

## The relationship between response rates and nonresponse bias – Regularities over time and over the data collection period

Declining response rates are a continuing problem for household surveys in many Western countries (Atrostic et al. 2001; de Leeuw and de Heer 2002; Dixon and Tucker 2010). The German General Social Survey (ALLBUS) – which has been conducted every other year since 1980 – also has been facing an increase in nonresponse in the past decades. Between 1994 and 2014, for instance, the response rate of ALLBUS decreased from 54% to 35%. The main reason for this decline was a rise in the number of refusals.

ALLBUS is a face-to-face survey of the adult population in Germany, covering a wide range of topics and aiming at charting the long-term trends in attitudes and behaviour (<http://www.gesis.org/en/allbus>). The decrease in response rates raised concerns whether the increase in nonresponse rates would be accompanied by an increase in nonresponse bias.

ALLBUS provides a good opportunity to analyze this relationship, as ALLBUS has used an identical study design since 1994 (target population of adults living in private households; samples of named individuals; 3.500 completed interviews; interview duration around 70 minutes; consistent calculation of response rates). In addition to analyzing this relationship over time, it is also possible to investigate the relationship using individual ALLBUS surveys and looking at changes in nonresponse bias over the course of the data collection period.

### Research questions

Using ALLBUS data, we investigate the following issues concerning the relationship between response rates and nonresponse bias:

- a) Is there a (negative) relationship between the response rates of the ALLBUS surveys and nonresponse bias over time?
- b) Does this relationship differ for different survey variables?
- c) Can any regularity be observed with respect to nonresponse bias in the course of the data collection period of individual ALLBUS surveys?

### Methods

To assess the quality of the ALLBUS surveys and to arrive at an indicator of nonresponse bias, the distributions of several demographic variables were compared with the respective data from official statistics. This is a straightforward and frequently used method for analyzing nonresponse bias (see Hartmann 1990, Hartmann & Schimpl-Neimanns 1992, Häder & Häder 1997, Koch 1998). The data from the German microcensus survey can be used as a valid external criterion to evaluate the net samples from ALLBUS, since participation in the microcensus is mandatory and its unit nonresponse rate is only about three percent. Comparisons of microcensus data with data from other official sources (like Census data, and data on educational achievement) confirm the good quality of the microcensus data (Hartmann, 1990). A drawback of the comparison with the microcensus is that only sociodemographic variables can be analyzed.

We use three different measures to compare the distributions of ALLBUS and the German microcensus for seven different variables:

- a) Index of dissimilarity for variable  $j$  with  $i$  categories  $D_j = \frac{1}{2} \sum_{i=1}^n |S_i - B_i|$
- b) Mean index of dissimilarity for  $v$  variables with  $i$  categories  $D = \left( \sum_{j=1}^v \frac{1}{2} \sum_{i=1}^n |S_i - B_i| \right) / v$

*For both a) and b) the following variables and categorizations were used: Gender (M/F); age (10-year age categories); education (3 categories: low/medium/high); occupational status (5 Cat.); marital status (married: Y/N); household size (1/2/3/4/5+ persons); in paid work (17 hours or more a week: Y/N)*

- c) (Multivariate) population based r indicator, using the following variables and categorizations: Gender (M/F); age (-40/40-60/60+ years); education (low vs. high); marital status (married: Y/N); in paid work (0-16 hours/17+ hours)

$$\hat{R}_r(\mathbf{p}) = 1 - 2\sqrt{\frac{1}{N-1} \sum_r d_i \hat{\rho}_i^{-1} (\hat{\rho}_i - \hat{\rho}_r)^2}$$

In what follows, we analyze the relationship between nonresponse rates and nonresponse bias from the two perspectives mentioned above (over the years from 1994 to 2014 AND within individual ALLBUS surveys looking at interim results over the data collection period).

