INSTITUT FÜR ARBEITSMARKT- UND BERUFSFORSCHUNG
Die Forschungseinrichtung der Bundesagentur für Arbeit

A DEEPER LOOK INTO EDUCATION BIAS IN WEB SURVEYS

Joint Statistical Meeting, Portland, OR

August, 4th, 2024

Mustafa Coban
Christine Distler
Mark Trappmann

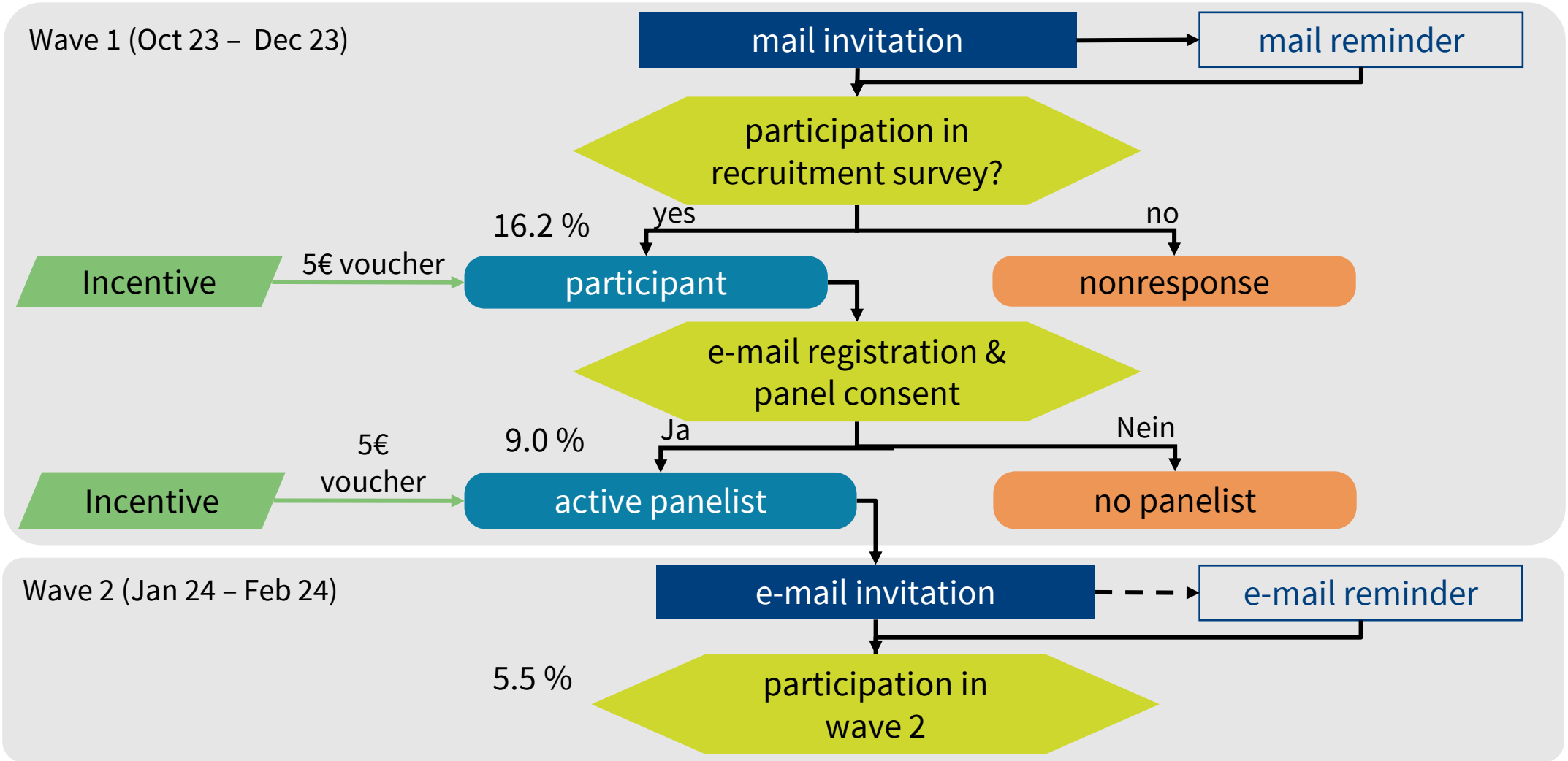
BACKGROUND

- IAB-OPAL started in 2023
  - New probability online panel for labour market research
  - Offline recruitment from admin data of the Federal Employment Agency
    - Integrated Employment Biographies (IEB): based on obligatory employer notifications for social insurance and unemployment benefit / insurance administration
    - Excellent coverage of target population: labour force minus civil servants & self-employed
- This sampling frame contains information usually not available for nonrespondents
  - Education, nationality, age
  - Employment history: times of unemployment, benefit receipt, occupational status, daily wage, ...

