

# Sociodemographic variables in surveys

## Increasing research potential through output harmonization

Silke Schneider<sup>1</sup> and Lennart Palm<sup>1</sup>

<sup>1</sup> GESIS - Leibniz-Institute for the Social Sciences

Measuring people's sociodemographic characteristics over time and in different contexts or studies is one of the keys of quantifying social inequalities and social change within any society. While different respondent-based studies usually focus on different topics, almost all of them collect data on some key features defining respondent's backgrounds and living situations.

Summarizing the measurement of sociodemographic characteristics however, the German survey landscape could be characterized as "same but different": studies measure the same concepts yet differ in their specific approach, despite the fact that the "Demographische Standards" (Hoffmeyer-Zlotnik et al., 2016) provide recommended questionnaire items since the early 1990s. This has several problematic implications: Firstly, it becomes quite difficult to combine knowledge about groups of respondents with the same sociodemographic characteristics across studies on a macro level, and thus to systematically accumulate knowledge on social inequalities and social change. Secondly, combining different data sets, e.g. in order to be able to analyze small groups or rare phenomena, can become a laborious effort which might still produce questionable results, since key caveats can be overlooked in survey documentation. Combining supposedly basic categories such as marital status across studies can already make simplification and therefore loss of information necessary, as there are multiple ways of measuring marital status as Table 1 shows.

**Table 1:** Overview of (proposed) measurements of marital status in selected studies by Schneider et al (2022)<sup>1</sup>

Legally possible categories of marital status in Germany	Categories in Dem. Standards, ALLBUS, G Panel, FReDA, GIP	Categories in MZ, SOEP	Categories in GLES, Covid-19-U	Categories in GEDA, NEPS	Categories in Best_FDM <sup>2</sup> (minimal)
Married, living with spouse	X	X*	X		
Registered partnership, living with partner	X	X**	X	X	X
Married, not living with spouse	X	X*	X	X	
Registered partnership, not living with partner	X	X**	X		
Divorced	X	X	X	X	X
Registered partnership, dissolved	X	X			
Widowed	X	X	X	X	X
Registered partner passed	X	X			
Single	X	X	X	X	X

ALLBUS: Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. <https://www.gesis.org/en/allbus/allbus-home>

Best\_FDM: connecting empirical social science research in societal crises. <https://wzb.eu/en/research/trans-sectoral-research/konsortswd/projects/bestfdm-vernetzung-empirischer-sozialwissenschaftlicher-forschung-in-gesellschaftlichen-krisen>

Covid-19-U: COVID-19 and Inequality - Survey Program. <https://www.exc.uni-konstanz.de/en/inequality/research/covid-19-and-inequality-surveys-program/>

Dem. Standards: Demographic Standards. [https://www.statistischebibliothek.de/mir/receive/DE-Monografie\\_mods\\_00003695](https://www.statistischebibliothek.de/mir/receive/DE-Monografie_mods_00003695)

FReDA: The German Family Demographic Panel Study. <https://www.freda-panel.de/FReDA/EN/Startseite.html>

G Panel: GESIS Panel. <https://www.gesis.org/en/gesis-panel/gesis-panel-home>

GIP: German Internet Panel. <https://www.uni-mannheim.de/en/gip/>

GLES: German Longitudinal Election Study. <https://www.gesis.org/wahlen/gles>

MZ: Microcensus. <https://www.gesis.org/en/missy/metadata/MZ/>

NEPS: National Educational Panel Study. <https://www.neps-data.de/Mainpage>

SOEP: Socio-Economic Panel. [https://www.diw.de/de/diw\\_01.c.412809.de/sozio-oeconomisches\\_panel\\_soep.html](https://www.diw.de/de/diw_01.c.412809.de/sozio-oeconomisches_panel_soep.html)

*\*, \*\*: MZ and SOEP don't differentiate between spouses/partners living together or not in their first marital status item, yet do so later on. Same number of stars indicate same category in marital status item.*

It would be desirable if harmonization of basic sociodemographic concepts was done once and for all to use. To facilitate output harmonisation of socio-demographic variables, KonsortSWD develops proposals for standard variables for selected socio-demographic attributes.

<sup>1</sup> Schneider et al (2022): <https://zenodo.org/record/6810973#.ZBhhUoSZOU>

<sup>2</sup> These are recent recommendations for questionnaire items by RatSWD, which were therefore not included in the review by Schneider et al (2022). For more details see: <https://zenodo.org/record/6810973#.ZBhhUoSZOU>

While our approach was based on the German survey data landscape, we leaned upon international standards such as ISCED for education or the EU's standardised key social variables for household net income or main activity status, to ensure international compatibility. To provide for the quality and usefulness of the proposed standard variables, we used three methods: Firstly, before developing our proposals, we reviewed existing survey instruments in several of Germany's leading studies (<https://doi.org/10.5281/zenodo.6810973>). Secondly, the proposals were discussed in a virtual roundtable meeting with researchers, study representatives and data users, as well as bilaterally with individual experts. Thirdly, based upon a multiple linear regression analysis approach, we validated our proposals both in a data-driven and a theory-driven way, using a broad set of up to 190 potential outcome variables. Based on these validation results and the feedback gained, we have refined our proposals.

Ideally, our proposals would be published within scientific use files of individual studies for easy access and usage.

Additionally, we hope our proposals could help implement an understanding for time constant variables and their specific needs in data collection when unavoidable changes to questionnaire items, which are repeatedly necessary to keep up with societal changes and respondent's life's and understanding, happen. Another benefit of this path of harmonisation could also be greater independency from survey mode switches due to technological changes and technological use patterns within societies if data needs of our proposals are considered when mode switches occur.

In this talk, we will showcase the standard variables for 3 socio-demographic attributes, namely education, marital status and main activity status to collect final feedback. Our proposals will be published in autumn 2023.

## Competing interests

The authors declare that they have no conflict of interest.

## References

1. Hoffmeyer-Zlotnik, J. H. P., Beckmann, K., Glemser, A., Heckel, C., von der Heyde, C., Schneider, S. L., Hanefeld, U., Herter-Eschweiler, R., & Kühnen, C. (2016). *Demographische Standards, Ausgabe 2016: Eine gemeinsame Empfehlung des ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und des Statistischen Bundesamtes*. Statistisches Bundesamt. [https://www.statistischebibliothek.de/mir/servlets/MCRFileNo-deServlet/DEMonografie\\_derivate\\_00001549/Band17\\_DemographischeStandards1030817169004.pdf](https://www.statistischebibliothek.de/mir/servlets/MCRFileNo-deServlet/DEMonografie_derivate_00001549/Band17_DemographischeStandards1030817169004.pdf)