## Results

### Nonresponse bias over the years

The mean index of dissimilarity across the seven variables remains quite stable over time, despite the response rates of ALLBUS were decreasing considerably (figure 1a). Figure 1b shows the relation between the response rate and the mean index of dissimilarity for 10 ALLBUS surveys between 1994 and 2014. Contrary to the usual expectation, we find a weak positive relationship between the response rate and the mean deviation from the external benchmark data.

Figure 1a: Response rates and Mean Index of Dissimilarity, by survey year (ALLBUS 1994-2014)

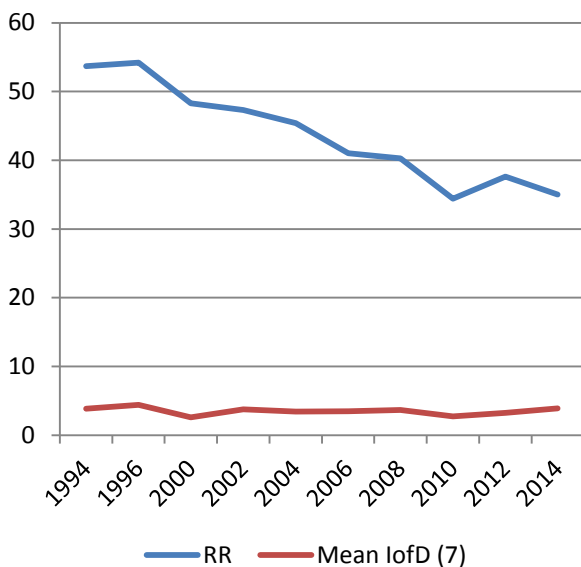
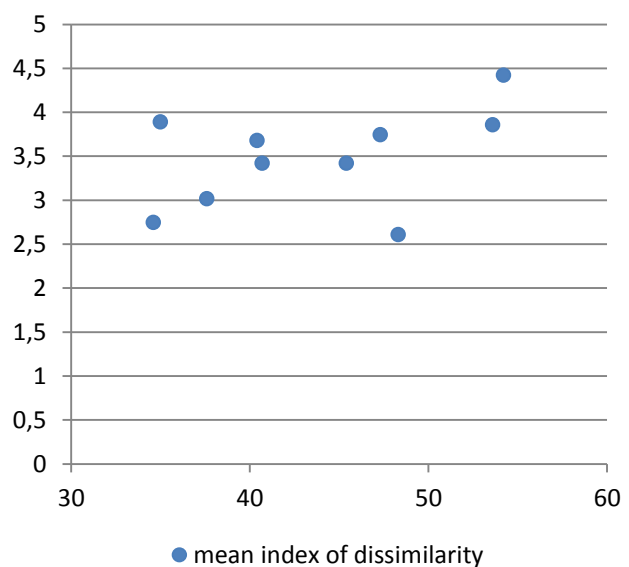
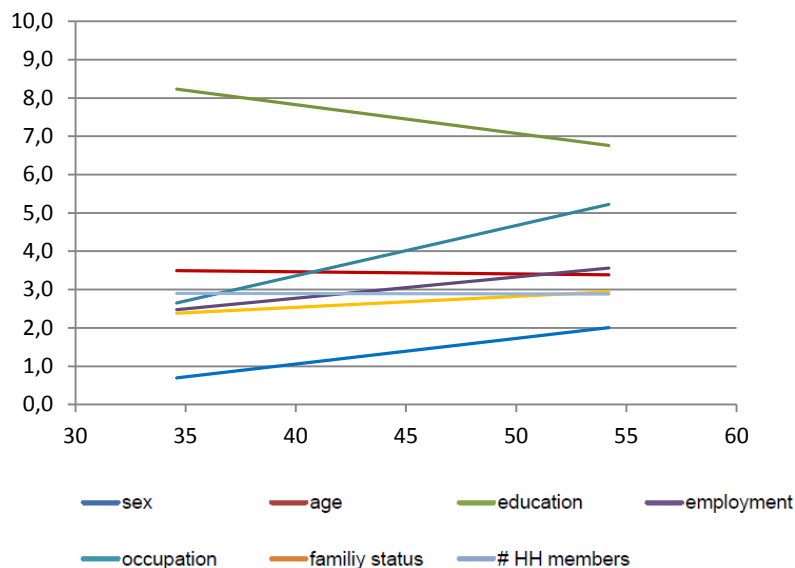


Figure 1b: Mean Index of Dissimilarity, by response rate (ALLBUS 1994-2014)



When we examine the 7 variables individually (figure 2), a positive relationship can be found for the variables gender, employment status, occupational status and marital status. For the variable household size neither a positive nor a negative relationship can be observed. Only for the variables age and education the relationship is negative (as common wisdom would expect).