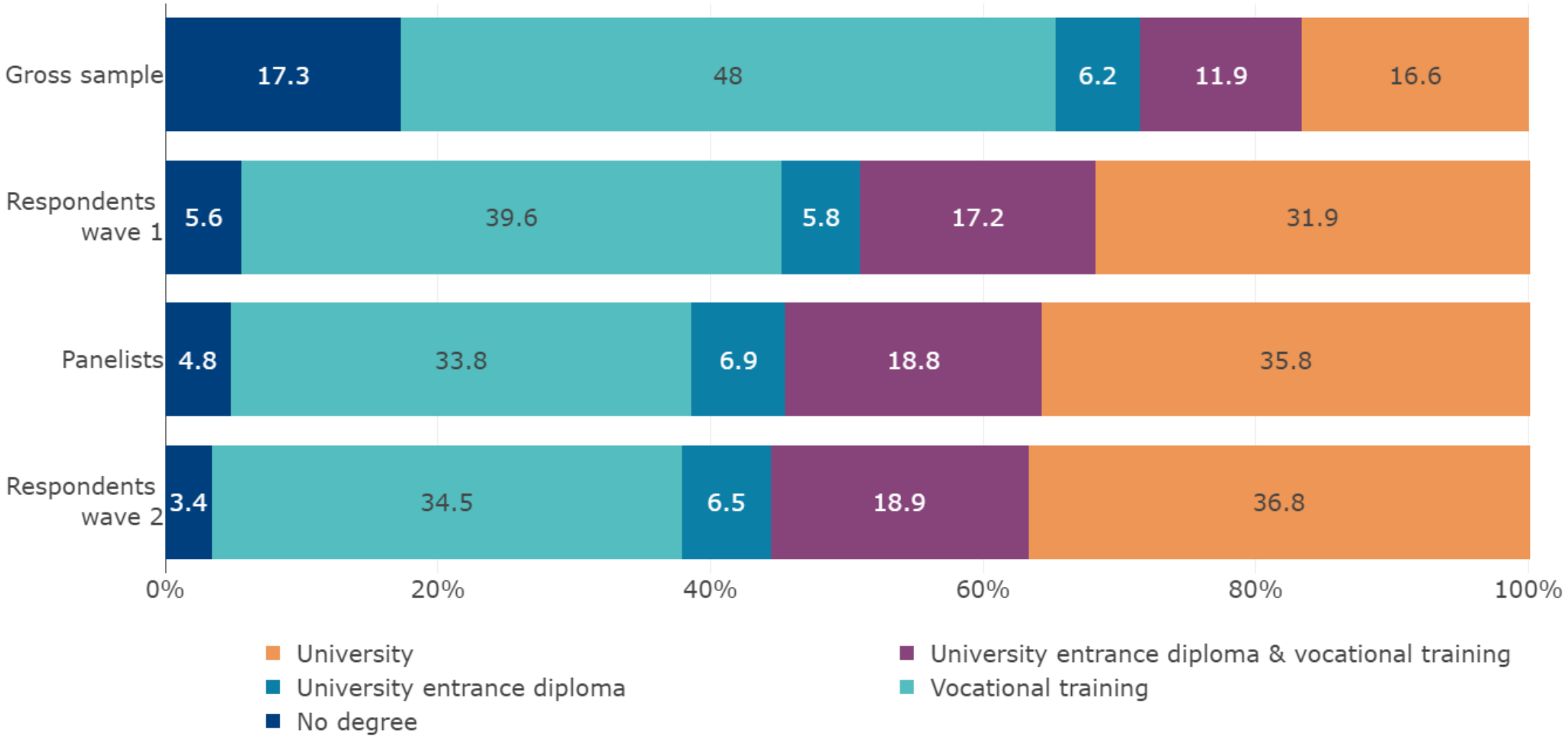
RESEARCH QUESTION

- Focus on education bias
  - Education bias frequently documented in all survey modes, particularly in web surveys (e.g. Schnell 1997, Luijkx et al. 2021, Stein et al. 2024)
  - Most web surveys use (weighting) adjustments via calibration/post-stratification to known population totals/proportions from benchmark statistics
    - Usually require MAR (Rubin 1976) mechanism
    - Access to usually unobserved variables allows us to test this assumption
- RQ1: How do the different stages of the recruitment process for an online panel contribute to education bias?
- RQ2: Are there specific subgroups within the low educated who are even less likely to participate?
- RQ3: Are there interaction effects between education and other predictors of nonresponse?

RECRUITMENT FOR IAB-OPAL



EDUCATION DISTRIBUTION AT DIFFERENT STAGES



ESTIMATION METHOD: PROBIT MODEL

Probit model: $\mathbb{P}(y = 1|\boldsymbol{x}', A) = \Phi(\boldsymbol{x}', A)$

(1) Model 1: No interaction effects

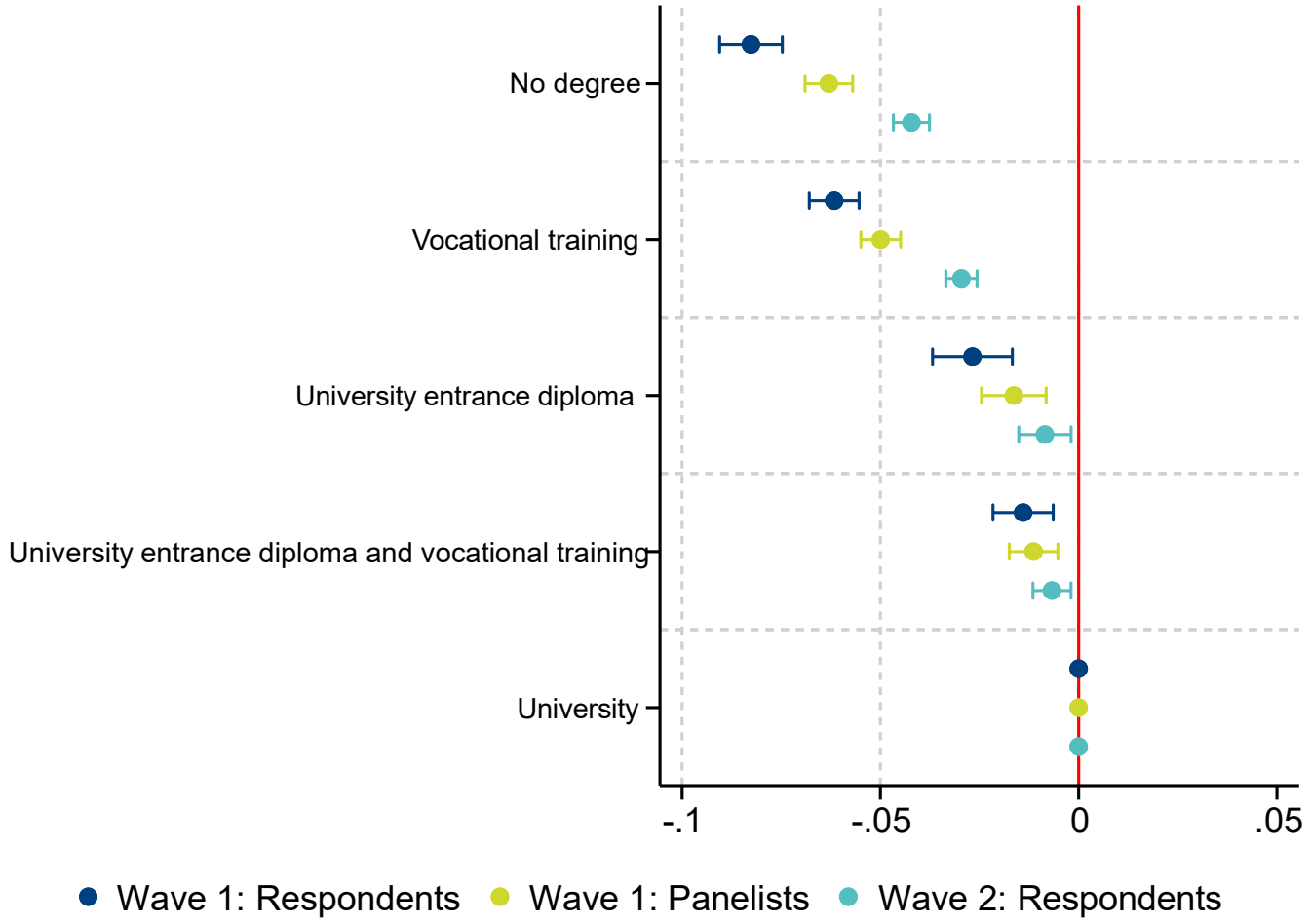
$$y^\star = \boldsymbol{x}'\boldsymbol{\beta} + \gamma \cdot education + \epsilon \quad , y = 1[y^\star > 0]$$

(2) Model 2: interaction effects with education

$$y^\star = \boldsymbol{x}'\boldsymbol{\beta} + \gamma \cdot education + (education \times \boldsymbol{x}')\boldsymbol{\delta} + \epsilon \quad , y = 1[y^\star > 0]$$

- Dependent variables: Participation in recruitment wave / active panelist / wave 2 response
- Explanatory variables from sampling frame:
  - Socio demographic: education, sex, age, nationality
  - Employment: employment status, occupational status, daily wage, supervisory function
  - Unemployment experience: benefit receipt, unemployment duration
  - Other controls: region, printing batch, shipping batch

