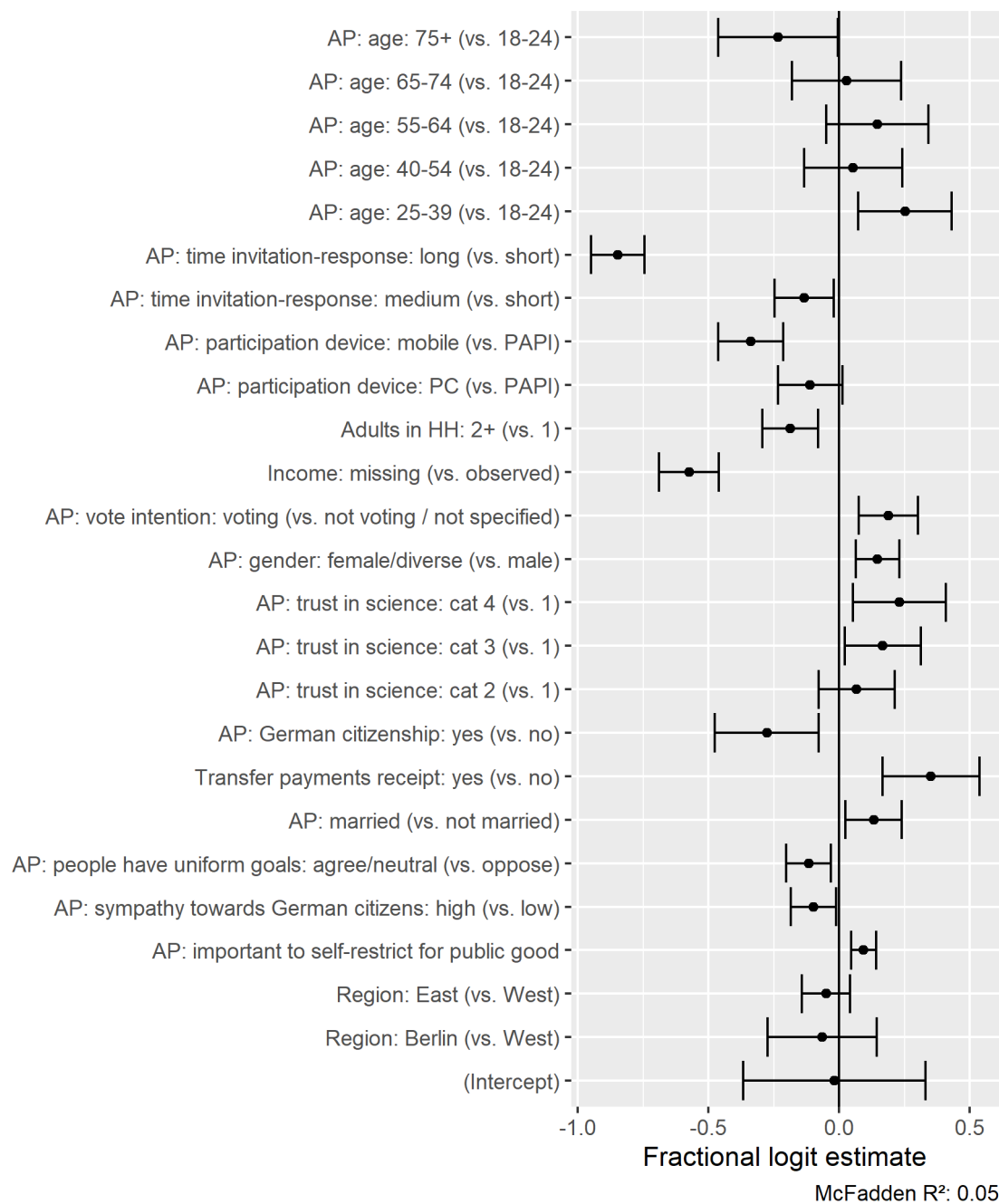
*3.4 Estimated models: Wave 1 Part 2**Response probability ($\hat{p}_4$)*

Logistic regression was used for modelling the response probability to W1P2 among those who participated in W1P1. In addition to W2, survey data and paradata were available as predictors for the model. To account for differences in response behavior mechanisms between APs and HMs, a binary indicator for being an HM was added to the model. Moreover, for all selected main effects, interaction terms with the HM binary indicators were evaluated via forward selection and added if they reduced the cross-validation error.

A coefficient plot for this model is displayed in Figure 5. The McFadden R^2 was 0.07.

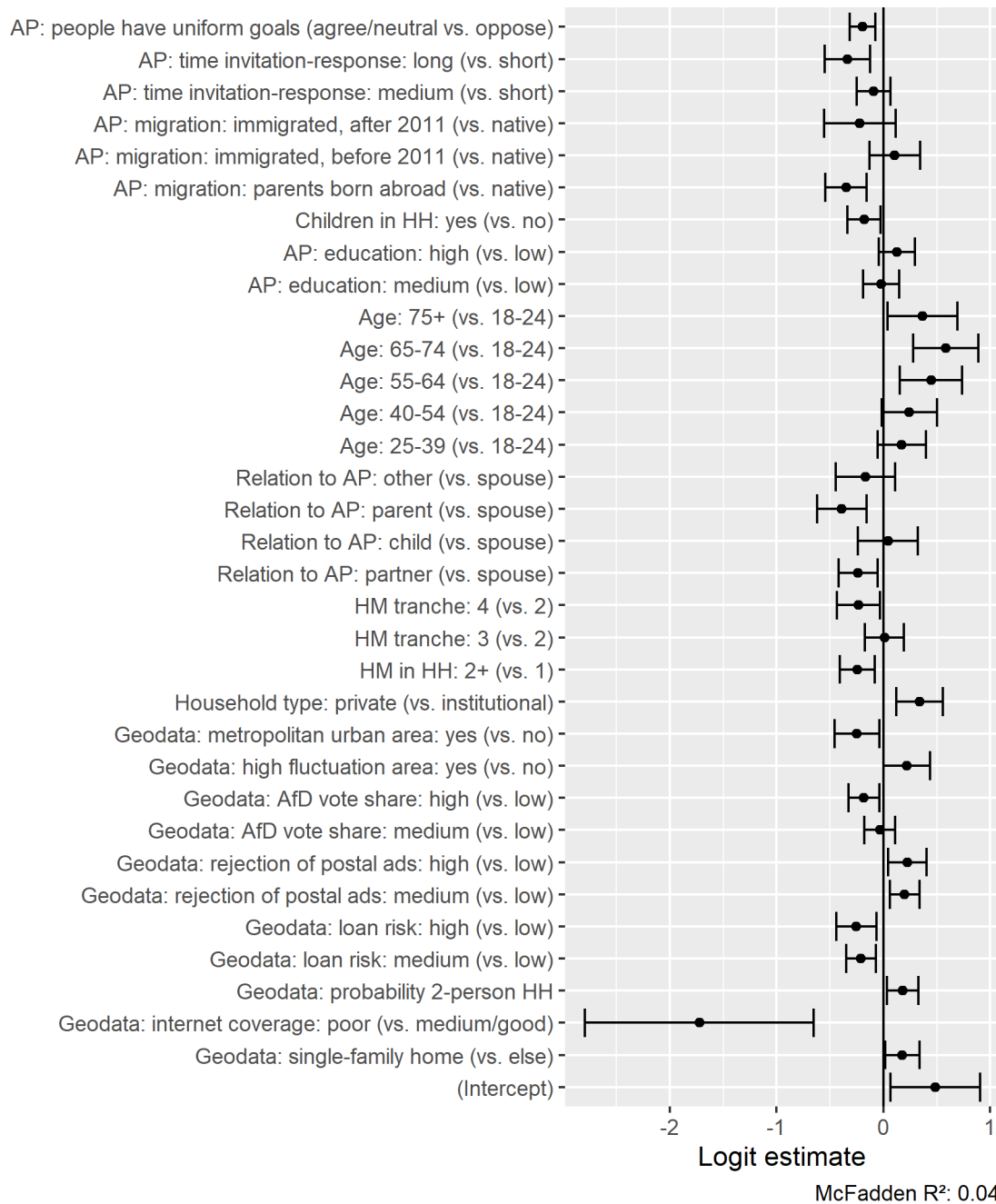
⁴ The majority of outcomes was exactly 0 or 1, though, meaning that most AP either named all their HM or none of them.

Figure 3: Fractional logistic regression of the proportion of named household members on selected predictor variables.



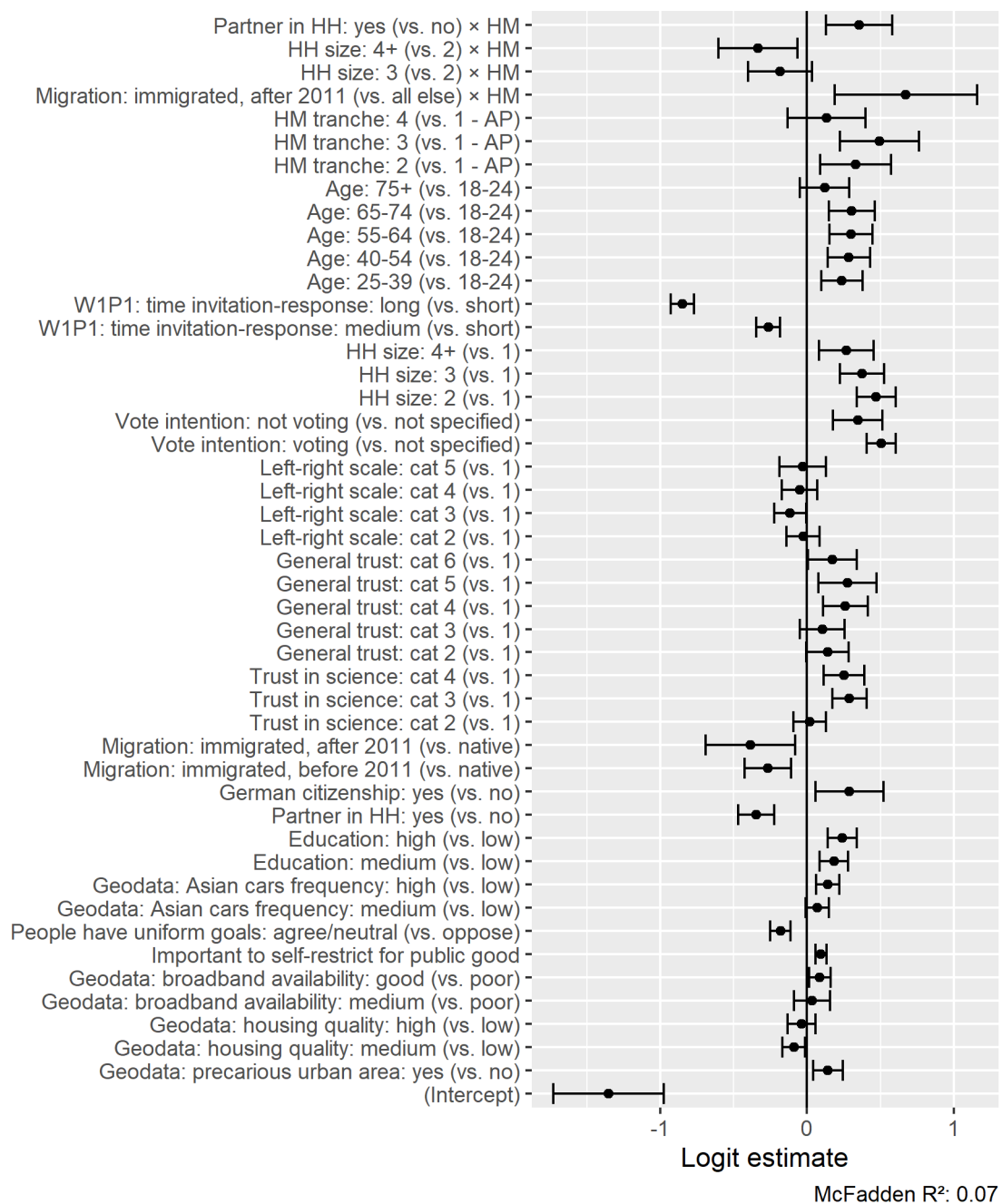
Note: AP = anchor person; cat = category; HH = household. Confidence intervals are shown at the 5% level of significance.