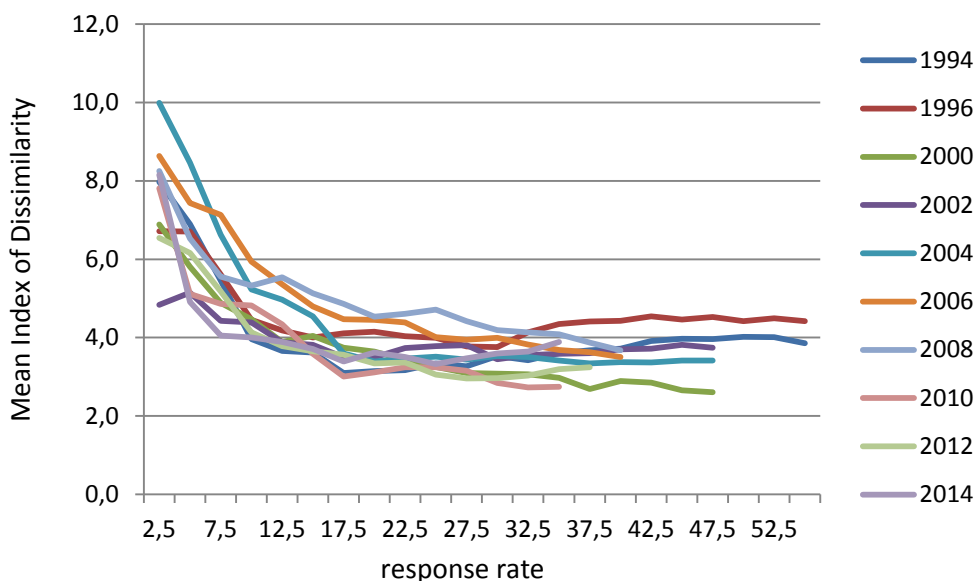
Figure 2: Mean Index of Dissimilarity (linearized) for individual variables, by Response rates (ALLBUS 1994-2014)



#### Nonresponse bias over the data collection period

Usually one would expect that nonresponse bias will become smaller over the data collection period of a survey, as more sample units are worked on and a higher response rate has been achieved. In order to analyze this relationship, we computed the indicators of nonresponse bias for interim stages of the individual ALLBUS surveys. For most ALLBUS surveys we can observe the expected pattern when we look at the mean index of dissimilarity (figure 3). The deviations from the microcensus become smaller at later points in data collection. However, we have to be aware that there is only a slight change in the size of the mean deviation once a response rate of around 20% has been achieved. In addition, in a few ALLBUS surveys from the 1990ies, the deviations even increase after having achieved a response rate of around 35% to 40%.

Figure 3: Mean Index of Dissimilarity, by response rate<sup>1</sup>, by year of survey



<sup>1</sup> Response rate = number of completed interviews at a certain point of time / gross sample size.

When we look at individual variables, different patterns can be observed. Low educated people become increasingly underrepresented during the data collection period. At the end of the data collection period, they are underrepresented in each ALLBUS survey (figure 4a). For the percentage of employed persons, it is just the other way round (figure 4b). In the course of the data collection period the percentage of employed people is increasing. This increase means, that in some ALLBUS surveys the direction of deviation changes from an underrepresentation of people in paid work at the beginning of the data collection period to an overrepresentation once data collection has been finalized.