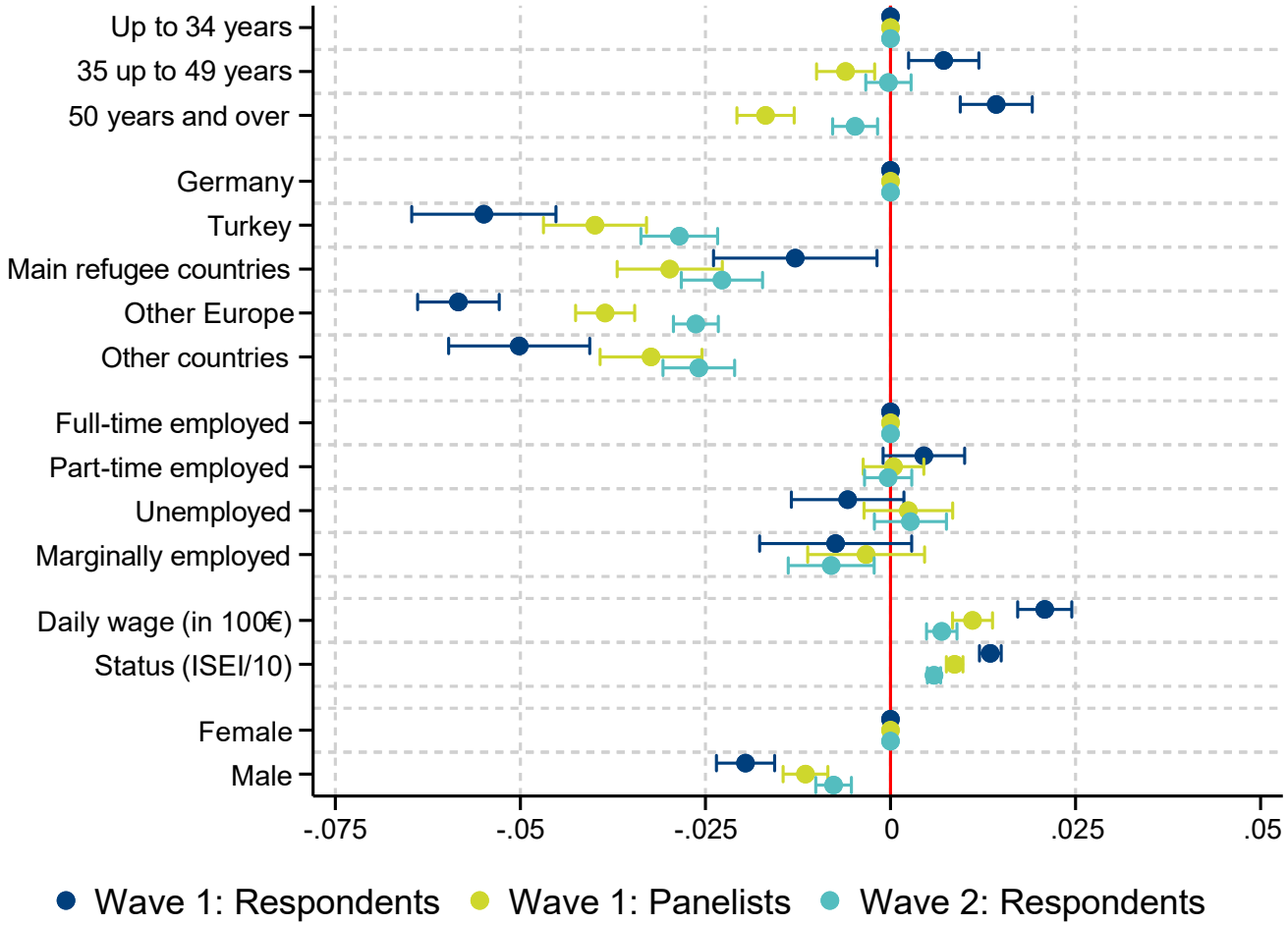
MODEL 1: AVERAGE MARGINAL EFFECTS (AME) OF EDUCATION



Notes: 95% confidence intervals; all control variables taken into account; panelists = respondents of wave 1 with panel consent and registration on the survey portal.

Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 115,610

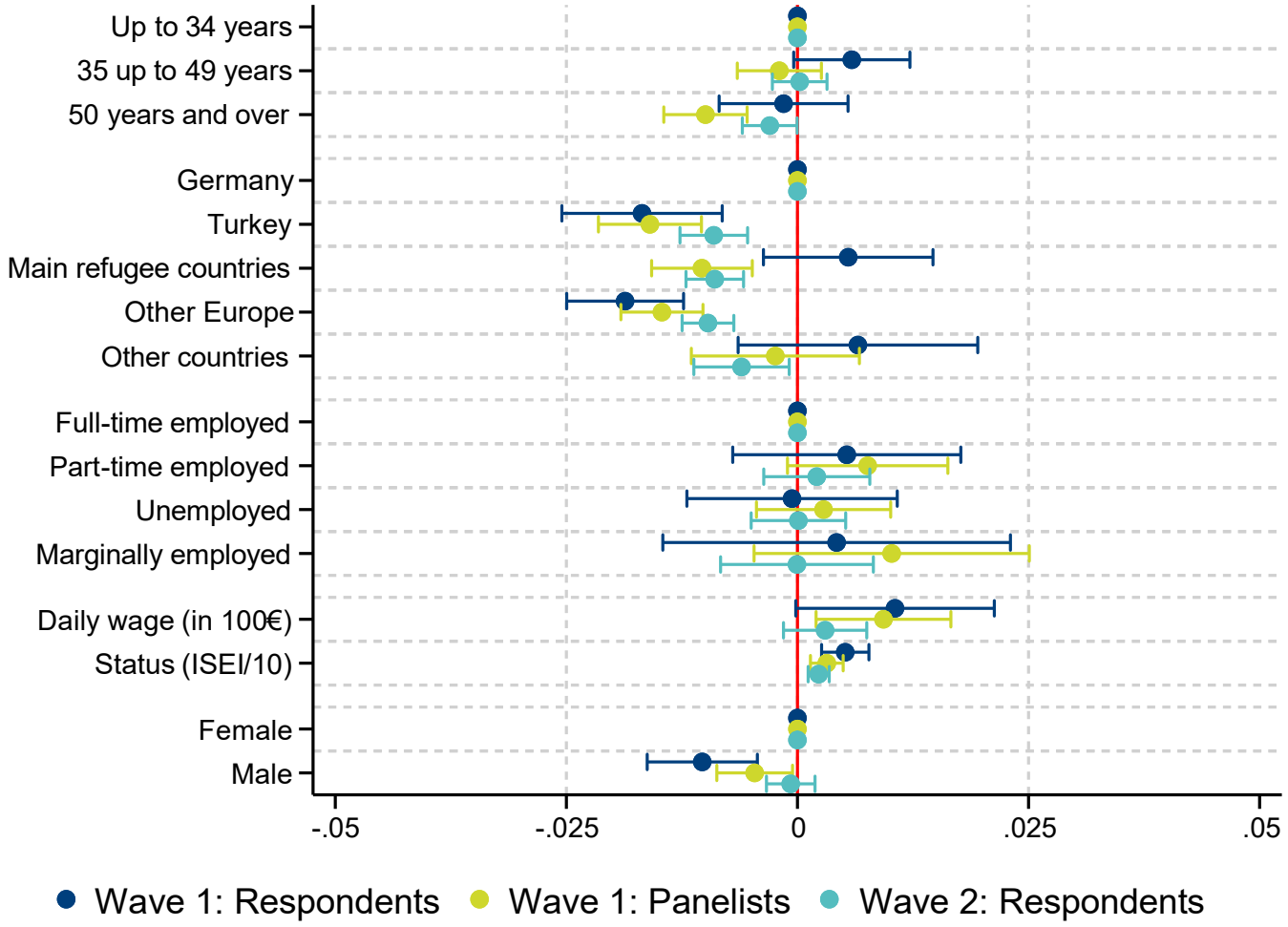
MODEL 2: AME OF OTHER PREDICTORS



Notes: 95% confidence intervals; all other control variables taken into account; panelists = respondents of wave 1 with panel consent and registration on the survey portal.

Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 115,610

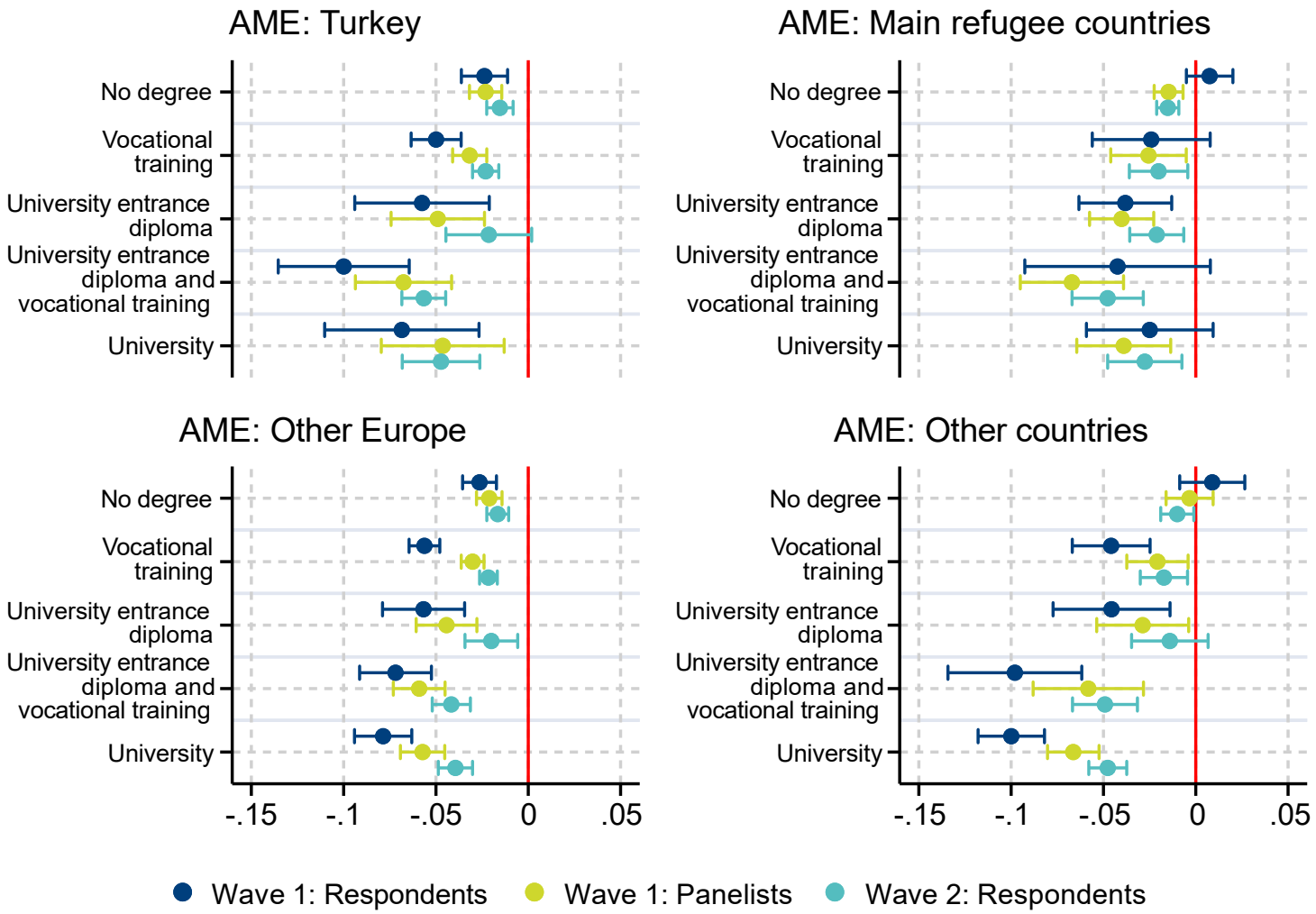
MODEL 1: AME OF OTHER PREDICTORS FOR LOWEST EDUCATION



Notes: 95% confidence intervals; all other control variables taken into account; panelists = respondents of wave 1 with panel consent and registration on the survey portal.

Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 19,995

MODEL 2: SELECTED INTERACTION EFFECTS: NATIONALITY × EDUCATION

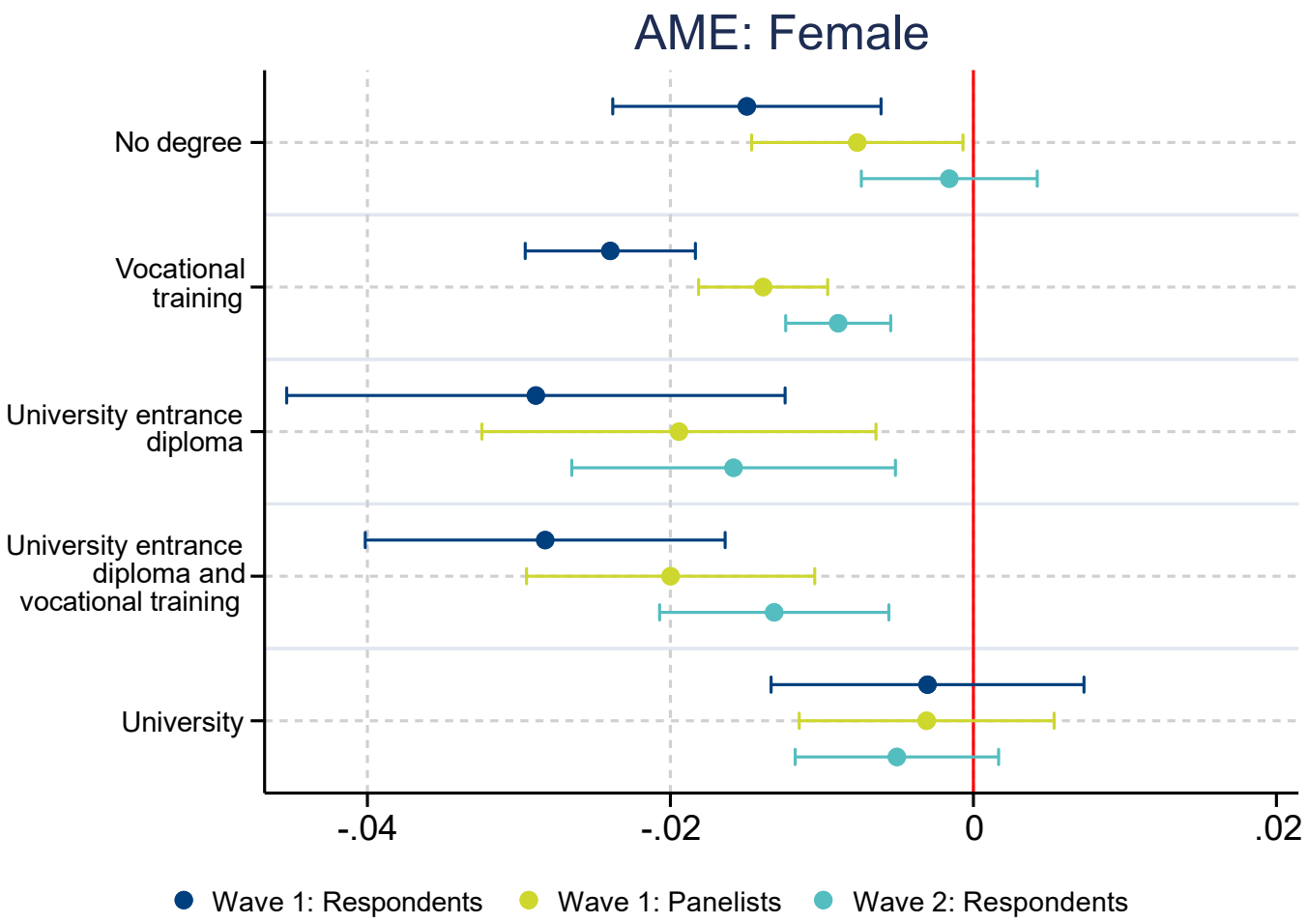


Notes: 95% confidence intervals; all other control variables taken into account; panelists = respondents of wave 1 with panel consent and registration on the survey portal.

Reference category: Germany

Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 115,610

MODEL 2: SELECTED INTERACTION EFFECTS: SEX × EDUCATION

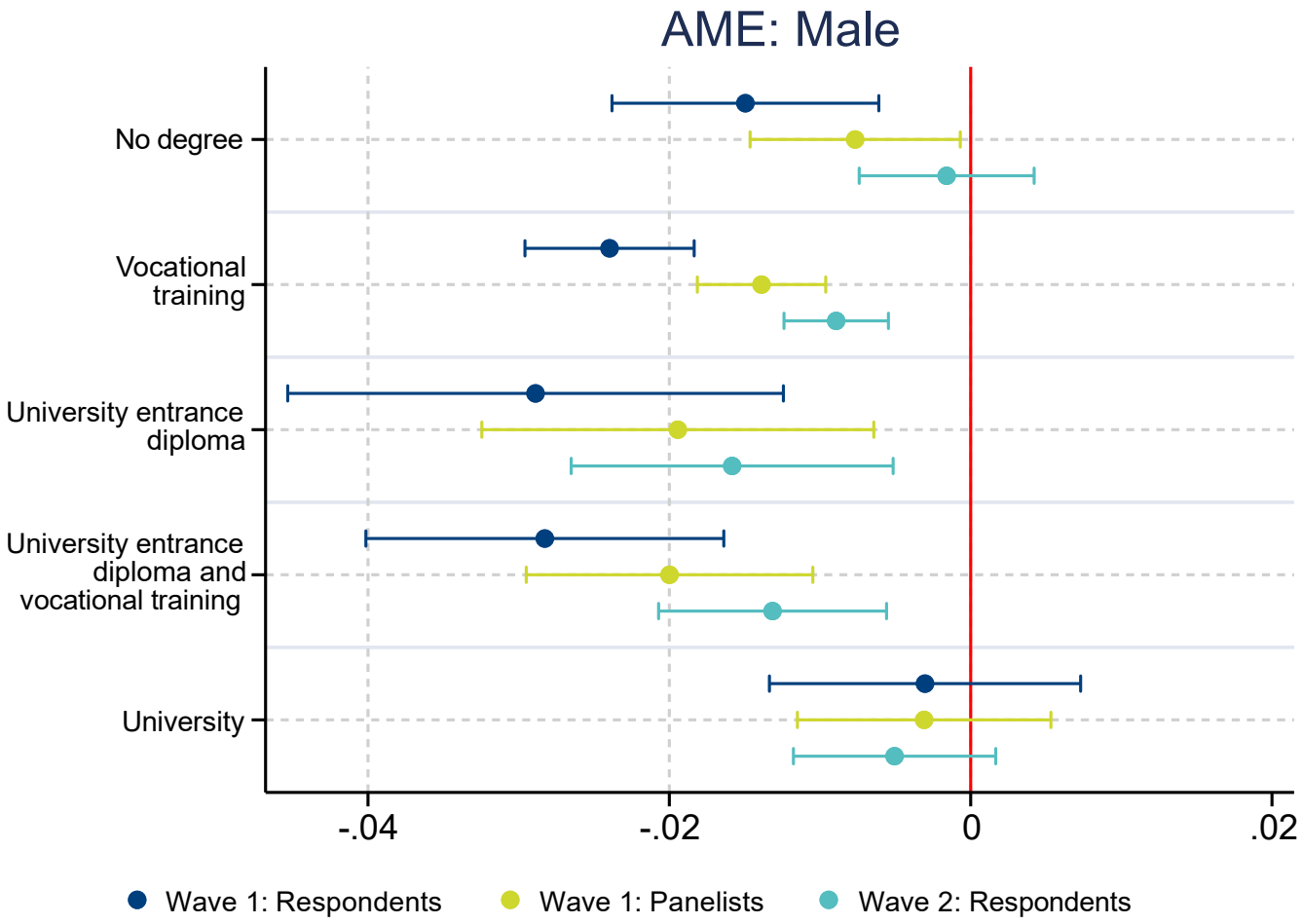


Notes: 95% confidence intervals; all other control variables taken into account; panelists = respondents of wave 1 with panel consent and registration on the survey portal.

Reference category: Male

Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 115,610

MODEL 2: SELECTED INTERACTION EFFECTS: SEX × EDUCATION



Notes: 95% confidence intervals; all other control variables taken into account; panelists = respondents of wave 1 with panel consent and registration on the survey portal.

Reference category: Female

Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 115,610

Conclusion

- Education bias cumulates at every stage of the recruitment process
- Unit-Nonresponse among low educated is stronger for
  - men
  - foreign nationals
  - Low wage and low occupational status
- Nationality plays a stronger roll for the highly educated
- Men less likely to participate unless high education
- Older respondents with high education more likely to participate, but less likely to become active panelists than their younger counterparts
- Implications
  - Strongest interaction effects with demographics: advice to use multidimensional margins for calibration adjustments
  - Use frame data to inform adaptive designs countering cumulative loss of low educated

THANK YOU!