Figure 4: Logistic regression of invited HMs' response (1) vs. nonresponse (0) on selected predictor variables.



Note: AP = anchor person; cat = category; HM = household member; HH = household. Confidence intervals are shown at the 5% level of significance.

Figure 5: Logistic regression of response (1) vs. nonresponse (0) to W1P2 among persons who participated in W1P1 on selected predictor variables.



Note: cat = category; HM = household member; HH = household; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

3.5 Estimated models: wave 2 and 3

Household retention probability ($\hat{p}_5$)

The probability of a household that was interviewed in W1P1 to still be in the panel in W2 or W3 was modeled using logistic regression. This model can rely on W1P1 survey data, paradata, and geodata. In addition, interaction terms with the number of adults in the household were evaluated via forward selection for all selected main-effect predictor variables and included in the model if they reduced the cross-validation error in order to account for the possibility that household-level panel attrition may work differently in larger households compared to 1-person households.

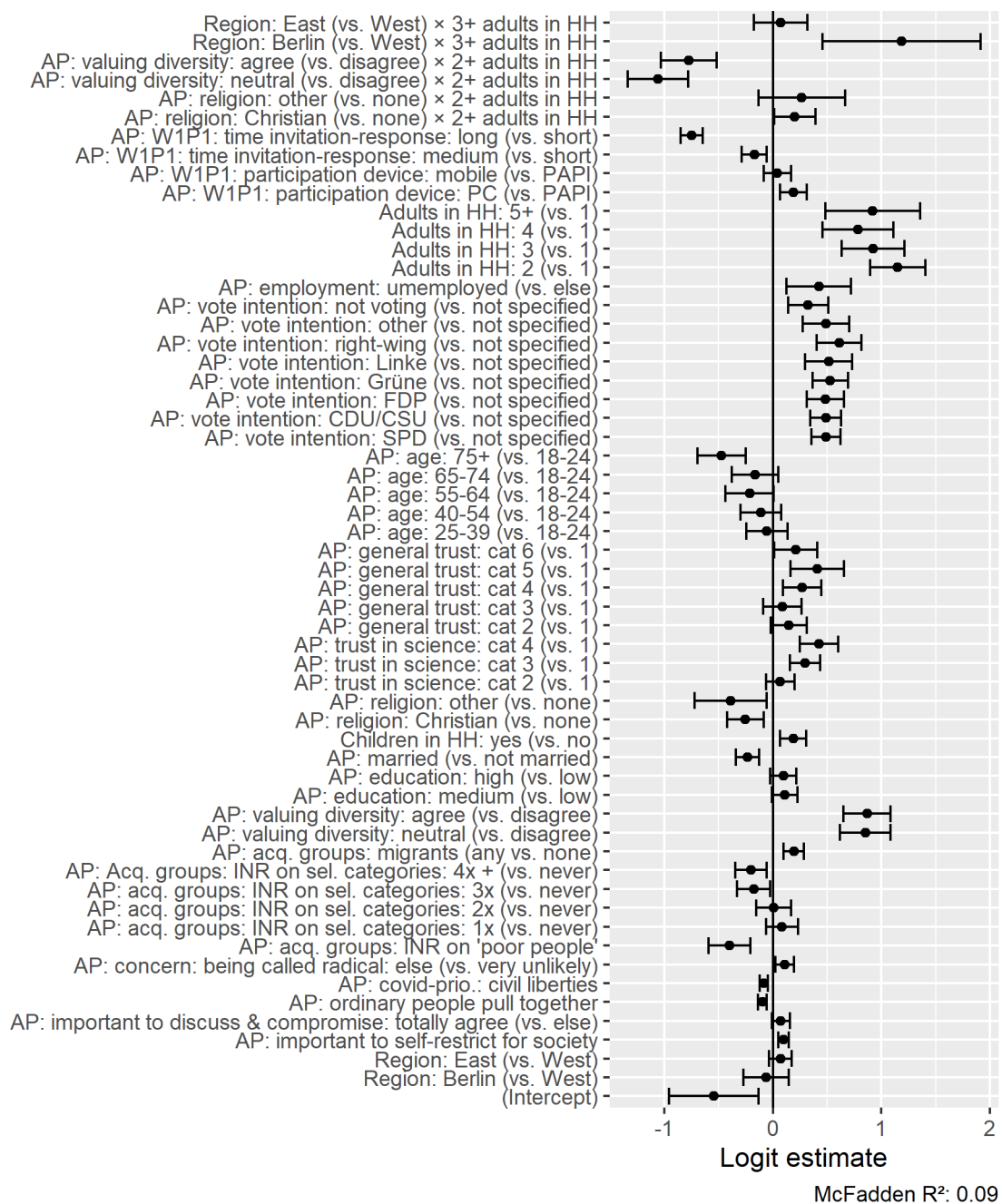
A coefficient plot for the W2 model is displayed in Figure 6 and for the W3 model in Figure 7. Both models have a McFadden R^2 of 0.09.

Proportion of persons of a household in the panel ($\hat{p}_6$)

As for $\hat{p}_2$ in W1P1, fractional logistic regression was used to estimate the proportion of persons in the household (excluding the AP) that could be invited to a follow-up wave. For this model, the dataset was restricted to households that were still part of the panel and had at least one person other than the AP aged 18 or older living in it. Candidate predictor variables for the stepwise selection were W1 survey data and present-wave geodata and paradata. Again, interaction terms of all selected variables with the number of persons aged 18 or older in the household were evaluated. Empirically, however, no interaction term for wave 2 and only one interaction term for wave 3 were selected into the models.

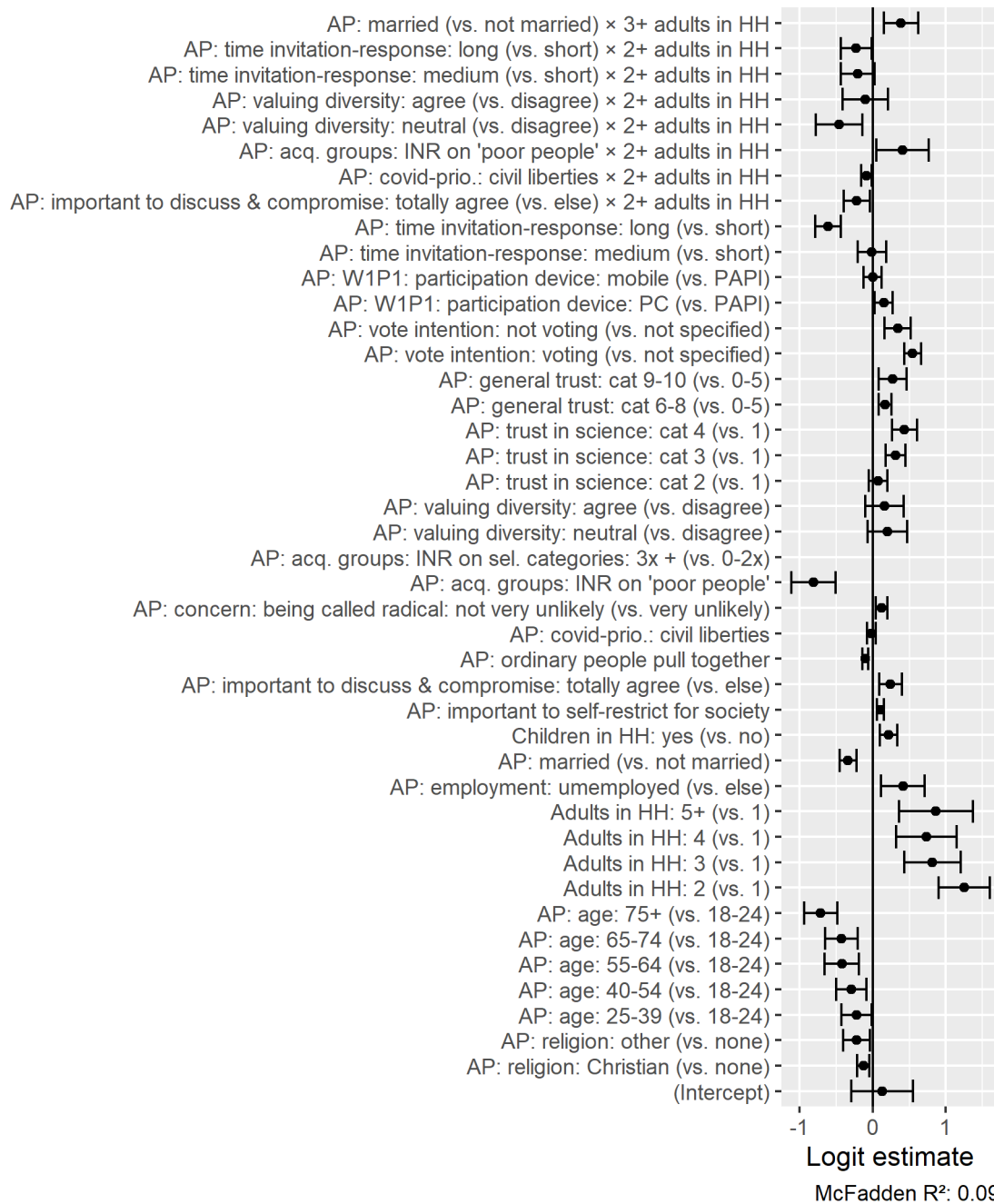
Figures 8 and 9 show coefficient plots for the W2 and W3 models, respectively. Both models have a McFadden R^2 of 0.04.

Figure 6: Logistic regression of W1P1 households to still be in the panel at W2 (1) vs. not being in the panel anymore (0) on selected predictor variables.