Figure 4a: Deviation from the microcensus for “Low education”, by response rate, by survey year (ALLBUS 1994 -2014)

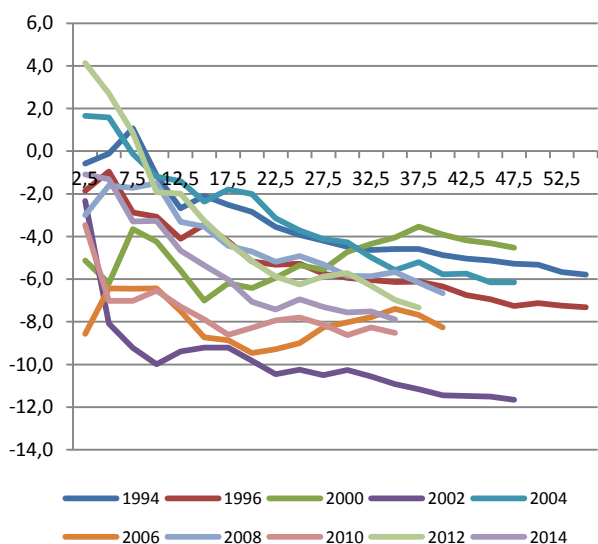
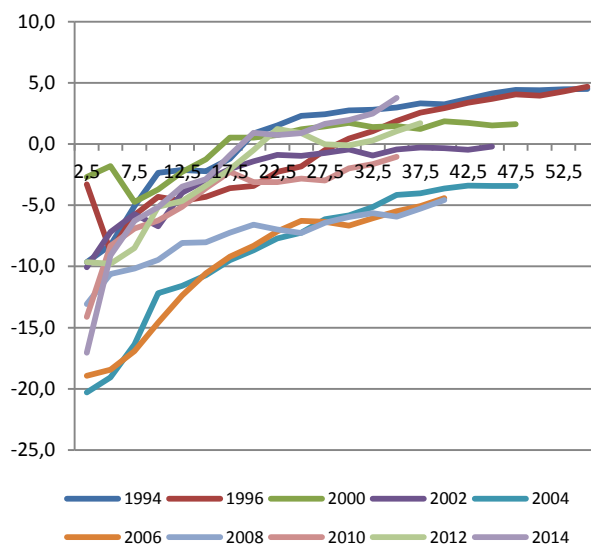
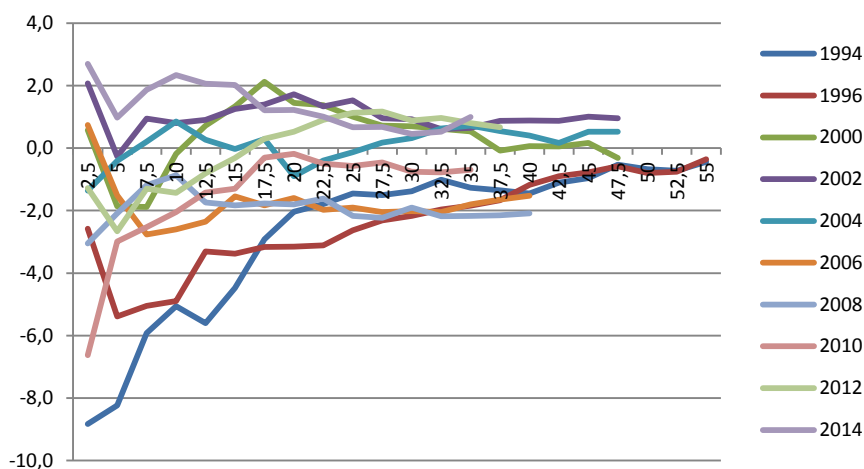


Figure 4b: Deviation from the microcensus for “persons in paid work more 17hrs week”, by response rate, by survey year (ALLBUS 1994 -2014)



For other variables, like age for instance, such a prediction is not possible (figure 4c; here percentage of respondents under 30 years). Whereas in some ALLBUS surveys this age group has been underrepresented at the beginning of data collection, in other surveys, this age group has been overrepresented at the early stages of fieldwork. In all surveys, however, the deviations become smaller and are very close to the results of the micro census at the end of the data collection period.