Mark Trappmann
Institute for Employment Research (IAB)
University of Bamberg
E-Mail: mark.trappmann@iab.de

REFERENCES

Luijkx, R., Jónsdóttir, G. A., Gummer, T., Ernst Stähli, M., Frederiksen, M., Ketola, K., Reeskens, T., Brislinger, E., Christmann, P., Gunnarsson, S. Þ., Hjaltason, Á. B., Joye, D., Lomazzi, V., Maineri, A. M., Milbert, P., Ochsner, M., Pollien, A., Sapin, M., Solanes, I., … Wolf, C. (2021). The European Values Study 2017: On the Way to the Future Using Mixed-Modes. *European Sociological Review*, *37*(2), 330–346. https://doi.org/10.1093/esr/jcaa049

Rubin, D.R. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592, https://doi.org/10.1093/biomet/63.3.581

Schnell, R. (1997). Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen. Leske + Budrich.

Stein, A. et al. (2024). Education Bias in 56 large-scale probability-based general population surveys in Germany. Unpublished Manuscript.

APPENDIX

EDUCATIONAL BIAS ACROSS THE RECRUITMENT PROCESS

Recruitment step	No degree	Vocational training	University entrance diploma	University entrance diploma & vocational training	University
Absolute Bias (in percentage points)					
Respondents in wave 1	-11.7	-8.4	-0.4	5.3	15.3
Panelists	-12.5	-14.2	0.7	6.9	19.2
Respondents in wave 2	-13.9	-13.5	0.3	7	20.2
Relative Bias (in percentage)					
Respondents in wave 1	-67.6	-17.5	-6.45	44.5	92.2
Panelists	-72.3	-29.6	11.3	58	116
Respondents in wave 2	-80.3	-28.1	4.84	58.8	122

Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 115,610

METHOD

Probit: Average Marginal Effects

(1) Continuous variable X_j:

$$\tau_{X_j} = \frac{\partial \mathbb{P}(y = 1|\boldsymbol{x}')}{\partial X_j}$$

(2) Discrete variable X_j:

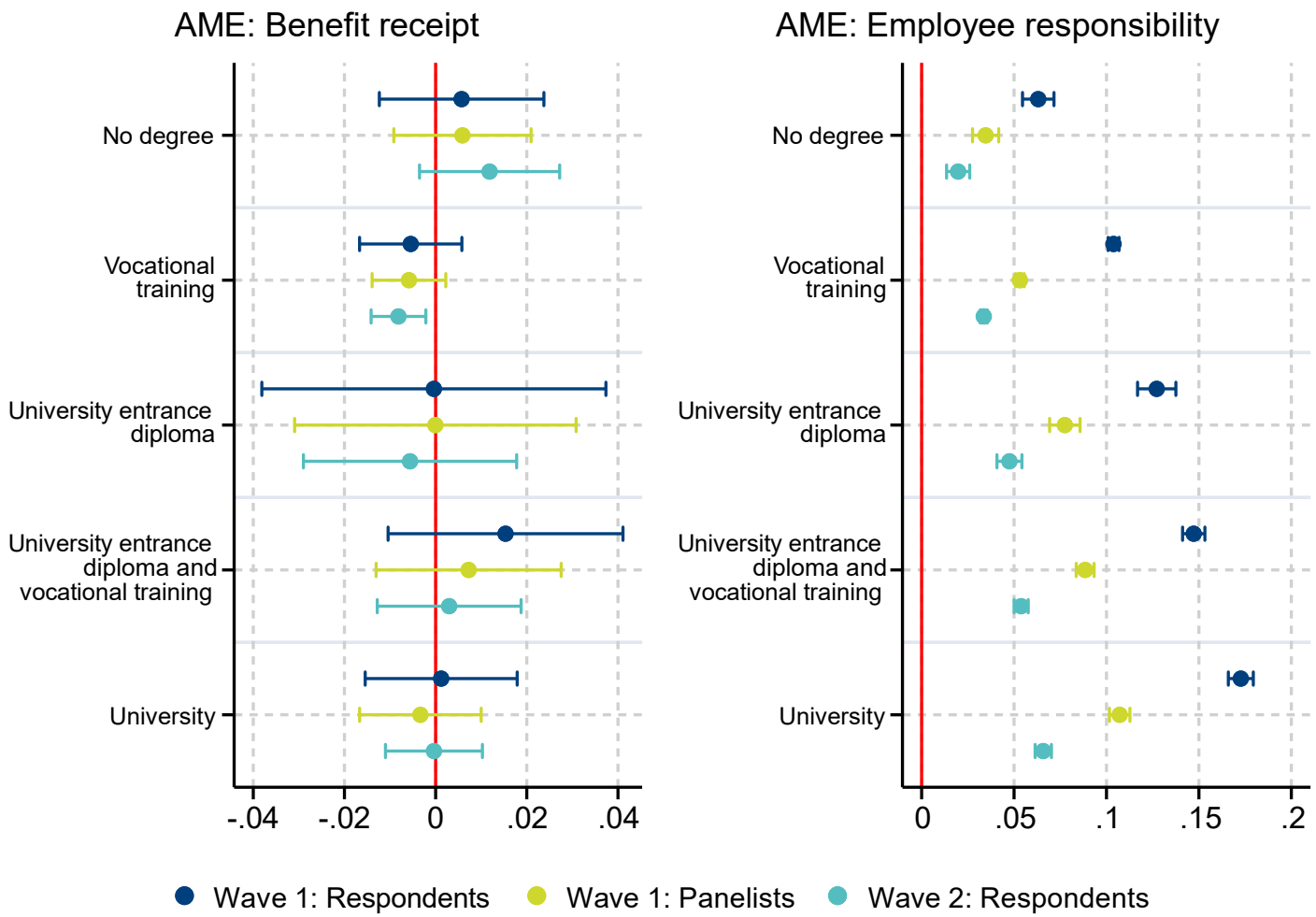
$$\tau_{X_j} = \mathbb{P}(y = 1|\boldsymbol{x}')|_{x_j=1} - \mathbb{P}(y = 1|\boldsymbol{x}')|_{x_j=0}$$

(3) Discrete variable X_j conditioned on discrete variable X_k :

$$\tau_{X_j} = \mathbb{P}(y = 1|\boldsymbol{x}')|_{x_j=1,x_k=d} - \mathbb{P}(y = 1|\boldsymbol{x}')|_{x_j=0,x_k=d}$$