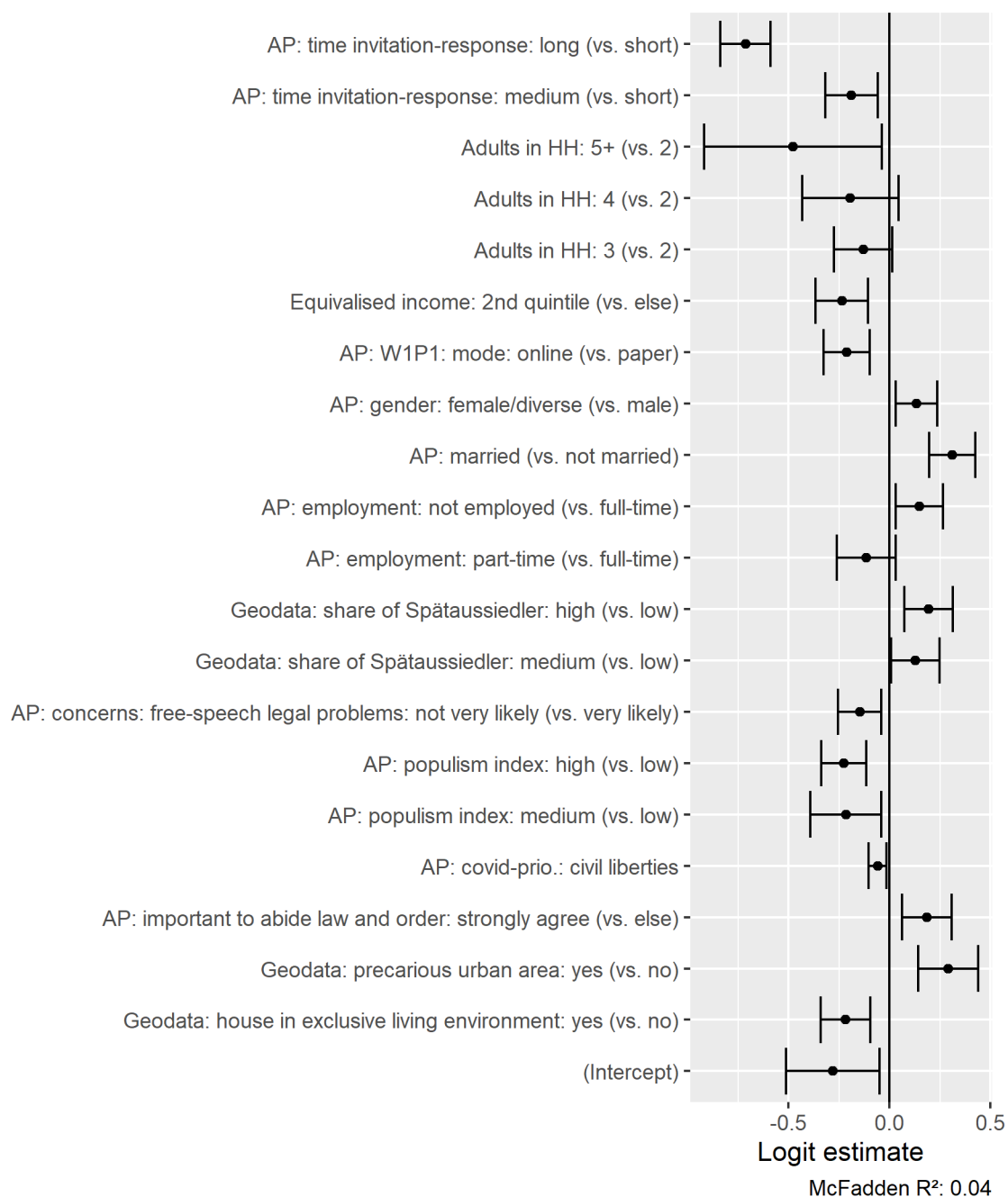
Note: AP = anchor person; cat = category; HH = household; INR = item nonresponse; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

Figure 7: Logistic regression of a W1P1 household to still be in the panel at W3 (1) vs. not being in the panel anymore (0) on selected predictor variables.



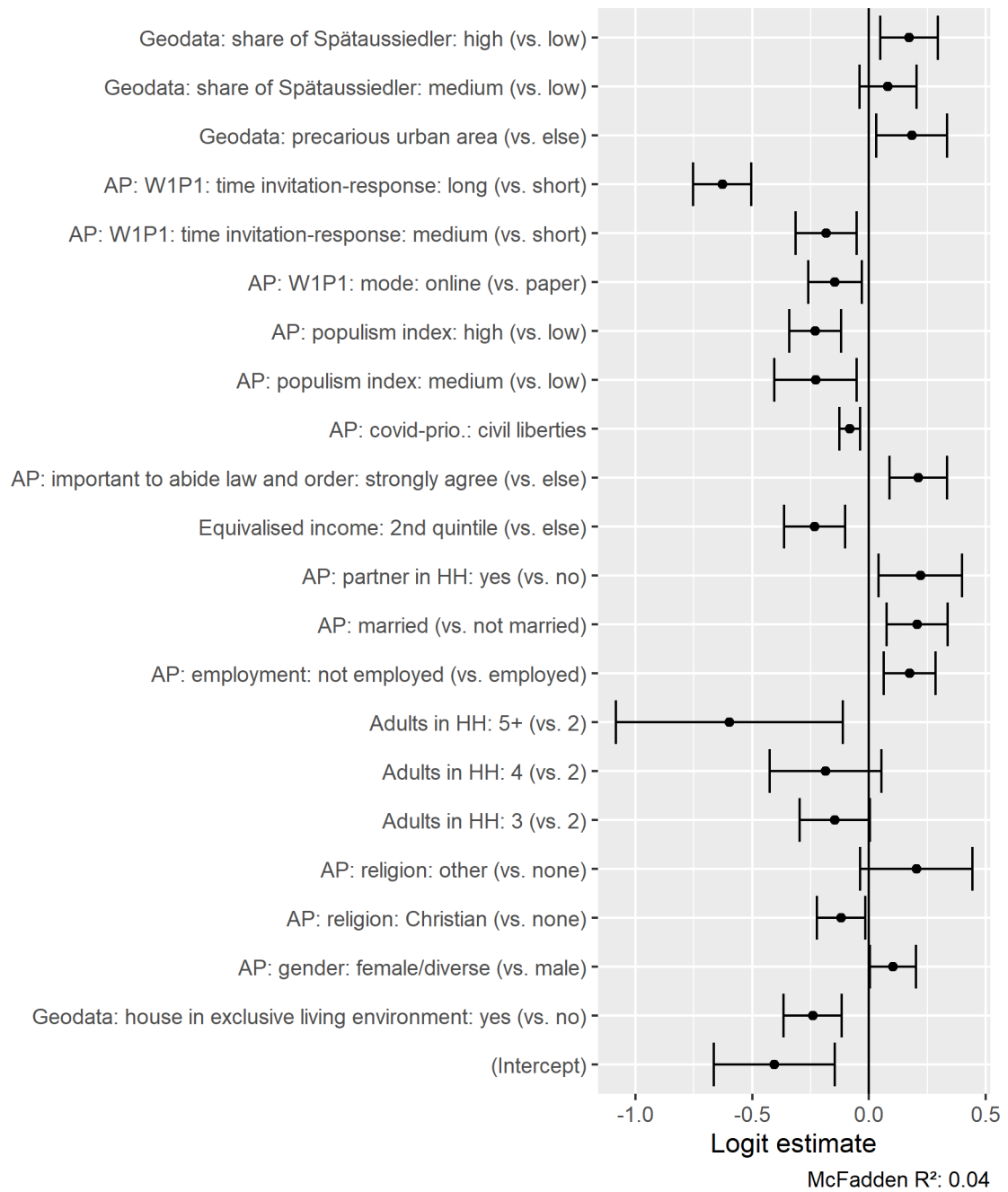
Note: AP = anchor person; cat = category; HH = household; INR = item nonresponse; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

Figure 8: Fractional logistic regression of W2 households with more than one person aged 18 or older of the proportion of W2 panel members to the total number of persons aged 18 or older living in the household (excl. AP) on selected predictor variables.



Note: AP = anchor person; HH = household; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

Figure 9: Fractional logistic regression of W3 households with more than one person aged 18 or older of the proportion of W3 panel members to the total number of persons aged 18 or older (excl. AP) on selected predictor variables.



Note: AP = anchor person; HH = household; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

Anchor persons and household members response probability ($\hat{p}_7$)

Logistic regression was used to estimate the response probability of APs and HMs that were still part of the panel's gross sample at a given wave. W1 survey data as well as present-wave paradata and geodata could be used for these models. As in W1P2, a dummy for being an HM was added, and interaction terms of all selected predictor variables with the HM dummy were tested and added if necessary.

Figures 10 and 11 show coefficient plots for the W2 and W3 models of $\hat{p}_7$, both having a McFadden R^2 of 0.06.

New household member response probability ($\hat{p}_8$)

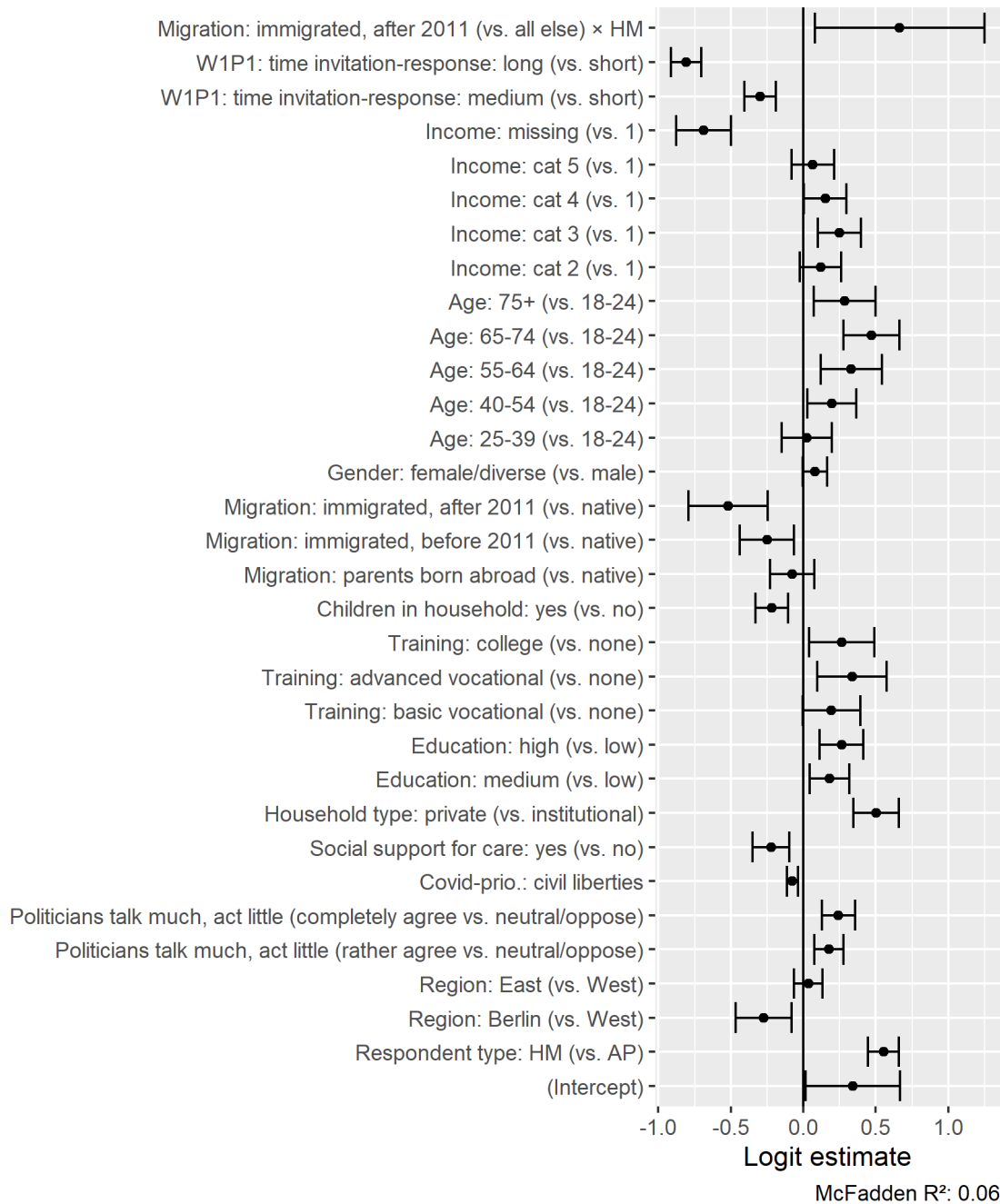
Models for the response probability of invited new household members were estimated using logistic regression and could rely on the current AP's survey data, paradata, and geodata, as well as limited information available on NHMs (gender, age, relation to the AP) provided by the AP.

