Supplementary Material for "Cash Transfers, Mental Health, and Agency: Evidence from an RCT in Germany"

Sandra Bohmann, Susann Fiedler, Maximilian Kasy,
Jürgen Schupp, Frederik Schwerter

April 7, 2025

Contents

List of Figures

List of Tables

S1 Experimental Design

We estimate the causal effect of a regular, guaranteed, and unconditional cash-transfer program on recipients' well-being in a preregistered randomized controlled trial. The RCT was implemented by our partner, the German NGO "Mein Grundeinkommen e.V.", which is financed through private donations. Prior to the RCT, Mein Grundeinkommen made regular cash transfers of EUR 1,000 per month for a single year—which are not evaluated in this paper—to 818 randomly assigned applicants, making Mein Grundeinkommen a credible partner to finance the basic income in our RCT.

S1.1 Sampling and treatment assignment

In the following we describe in detail the multi-step sampling and treatment assignment procedure used to construct our study sample. The steps in this procedure are (i) a public call and voluntary registration of potential participants, (ii) selection of a subsample based on demographic and economic eligibility criteria, (iii) stratified sampling of eligible registrants to construct a representative baseline sample, members of which were then invited to fill out a longer baseline survey, (iv) blocking of participants in the baseline sample who have a completed survey, based on a rich set of baseline covariates, and random assignment to treatment within blocks, and (v) selection of a representative subsample of blocks based on the budget constraints of the study.

S1.1.1 Sampling

Signup call and registrations In August 18, 2020, MG and the German Institute for Economic Research (DIW Berlin) publicly announced the launch of the RCT during Spring/Sommer 2021 and made a public call to register to participate in the RCT. The announcement included a description of the main features of the study: Selected participants of the study would be randomly assigned to a treatment group or a control group; treatment and control groups would participate in biannual online surveys; members of the treatment group would receive monthly payments of 1,200.00 EUR for three years; members of the control group would receive monetary incentives to complete the surveys; additional research activities may be offered. During signup, we collected the following screening information: Age, gender, education, monthly net income, number of people living in their household, number of kids, zip code, and their general attitude towards universal basic income. Between August 18 and December 10 in 2020, 2,048,370 potential participants registered in response to this public signup call.

Eligibility criteria We then invited a subsample of registered individuals (called "baseline sample") to complete the baseline survey. Selection into the baseline sample is based on the following eligibility criteria with respect to participants' demographic and socioeconomic characteristics. These eligibility criteria were largely determined by our implementation partner,

MG.

1. Participants have to be between 21 and 40 years old.

2. Households of size greater than one, and individuals with dependent children, are excluded from participation.

   Participants of our study whose household size changes, or who have a child, will, however, not lose their participation status.

3. Participants are required to be German residents and to have a monthly net income between 1,100.00 and 2,600.00 EUR.

4. Individuals who, at the time of the baseline survey, were receiving social benefits for long term unemployment are excluded from participation.[1]

   Participants of our study who transition to unemployment and receipt of social transfers will, however, not loose their participation status.

Baseline sample Among the potential participants who satisfied these criteria, our implementation partner next sampled 20,000 individuals who were invited to participate in a baseline survey. Sampling of these individuals was based on the following criteria. First, the sample was supposed to contain an equal number of proponents and opponents of a universal basic income. Second, potential participants in both of these groups were sampled using a weighted sampling procedure to generate a sample that is close to being representative for the (eligible) German population, and similar across both groups, in terms of age, gender, income, education, employment status, and state ("Bundesland").[2]

Baseline survey Before the invitations to the baseline survey were sent out, one person requested to be excluded from the RCT. The baseline survey resulted in 14,420 completed surveys. Of the remaining invitations,

- 51 invitations were sent to recipients with multiple registrations. These participants were in turn excluded since potential participants were allowed to register only once.
- 3,359 invitations were sent to recipients who subsequently never started the baseline survey.
- 328 invitations were sent to recipients who then started but did not complete the baseline survey.
- 1,841 recipients completed the survey, but did not sign the required data sharing consent forms.

[1] Given current benefit eligibility rules, such social benefits would have been cut by up to the full amount of the cash transfer by MG, if these individuals were to participate in our study. The net transfer to such individuals would thus have been significantly below the expenditure for MG.

[2] The details of this sampling procedure do not affect, in any case, the internal validity or correctness of inference for the study design described below.