... with $d \in \mathbb{D}_{x_k}$

MODEL 2: SELECTED INTERACTION EFFECTS: EDUCATION × BENEFIT RECEIPT, EDUCATION × SUPERVISORY FUNCTION

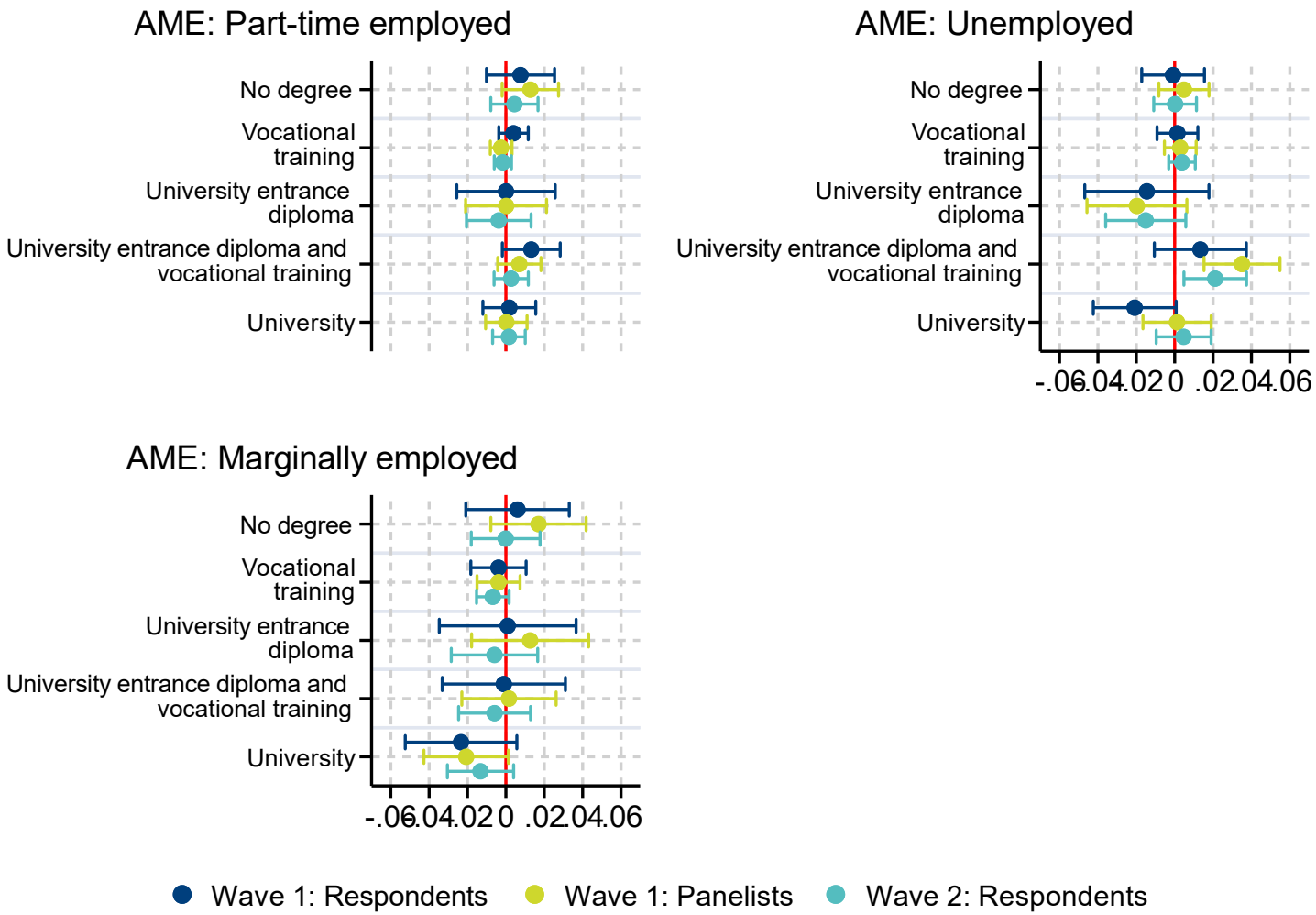


Notes: 95% confidence intervals; all other control variables taken into account; panelists = respondents of wave 1 with panel consent and registration on the survey portal.

Reference category: No benefit receipt and no employee responsibility

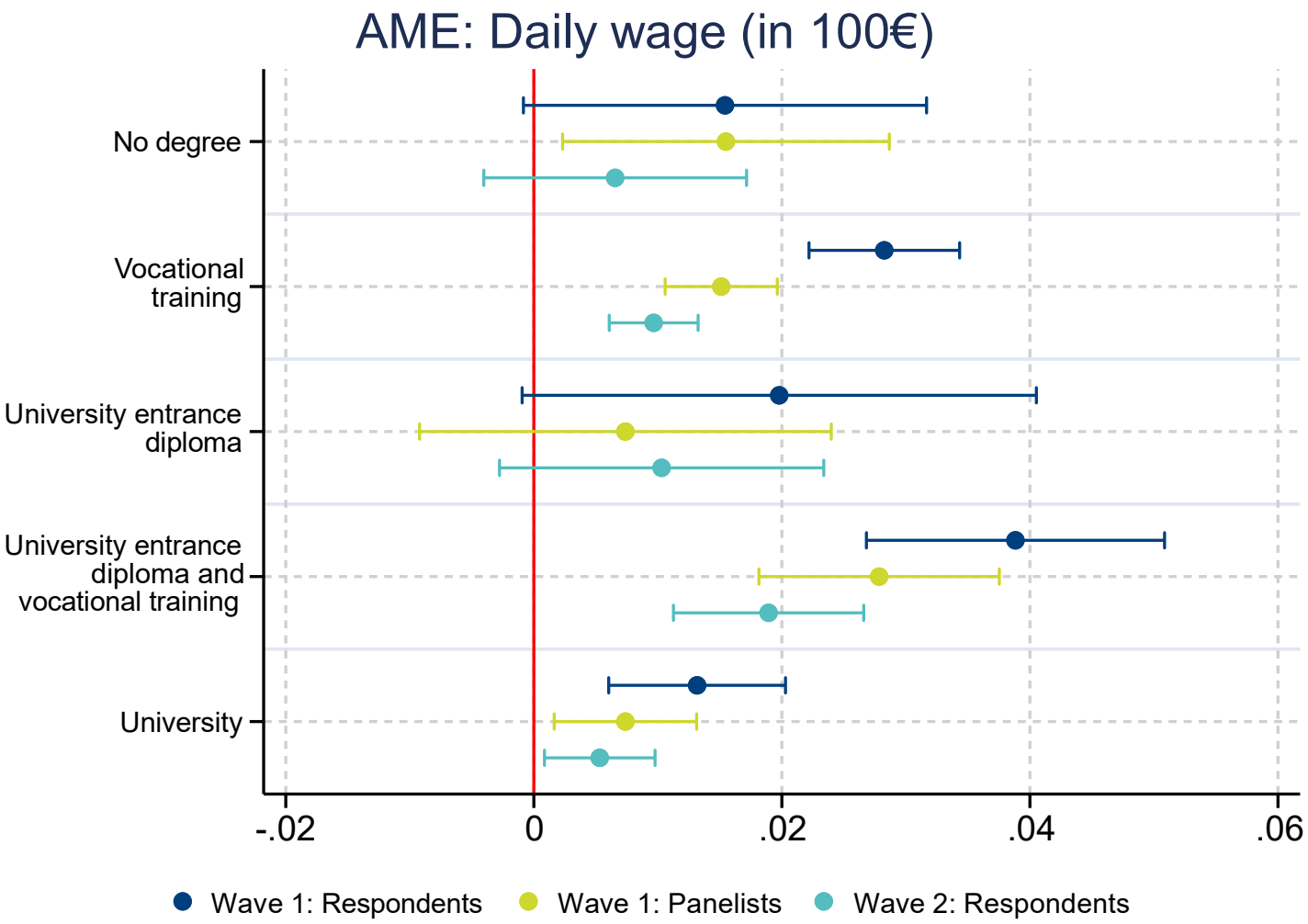
Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 115,610