Figures 12 and 13 show coefficient plots for the W2 and W3 models of $\hat{p}_8$. Since typically only a few hundred NHMs are invited each wave, relatively few predictor variables were selected into the model. Still, both models show a relatively high McFadden R^2 of 0.13 and 0.14, indicating that nonresponse among invited NHMs may be quite selective. The most prominent effect in this model is that later tranches of NHMs tend much more towards nonresponse than the respective earlier tranche. This might reflect two underlying mechanisms: First, these individuals have less time to respond before the field closes, potentially resulting in a lower response rate within this group. Second, and perhaps more importantly, individuals in a later tranche were in this tranche because their respective AP responded relatively late, potentially only after being reminded, suggesting that later tranches may to some extent represent households that are hard-to-survey.

Original-anchor-person response probability ($\hat{p}_9$)

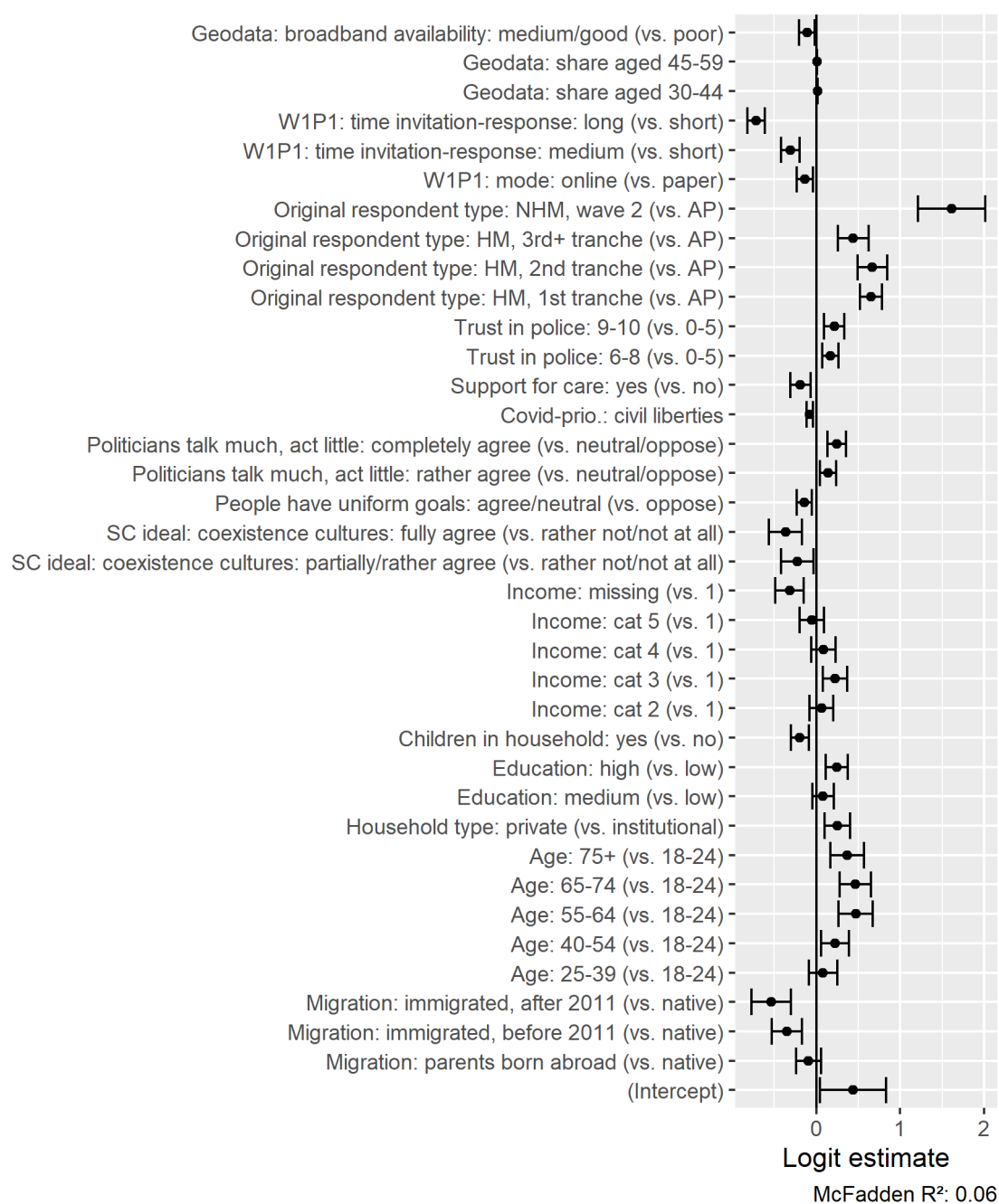
The probability to respond to a wave among all anchor persons of W1P1, as utilized for the generation of *PHRF_AP*, was estimated using logistic regression. Candidate predictors were all available data from W1P1. The coefficient plots for these models are shown in Figures 14 and 15. The McFadden R^2 was 0.09 for wave 2 and 0.07 for wave 3.

Figure 10: Logistic regression of response (1) vs. nonresponse (0) to W2 among anchor persons and household members invited to W2 on selected predictor variables.



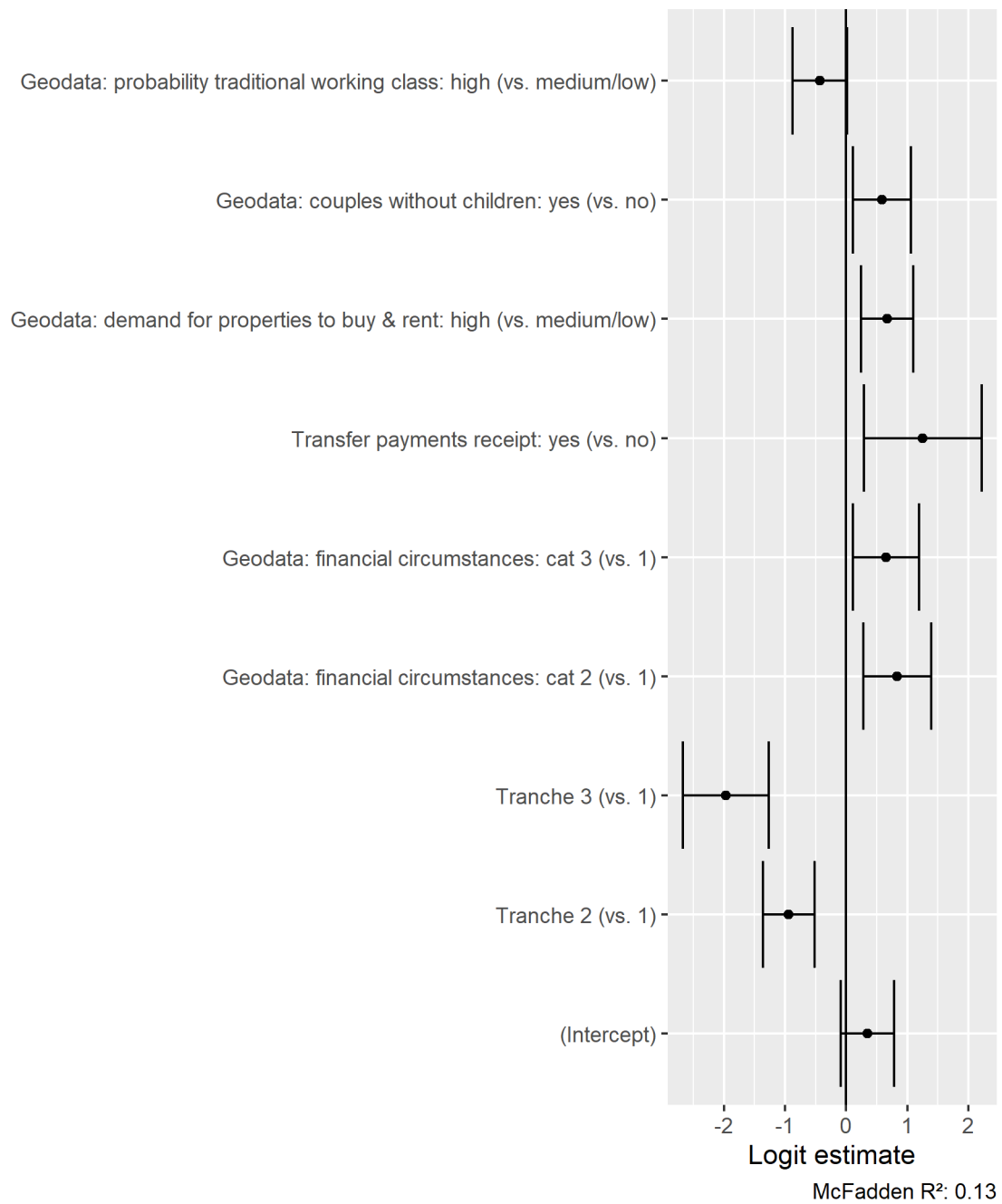
Note: cat = category; HH = household; HM = household member; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

Figure 11: Logistic regression of response (1) vs. nonresponse (0) to W3 among anchor persons and household members invited to W3 on selected predictor variables.