Figure 4c: Deviation from the microcensus for “people less 30years”, by response rate, by survey year (ALLBUS 1994 -2014)



Population based r-indicators provide an alternative way to analyze nonresponse bias by way of comparison with external benchmark data. Population based r indicators allow to analyze several variables simultaneously. In figure 5a the population based r-indicators for 3 different combinations of variables are displayed. For the sake of comparison, the mean index of dissimilarity is shown in figure 5b for the same variables as used in figure 5a.

Figure 5a: three different population based r indicators, by response rate (completed cases), (ALLBUS 2012)

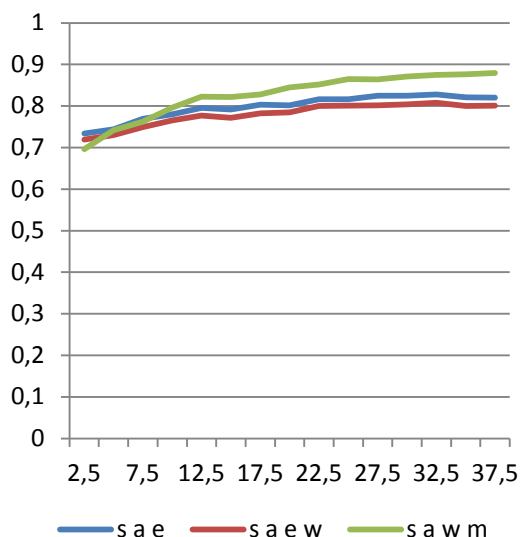
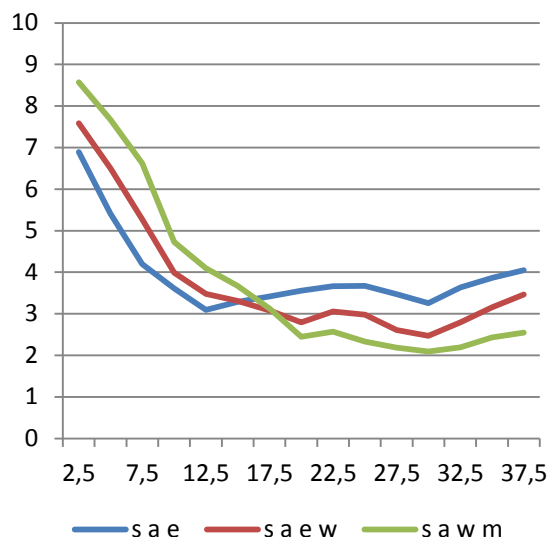


Figure 5b: Mean Index off dissimilarity for the same variables as used in figure 5a, by realized number of cases, (ALLBUS 2012)



s = sex / a = age / e = education / w = work/employment status / m= marital status

Both figures show the same main message: the deviations of the ALLBUS samples from the external benchmark data become smaller in the course of data collection. Near the end of the field time, however, a small increase in the size of the deviations can be observed (except for the r-indicator for the combination of s a w m).

## Summary

- No indication that the decrease in ALLBUS response rates over the past 20 years comes along with a decline in the quality of sample composition.
- Looking at data collection processes for individual ALLBUS surveys, sample composition becomes better until a response rate of around 20-25% has been reached. Beyond this level only marginal changes can be observed. In a few ALLBUS surveys from the 1990ies, the deviations even increase after having achieved a response rate of around 35% to 40%.

## Points for discussion

- We found a few regularities in sampling composition over the data collection period. Is there any evidence in this respect from other studies?
- Sample composition for some variables is getting better over the data collection period, for others it becomes worse. How to cope with that issue, if the aim is to achieve a more balanced sample composition at the end of the data collection period?
- Should indications of bias at an early point of time in the data collection period prompt more active fieldwork management? How should this be implemented?
- Are there any other experiences in the calculation of population based r-indicators over the data collection period?