MODEL 2: SELECTED INTERACTION EFFECTS: EDUCATION × EMPLOYMENT STATUS



Notes: 95% confidence intervals; all other control variables taken into account; panelists = respondents of wave 1 with panel consent and registration on the survey portal.
Reference category: Full-time employed
Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 115,610

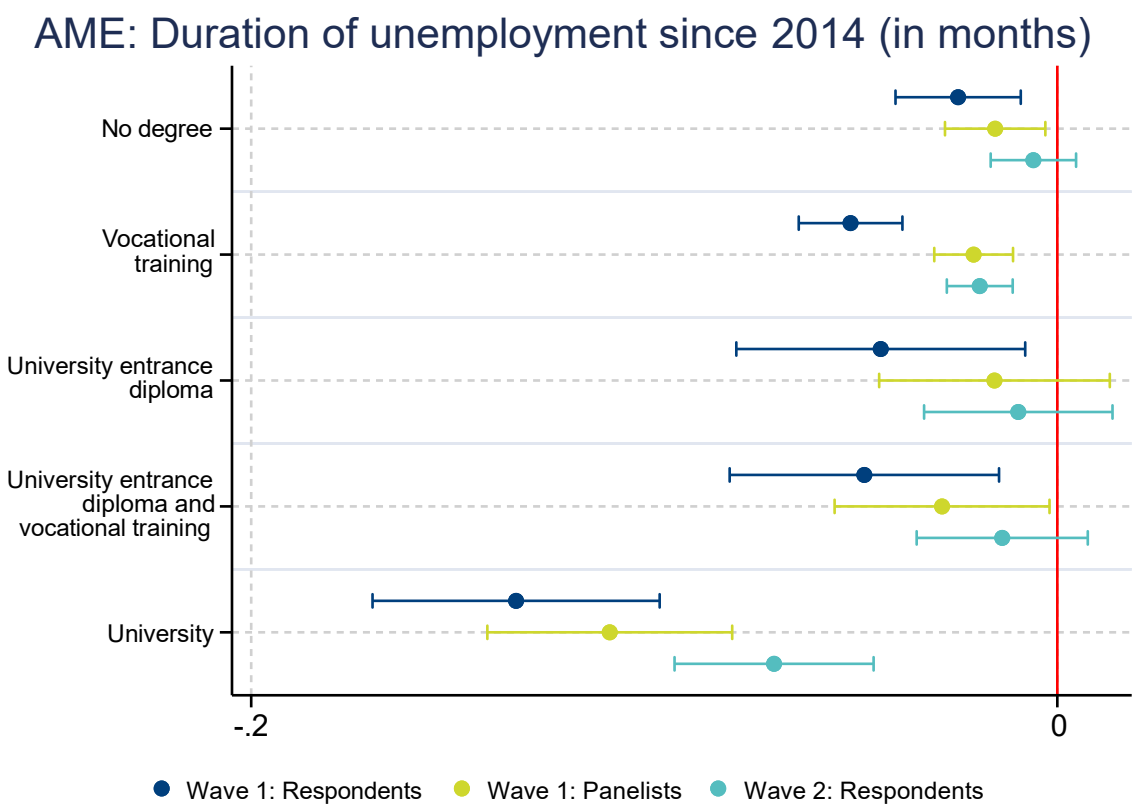
MODEL 2: SELECTED INTERACTION EFFECTS: EDUCATION × DAILY WAGE

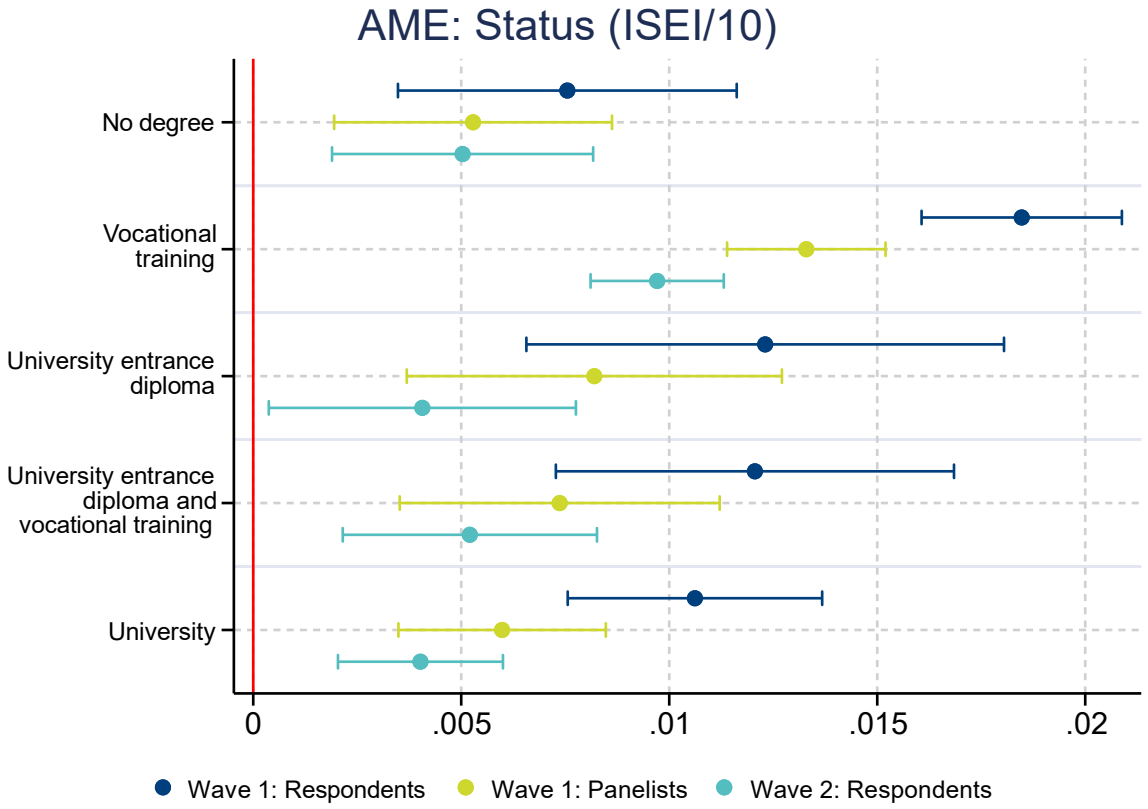


Notes: 95% confidence intervals; all other control variables taken into account; panelists = respondents of wave 1 with panel consent and registration on the survey portal.

Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 115,610

MODEL 2: SELECTED INTERACTION EFFECTS: EDUCATION × UNEMPLOYMENT DURATION, EDUCATION × STATUS





Notes: 95% confidence intervals; all other control variables taken into account; panelists = respondents of wave 1 with panel consent and registration on the survey portal.

Source: Integrated Employment Biographies 2021, IAB-OPAL wave 1-2, own calculations. N = 115,610