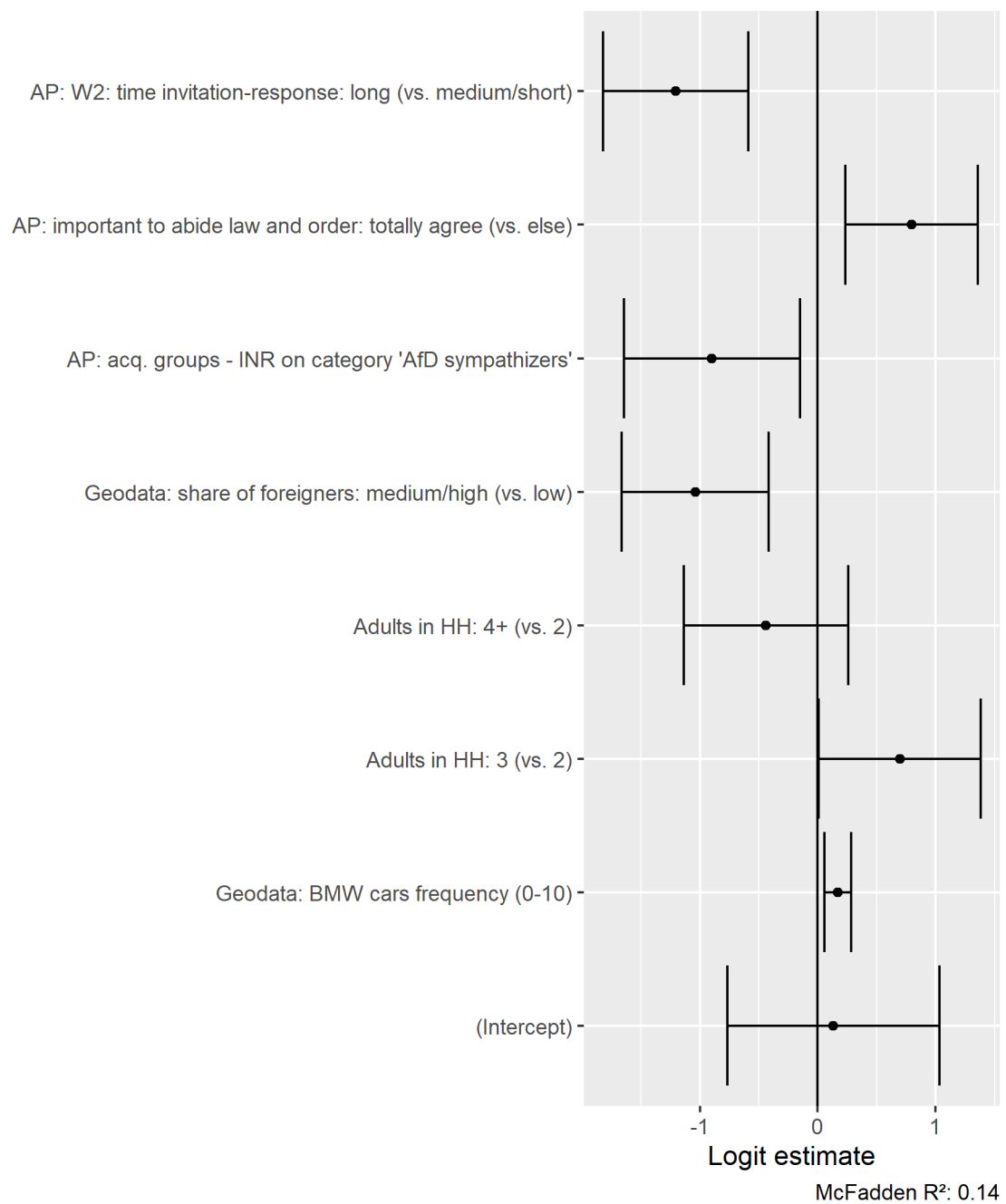
Note: AP = anchor person; cat = category; HH = household; HM = household member; NHM = new household member; SC = societal cohesion; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

Figure 12: Logistic regression of response (1) vs. nonresponse (0) to W2 among new household members invited to W2 on selected predictor variables.



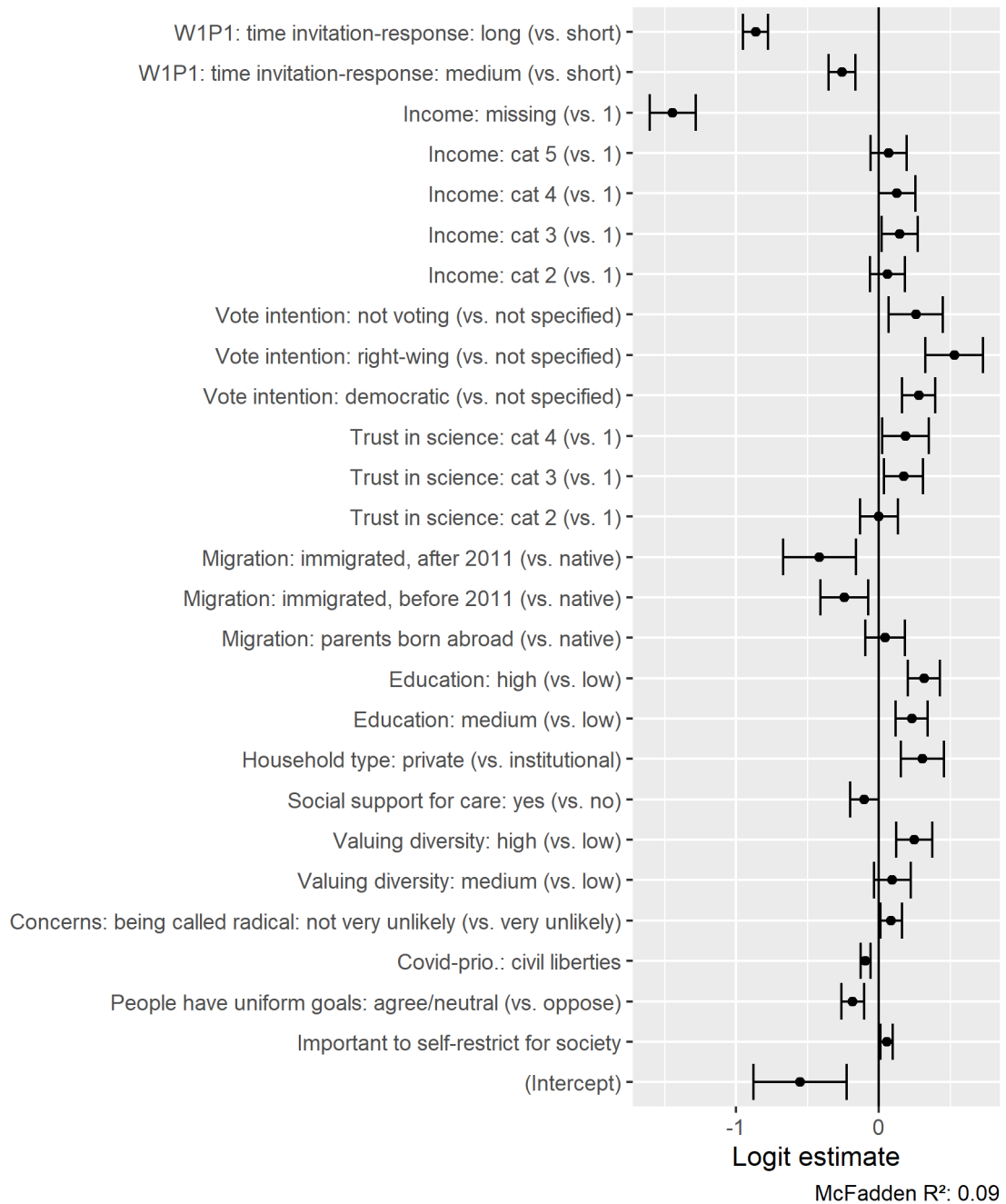
Note: cat = category. Confidence intervals are shown at the 5% level of significance.

Figure 13: Logistic regression of response (1) vs. nonresponse (0) to W3 among new household members invited to W3 on selected predictor variables.



Note: acq. = acquaintance; AP = anchor person; cat = category; HH = household; W2 = wave 2. Confidence intervals are shown at the 5% level of significance.

Figure 14: Logistic regression of response (1) vs. nonresponse (0) to W2 among original W1P1 anchor persons who participated in W1P1 on selected predictor variables.



Note: cat = category; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

Figure 15: Logistic regression of response (1) vs. nonresponse (0) to W3 among original W1P1 anchor persons who participated in W1P1 on selected predictor variables.



Note: acq. = acquaintance; cat = category; INR = item nonresponse; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

Household-level staying probability ($\hat{p}_{10}$)

The staying household-level staying probability, i.e. the probability of a household whose AP participated in the previous wave to have their AP participate also in the present wave, was estimated using logistic regression. All data from the previous wave could be used as candidate predictors for these models.

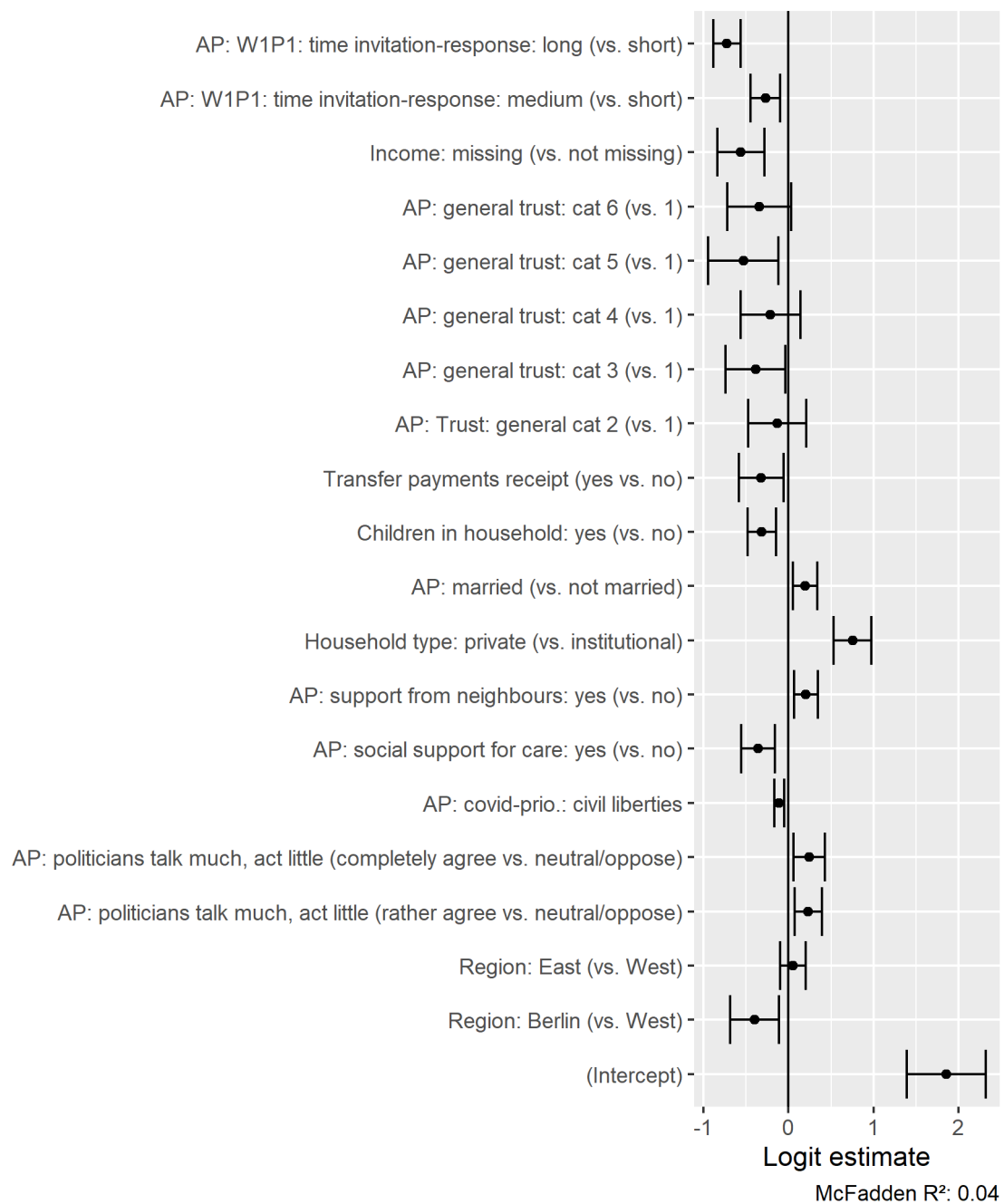
Figures 16 and 17 show coefficient plots for the W2 and W3 models. McFadden R^2 is 0.04 for W2 and 0.05 for W3.

Individual-level staying probability ($\hat{p}_{11}$)

Finally, the probability of an individual participating in the previous wave to also participate in the present wave was, again, estimated via logistic regression. As for the household-level staying probability, all data from the previous waves were available as candidate predictors. In addition, an HM binary indicator was added, and interaction terms of this indicator with all other selected predictor variables were evaluated and, if necessary, added to the model.

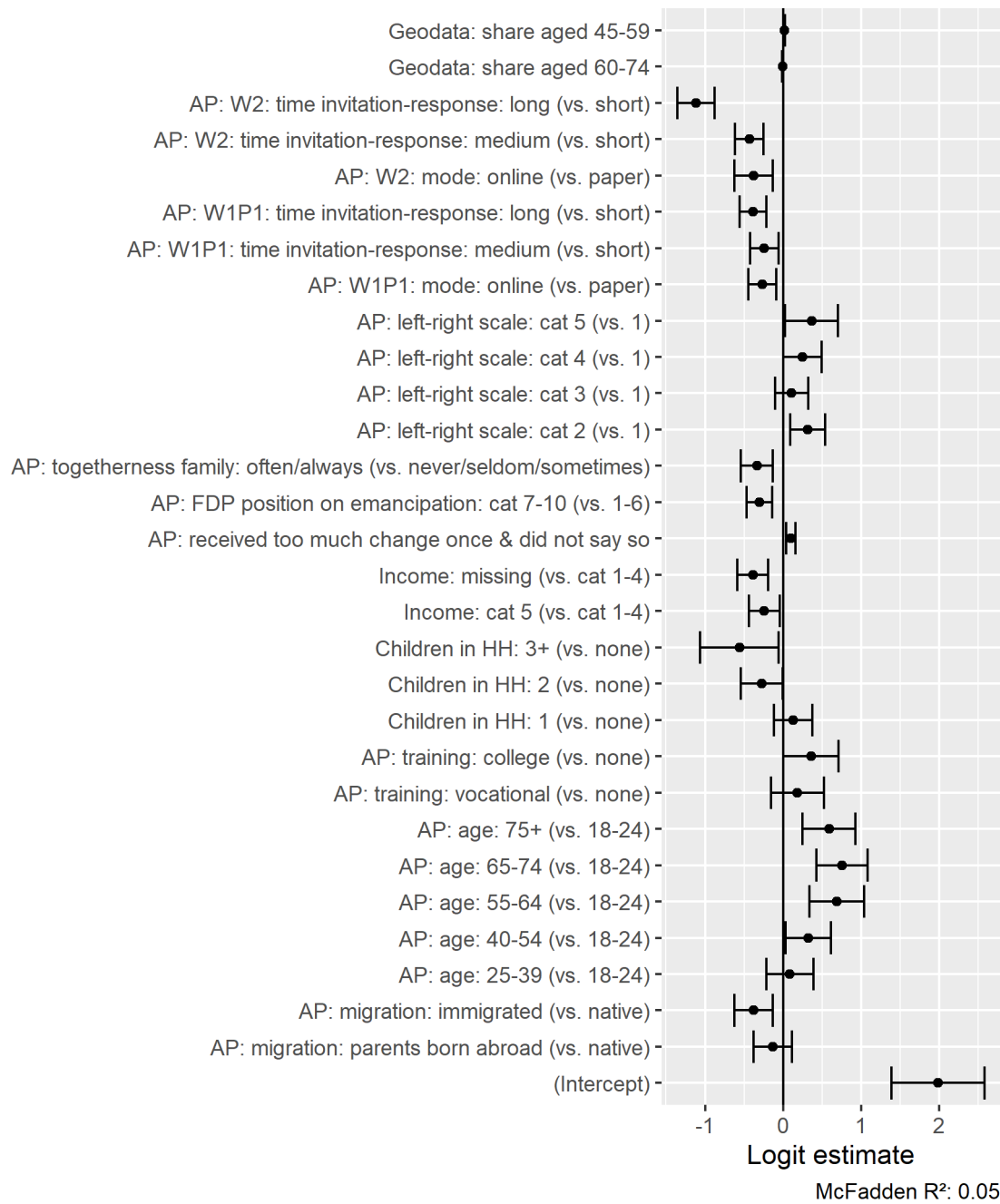
Figures 18 and 19 display coefficient plots for W2 and W3, respectively. McFadden R^2 for these models was 0.04 and 0.06.

Figure 16: Logistic regression of anchor person response (1) vs. anchor person nonresponse (0) to W2 among households whose anchor person participated in W1P2 on selected predictor variables.



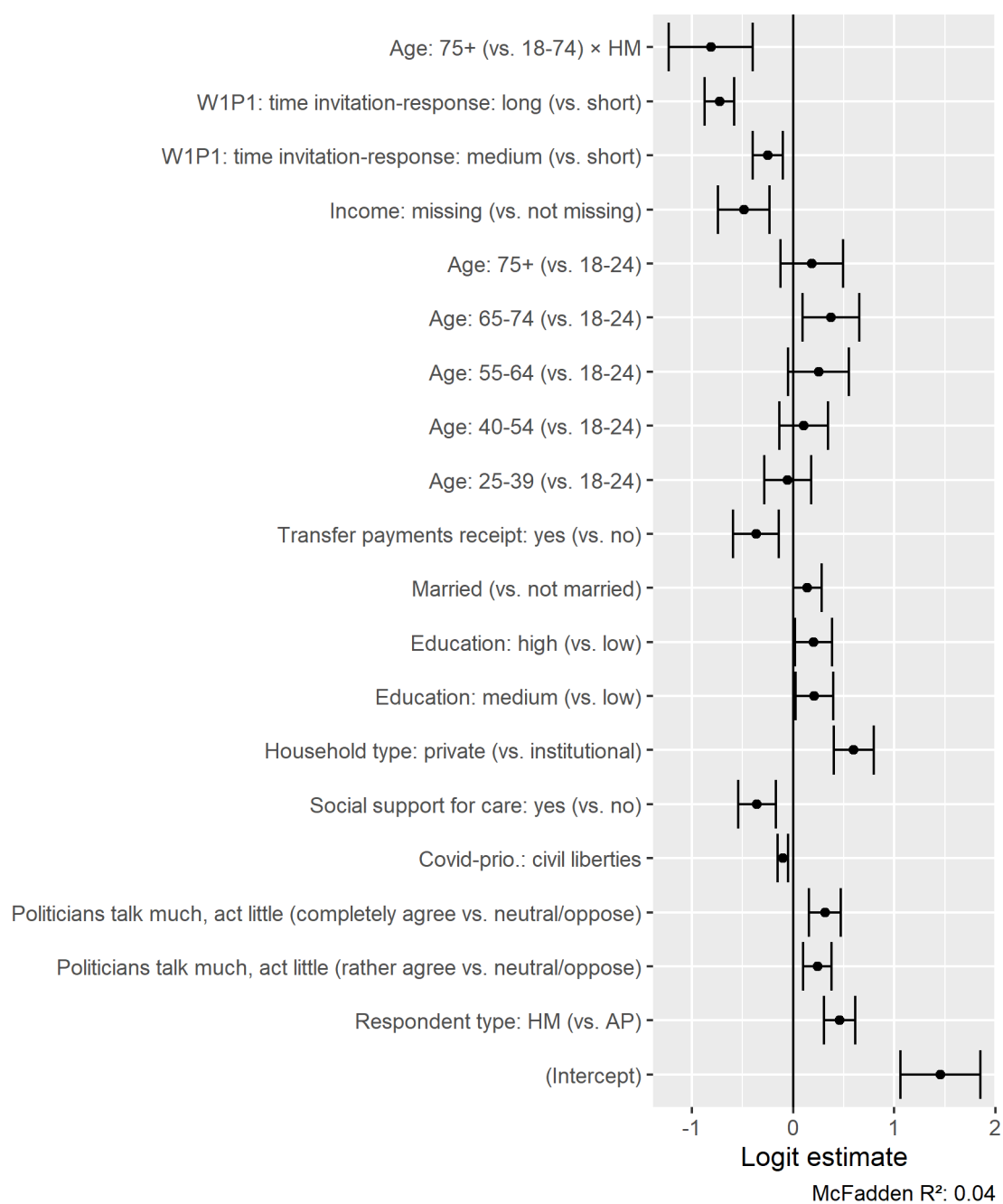
Note: AP = anchor person; cat = category; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

Figure 17: Logistic regression of anchor person response (1) vs. anchor person nonresponse (0) to W3 among households whose anchor person participated in W2 on selected predictor variables.



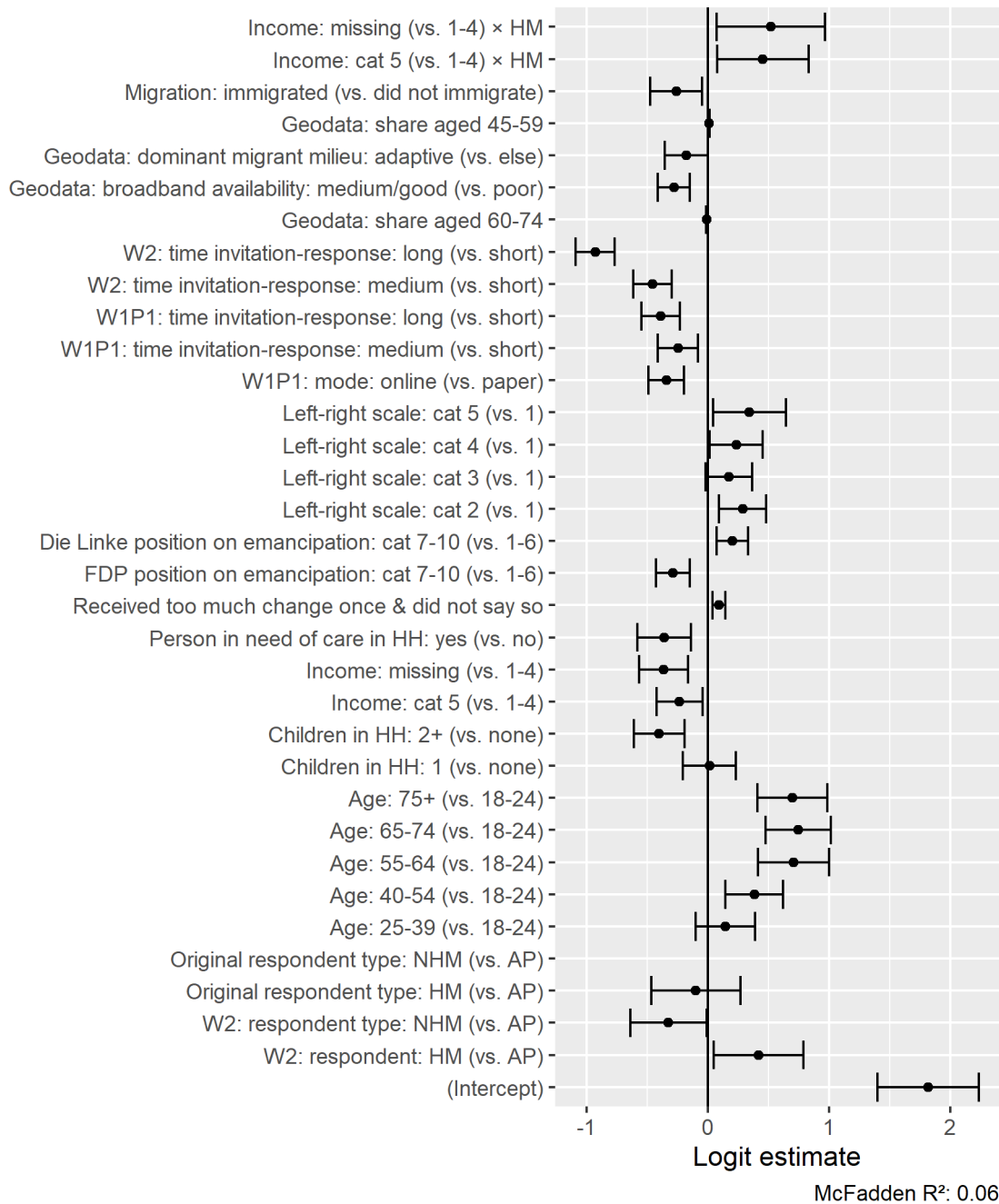
Note: AP = anchor person; cat = category; W1P1 = wave 1 part 1; W2 = wave 2. Confidence intervals are shown at the 5% level of significance.

Figure 18: Logistic regression of response (1) vs. nonresponse (0) to W2 among persons who participated in W1P2 on selected predictor variables.



Note: AP = anchor person; HM = household member; W1P1 = wave 1 part 1. Confidence intervals are shown at the 5% level of significance.

Figure 19: Logistic regression of response (1) vs. nonresponse (0) to W3 among individuals who participated in W2 on selected predictor variables.



Note: AP = anchor person; cat = category; HH = household; HM = household member; NHM = new household member; W1P1 = wave 1 part 1; W2 = wave 2. Confidence intervals are shown at the 5% level of significance.

4. Empirical distribution of cross-sectional nonresponse weights across the generation process

This section examines how the nonresponse weights are distributed, and how each step of the weighting procedure shapes that distribution. Table 2 shows the distributions of the weights together with the coefficient of variation (CV) — a summary measure of how spread out the weights are, both for the design weights and for the *HHRF* nonresponse weights at each stage: (a) raw weights, which include the estimated response probabilities but no trimming, (b) the same weights after trimming, and (c) final weights after raking, shown separately for W1P1, W1P2, W2, and W3. The CV shows that incorporating response probabilities raises weights overall — as expected, since this step compensates for nonresponse — and moderately increases their variation. Trimming then reduces this variation somewhat by cutting back the largest weights, while raking increases it again by aligning the sample with known population margins. This pattern repeats in later waves, which additionally shows a general tendency toward overall larger weights and greater variation in weights over time. Both observations may be attributed to panel attrition after the recruitment wave.

Table 3 shows the same picture for the person-level *PHRF* weights. The pattern is similar to that at the household level: incorporating response probabilities increases both the average weight and its variation, trimming reduces variation by capping the largest values, and raking increases variation again. Moreover, weights grow in size and variation across waves. Person-level weights vary somewhat more than household-level weights, which required more pronounced trimming. Even so, raking at the individual level introduces a notable increase in variation, particularly in wave 3, to align the sample's gender, age, region, and education distributions with the population.

Table 4 presents the equivalent figures for the anchor-person weights, *PHRF_AP*. The same general patterns hold as for *HHRF* and *PHRF*, but with two notable differences. First, *PHRF_AP* weights tend to be somewhat larger, since they represent the same target population as *PHRF* while relying only on the smaller sample of original anchor-persons. Second, they vary less, which also makes a gentler trimming approach possible.

Table 2: Distribution of household-level nonresponse weights (*HHRF*) in W1P1, W1P2, W2, and W3 across the various stages of their generation. CV = coefficient of variation.

	Min.	1%	25%	Median	75%	99%	Max.	CV
Design weights (<i>DESIGN</i>)	70.1	302.3	574.8	1149.1	1149.1	2297.8	2297.8	0.558
W1P1 (n=13,053)								
Raw weights before trimming	267.4	668.4	1,587.3	2,660.8	3,695.4	12,356.9	24,364.0	0.740
Raw weights after trimming	267.4	668.4	1,587.3	2,660.8	3,695.4	12,356.9	14947.2	0.711
Final weights after raking (<i>HHRF</i>)	90.0	327.6	1,321.1	2,331.1	3,636.1	13,626.6	25,208.0	0.888
W1P2 (n=6,669)								
Raw weights before trimming	192.6	627.0	2,408.5	4,076.3	7,174.2	34,298.5	109070.6	1.129
Raw weights after trimming	192.6	627.0	2,408.5	4,076.3	7,174.2	34,298.5	39,682.0	1.022
Final weights after raking (<i>HHRF</i>)	225.5	709.9	2,574.5	4,239.3	7,276.9	32,107.5	42,127.9	0.949
W2 (n=6,128)								
Raw weights before trimming	195.2	687.6	2,552.6	4,254.3	7,498.9	37,578.8	236,395.8	1.248
Raw weights after trimming	195.2	687.6	2,552.6	4,254.3	7,498.9	20,771.2	23,879.5	0.843
Final weights after raking (<i>HHRF</i>)	320.0	832.6	2,812.6	4,738.9	8,393.8	24,546.4	42,114.9	0.865
W3 (n=5,562)								
Raw weights before trimming	157.4	673.6	2,712.7	4,663.3	8,235.5	41,451.6	149,018.1	1.230
Raw weights after trimming	157.4	673.6	2,712.7	4,663.3	8,235.5	26,716.9	29,039.6	0.912
Final weights after raking (<i>HHRF</i>)	176.6	862.4	2,890.5	5,047.6	9,058.5	30,235.3	48,360.8	0.928

Table 3: Distribution of person-level nonresponse weights (*PHRF*) in W1P1, W1P2, W2, and W3 across the various stages of their generation. CV = coefficient of variation.

	Min.	1%	25%	Median	75%	99%	Max.	CV
Design weights (<i>DESIGN</i>)	70.1	226.8	460.0	1149.1	1149.1	2297.8	2297.8	0.556
W1P1 (n=17,027)								
Raw weights before trimming	90.0	363.3	1,553.0	2,652.8	4,929.6	17,081.1	63,451.1	0.933
Raw weights after trimming	90.0	363.3	1,553.0	26,52.8	4,929.6	12,329.4	13,860.5	0.798
Final weights after raking (<i>PHRF</i>)	108.0	374.3	1524.3	2739.3	5257.9	18135.7	31523.1	0.949
W1P2 (n=9,168)								
Raw weights before trimming	208.8	664.1	2,611.7	4,734.2	9,254.9	41,552.2	215,205.2	1.173
Raw weights after trimming	208.8	664.1	2,611.7	4,734.2	9,254.9	27,109.8	30,033.7	0.908
Final weights after raking (<i>PHRF</i>)	201.9	624.6	2,483.8	4,706.2	9,434.1	37,223.2	53,866.9	1.044
W2 (n=8,642)								
Raw weights before trimming	217.5	706.8	2,751.6	4,954.5	9,633.6	51,636.5	545,694.5	1.439
Raw weights after trimming	217.5	706.8	2,751.6	4,954.5	9,633.6	25,937.1	31,026.4	0.877
Final weights after raking (<i>PHRF</i>)	154.3	572.8	2,424.2	4,649.5	9,729.4	41,872.2	87,190.7	1.137
W3 (n=7,992)								
Raw weights before trimming	166.4	696.6	2942.6	5,422.2	10,487.5	58,345.6	332,111.8	1.362
Raw weights after trimming	166.4	696.6	2942.6	5,422.2	10,487.5	25,460.4	32,748.6	0.886
Final weights after raking (<i>PHRF</i>)	122.7	491.3	2,356.1	4,854.6	10,366.2	49,368.1	105,785.8	1.216

Table 4: Distribution of person-level nonresponse weights (*PHRF_AP*) in W1P1, W1P2, W2, and W3 across the various stages of their generation. CV = coefficient of variation.

	Min.	1%	25%	Median	75%	99%	Max.	CV
Design weights (<i>DESIGN_AP</i>)	905.9	905.9	905.9	2297.8	2297.8	2297.8	2297.8	0.349
W1P1 (n=13,053)								
Raw weights before trimming	1,465.7	1,816.5	2,941.9	5,311.5	6,393.3	15,919.5	33,819.9	0.516
Raw weights after trimming	1,465.7	1,816.5	2,941.9	5,311.5	6,393.3	15,919.5	23,960.6	0.515
Final weights after raking (<i>PHRF_AP</i>)	919.2	1,402.7	2,960.2	4,686.6	6,543.9	17,975.7	34,373.9	0.623
W1P2 (n=6,519)								
Raw weights before trimming	1,249.6	2,111.6	5,274.4	7,965.8	12,287.9	46,708.0	145,461.7	0.878
Raw weights after trimming	1,249.6	2,111.6	5,274.4	7,965.8	12,287.9	36,233.3	40,245.6	0.720
Final weights after raking (<i>PHRF_AP</i>)	1,115.7	1,909.8	5,039.7	7,808.6	12,596.7	50,837.1	67,028.2	0.860
W2 (n=5,874)								
Raw weights before trimming	1,341.1	2,223.3	5,504.8	8,284.5	13,092.7	64,356.5	581,937.1	1.310
Raw weights after trimming	1,341.1	2,223.3	5,504.8	8,284.5	13,092.7	62,455.0	73,137.4	0.901
Final weights after raking (<i>PHRF_AP</i>)	991.4	1,868.2	5,006.3	7,742.2	13,222.1	74,845.7	124,303.1	1.090
W3 (n=5,284)								
Raw weights before trimming	1,470.9	2,494.4	6,094.2	9,352.0	14,772.9	67,916.9	276,932.2	1.051
Raw weights after trimming	1,470.9	2,494.4	6,094.2	9,352.0	14,772.9	60,586.1	68,284.1	0.844
Final weights after raking (<i>PHRF_AP</i>)	823.2	1,622.7	4,871.1	8,081.6	14,696.3	80,154.1	142,213.8	1.166

Who are the households with the highest weights?

As described above, the SCP *PHRF* weights involve a high degree of variability, reflecting both differential selection probabilities and differential nonresponse probabilities. This implies that a relatively small number of observations can exert substantial influence on estimates, which raises the question of who these respondents with very large weights are.

Table 5. Relative frequencies of selected features in the top 1% quantile of the distribution of individual-level weights (*PHRF*) in wave 1 compared to the wave 1 realized sample.

Feature	Relative frequency	
	<i>PHRF</i> top 1% quantile (n=170)	Full sample (n=17,027)
Respondent type: household member	46.5%	23.3%
Only adult in household	46.5%	19.7%
Living in Western Germany	94.7%	64.5%
Municipality size < 20,000	21.8%	45.6%
20,000 < Municipality size < 100,000	42.9%	20.5%
Male	61.2%	49.2%
Low education (Hauptschule or less)	97.1%	19.0%
No German citizenship	50.6%	6.0%
Unemployed	8.2%	1.9%
Age < 25	5.3%	9.1%
Age > 80	1.7%	6.5%
Not intending to vote in federal elections (excl. not allowed to vote)	11.3%	4.0%
Trust in science: <5 on a 10-pt. scale	14.3%	10.9%
Subjective position on social ladder: <5 on a 10-pt. scale	26.4%	15.0%

Table 5 compares the respondents in the top 1% quantile of W1P1 *PHRF* weights against the overall sample of respondents on selected characteristics. It shows that the respondents with the highest weights are:

- Almost entirely persons living in Western Germany (former FRG), due to the oversampling of households in Eastern Germany (see section 1 on design weights)
- Exceedingly often persons in small households, also due to the sampling design (see section 1 on design weights)
- In multi-person households, exceedingly often household members rather than anchor persons, because household members pass through more stages of nonresponse (see above)
- Overwhelmingly socio-economically disadvantaged: They almost entirely have a low school degree, a majority has no German citizenship, and a large share compared to

full sample is unemployed or consider themselves on the lower end of the social hierarchy.

- Disproportionately often male
- Less trusting in science
- Less likely to vote in federal elections (if allowed to vote)

These features of respondents with high weights therefore reflect either characteristics of the SCP's sampling design (lower selection probabilities for small households and households in Western Germany, lower response probability for HMs compared to APs) or known determinants of survey nonresponse: education (e.g., Stein et al., 2026), non-citizens (e.g., Wolf et al., 2021), age (e.g., Wolf et al., 2021), economic disadvantage (e.g., Felderer et al., 2019), gender (e.g., Felderer et al., 2019), trust in science (e.g., Silber et al., 2022), and political interest (e.g., Kmetty et al., 2022).

5. Discussion

The SCP provides three cross-sectional weights: *HHRF*, *PHRF*, and *PHRF_AP*. The right choice depends on the unit and population of interest. All three adjust for unequal selection probabilities and selective nonresponse, but they differ in the level of analysis they support and in the sample they draw on.

HHRF is the appropriate weight for household-level analyses. It applies to all households that participated in a particular wave, in the sense that their anchor person responded and, consequently, had the chance to answer the household-related questions of the household questionnaire module.

PHRF serves analyses at the person level and should be used whenever all participating individuals in a wave (anchor persons, household members and new household members alike) are utilized to infer to the German population.

PHRF_AP, by contrast, restricts inference to the original random sample of anchor persons, excluding individuals recruited as household members or new household members entirely. Like *HHRF*, *PHRF_AP* also mostly has only one observation per household, but extrapolates to the population of adult individuals rather than households in Germany. Hence, *PHRF_AP* is the right choice when using only the original anchor persons to infer to individuals (rather than households) in Germany. This comes at the cost of a smaller effective sample size, but because anchor persons are recruited through the higher-quality initial sampling process and pass through fewer stages of nonresponse than household members and new household members, the original anchor persons sample carries a lower risk of nonresponse bias. This

means that first, *PHRF_AP* varies less than *PHRF*, and second, there is a lower risk of residual nonresponse bias that the nonresponse weights cannot account for. Analyses with the AP sample using *PHRF_AP* weights may therefore also serve as a robustness check for analyses based on the whole sample using *PHRF* weights.

For longitudinal analyses, *HBLEIB* and *PBLEIB* provide inverse staying probabilities at the household and person level, respectively. Because both presuppose participation in the previous wave — a respondent who skipped a wave receives a weight of zero — multiplying *HBLEIB* or *PBLEIB* weights across several waves effectively drops anyone with a temporary gap in participation from the analysis. Users interested in a panel sample with intermittent nonresponse should therefore weigh this trade-off carefully before combining these weights across multiple waves.

The SCP weights have limitations. First, nonresponse weighting reduces, but cannot fully eliminate, nonresponse bias. It can only correct for non-random nonresponse to the extent that either predictor data are available for both respondents and nonrespondents and can be successfully exploited in nonresponse models, or population margins are known and can be enforced through raking. Neither condition is well met, for example, for attitudinal variables when constructing nonresponse weights at recruitment. Thus, nonresponse at recruitment that is related to attitudes, such as trust in institutions or political views, may not be fully captured by nonresponse weights.

Second, the cross-sectional weights are defined for the broadest possible group of interest for empirical analyses: anyone who answered at least one question in a particular wave, including partial interviews (breakoffs). Two consequences follow. First, the weights are not perfectly tailored to analyses restricted to specific subgroups in the population, such as pensioners or low-income households. Especially if the response mechanism within such groups were fundamentally distinct from the response mechanism observed in the general population, this may be a problem when using the provided weights. Second, weights are not perfectly suited either for analyses restricted to only complete interviews, or to the overlap of respondents across several waves. These subsets of the sample may involve additional selective nonresponse into survey breakoff or intermittent nonresponse, which may yield additional nonresponse bias. Third, even after trimming, the cross-sectional nonresponse weights vary. This variability stems not only from the sampling design and the successive stages of selection, but also from pronounced differences in nonresponse by citizenship status and education, and from the weighting scheme's aim of accounting for all relevant stages and predictors of nonresponse. Because such variability can affect the precision of estimates, users may wish to conduct sensitivity analyses — for example, re-estimating *PHRF*-based

models with *PHRF_AP* weights — to assess whether their conclusions depend on the specific choice of nonresponse weights.

Finally, unweighted analyses, although potentially more biased, can be more efficient than weighted ones, particularly given the variability of SCP weights. Where unit or panel nonresponse is not a central concern, users may therefore reasonably forgo nonresponse weighting altogether. This may be justified, for instance, when estimates are not meant to generalize to the German population, or when there is good reason to believe that the effects under study are homogeneous across the population (Kohler et al., 2019).

References

- Burgette, L. F., & Reiter, J. P. (2010). Multiple imputation for missing data via sequential regression trees. *American Journal of Epidemiology*, 172(9), 1070-1076. <https://doi.org/10.1093/aje/kwq260>
- Felderer, B., Kirchner, A., & Kreuter, F. (2019). The Effect of Survey Mode on Data Quality: Disentangling Nonresponse and Measurement Error Bias. *Journal of Official Statistics*, 35(1), 93-115. <https://doi.org/10.2478/jos-2019-0005>
- Gerlitz, J. Y., Axenfeld, J. B., Cornesse, C., Groh-Samberg, O., Kroh, M., Lengfeld, H., Liebig, S., Minkus, L., Reinecke, J., Richter, D., Teichler, N., Trautmüller, R. & Zinn, S. (2024). The German Social Cohesion Panel (SCP): Theoretical background, instruments, survey design, and analytical potential. OSF Preprints. <https://doi.org/10.31219/osf.io/tew78>
- Groh-Samberg, O., Axenfeld, J. B., Gerlitz, J.-Y., Cornesse, C., Kroh, M., Lengfeld, H., Reinecke, J., Teichler, N., Trautmüller, R., & Zinn, S. (2026). German Social Cohesion Panel 2021-2023 – Wave 1-3. Dataset and Documentation. <https://doi.org/10.60532/scp.2021-23.w1-3.v1>
- Grund, S., Lüdtke, O., & Robitzsch, A. (2024). Multiple imputation of missing data in large studies with many variables: A fully conditional specification approach using partial least squares. *Psychological Methods*. <https://psycnet.apa.org/doi/10.1037/met0000694>
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.
- Haziza, D., & Beaumont, J.-F. (2017). Construction of weights in surveys: a review. *Statistical Science*, 32(2), 206-226. <https://doi.org/10.1214/16-STS608>
- Kmetty, Z., & Stefkovics, Á. (2022). Assessing the effect of questionnaire design on unit and item-nonresponse: Evidence from an online experiment. *International Journal of Social Research Methodology*, 25(5), 659-672. <https://doi.org/10.1080/13645579.2021.1929714>
- Kohler, U., Kreuter, F., & Stuart, E. A. (2019). Nonprobability sampling and causal analysis. *Annual Review of Statistics and Its Application*, 6(1), 149-172. <https://doi.org/10.1146/annurev-statistics-030718-104951>
- Kroh, M., Siegers, R., & Kühne, S. (2015). Gewichtung und Integration von Auffrischungsstichproben am Beispiel des Sozio-oekonomischen Panels (SOEP). In J. Schupp & C. Wolf (Eds.), *Nonresponse Bias. Schriftenreihe der ASI - Arbeitsgemeinschaft*

Sozialwissenschaftlicher Institute (pp. 409-444). Springer VS. https://doi.org/10.1007/978-3-658-10459-7_13

Kroh, M., Kühne, S., Jacobsen, J., Siegert, M., & Siegers, R. (2017). *Sampling, nonresponse, and integrated weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4) – Revised version* (SOEP Survey Papers No. 477, Series C). Deutsches Institut für Wirtschaftsforschung (DIW Berlin), SOEP.

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Robitzsch, A. and Grund, S. (2023). *miceadds: Some Additional Multiple Imputation Functions, Especially for 'mice'*. R package version 3.16-18.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Schmidt, S. C. E., & Woll, A. (2017). Longitudinal drop-out and weighting against its bias. *BMC Medical Research and Methodology*, 17, 164. <https://doi.org/10.1186/s12874-017-0446-x>

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, 22(3), 278-295. <https://doi.org/10.1177/0962280210395740>

Silber, H., Moy, P., Johnson, T. P. Neumann, R., Stadtmüller, S. & Repke, L. (2022). Survey participation as a function of democratic engagement, trust in institutions, and perceptions of surveys. *Social Science Quarterly*, 103(7), 1619-1632. <https://doi.org/10.1111/ssqu.13218>

Stein, A., Gummer, T., Naumann, E., Rohr, B., Silber, H., Auriga, R., Bergmann, M., Bethmann, A., Blohm, M., Cornesse, C., Christmann, P., Coban, M., Décieux, J-P., Gauly, B., Hahn, C., Helmschrott, S., Hochman, O., Lemcke, L., Naber, D., Pötzschke, S., Roßmann, J., Schanze, J.-L., Schmidt, T., Schneider, S. L., Spangenberg, H., Rettig, T., Trappmann, M., Weinhardt, M., & Weiß, B. (2026). Education bias in probability-based surveys in Germany: evidence and possible solutions. *International Journal of Social Research Methodology*, 29(2), 197–214. <https://doi.org/10.1080/13645579.2025.2508889>

Steinwede, A., Kemmerling, J., von der Burg, K. Kleudgen, M. & Aust, N. (infas) (2024). SCP 2021/22 W1 Supplement: Methodenbericht German Social Cohesion Panel - 1. Erhebungswelle 2021. German Social Cohesion Panel 2021/22 - Wave 1. Bremen and Berlin: RDC-RISC/SOCIUM, SOEP/DIW Berlin.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45, 1-67. <https://doi.org/10.18637/jss.v045.i03>

Wolf, C., Christmann, P., Gummer, T., Schnaudt, C., & Verhoeven, S. (2021). Conducting general social surveys as self-administered mixed-mode surveys. *Public Opinion Quarterly*, 85(2), 623-648. <https://doi.org/10.1093/poq/nfab039>