Amongst the 14,420 individuals who completed the baseline survey and gave consent, 8,971 participants are considered in the randomized block assignment discussed next. The remaining 5,449 individuals are dropped because their eligibility status with respect to their characteristics listed above in criteria 1-4 changed and/or they had missing responses in baseline variables that were used in the randomized block assignment.[3]

S1.1.2 Blocking and treatment assignment

Blocking We use the answers to the screening questions and baseline survey to sort participants into homogenous blocks. We used the following 28 variables for this:

- Subjective well-being (PC1): first component of a principle component analysis that is based on an all subjective well-being outcomes that we discuss in detail in Section S1.3.
- Age 29-32: a dummy variable =1 if individuals' age is between 29 and 32 years, =0 if individuals' age is below 29 or above 32 years.
- Age 33-40: a dummy variable =1 if individuals' age is between 33 and 40 years, =0 if individuals' age is below 32 years.
- Female: a dummy variable =1 if individuals' gender is female, =0 if individuals' gender is not female.
- German citizen: a dummy variable = 1 if individual is a german citizen, =0 if not =1
- UBI proponent: a dummy variable = 1 if individuals' general attitude towards universal basic income is positive, =0 if it is negative.
- Tenure: a dummy variable =1 if the individual has (at least one) tenured job, =0 if the individual has no tenured job.
- Education: Hauptschule: a dummy variable = 1 if highest education level qualifies for vocational training, =0 if not =1.
- Education: Realschule: a dummy variable = 1 if highest education level qualifies for high school, =0 if not =1.
- Education: Fachabitur: a dummy variable = 1 if highest education level qualifies for vocational academy, =0 if not =1.

[3]Additionally, our implementation partner selected a group of 15 individuals who will be treated (that is, who will receive the basic income). These additional individuals indicated in the baseline survey that they were willing to participate in qualitative surveys (which are not conducted by the authors of this manuscript and are not part evaluated in this manuscript) and in interviews with journalists to publicly share their own experiences with the basic income *during* the RCT. Since any public appearance of these participants *may* bias their responses in our online surveys, we exclude these "media participants" from our study.

- Education: Abitur: a dummy variable = 1 if highest education level qualifies for university, =0 if not =1.
  (Note that the omitted education category is college or more.)
- Net monthly income: net monthly income available to the individual.
- Monthly saving: amount of money saved per month.
- Assets: individuals' total assets.
- Debt: individuals' level of debt.
- High financial security: a dummy variable = 1 if individual states that she could finance herself (with help of others but absent social security benefits) for one year without receiving any income, =0 if not =1.
- Working for money: a dummy variable = 1 if individual works and receives a financial compensation in return, =0 if not =1.
- In training or education: a dummy variable = 1 if individual is in vocational training or receives higher education (undergraduate, graduate, or doctoral level), =0 if not =1.
- In vocational training: a dummy variable = 1 if individual is in vocational training, =0 if not =1.
- Searching work: a dummy variable = 1 if looking for a job, =0 if not looking for a job.
- Sick days: number of workdays missed because of health.
- Weekly hours worked: number of hours worked per week
- Political preferences (PC1): first component of a principle component analysis that is based on an individual's response to how likely (in percent) it is that they vote for either party currently in the German parliament.
- Political preferences (PC2): second component of a principle component analysis that is based on an individual's response to how likely (in percent) it is that they vote for either party currently in the German parliament.
- Body mass index.
- Transfers to others: how much money did the individual give to family members or friends (or others) in 2020.
- Donation in 2020: how much money was donated in 2020.
- Binary gender: a dummy variable =1 if binary gender, =0 if not =1

Pairwise distances between observations are calculated using the Mahalanobis distance.[4] We construct blocks containing 32 observations each. The blocks are chosen to minimize the total sum of distances between pairs of observations within blocks. We do so using the R package *blockTools*[1]. We then discard all blocks with a maximum within-block distance greater than 14 (to avoid poorly matched observations), as well as one block with less than 32 observations.

Random assignment within blocks Within each block, treatment is assigned uniformly at random. We assign 2 out of the 32 observations in a block to the treatment group, 26 observations to the control group, and the remaining 4 observations to a "reserve," which is to be sampled in case of attrition of observations from the treatment or control group.

These numbers are chosen based on the following considerations: We want two treated units per block, in order to be able to calculate standard errors for the sample average treatment effect; cf.[2] and our discussion of inference below. We don't want more treated units per block, to keep blocks as homogenous as possible. The budget constraints of our implementation partner are furthermore such that we can survey 13 control units for every treated individual.

Lastly, because we have 107 treated individuals in total (an odd number), one additional individual from one block is chosen at random to participate in the treatment.

Weighted sampling of blocks This procedure results in 273 blocks, while our project budget allows for 53 blocks. These blocks are furthermore not fully representative for the baseline sample, because not all individuals who were invited to participate in the baseline survey passed eligibility and had non-missing responses in the questions we used for blocking (see above) and because of our discarding of poorly matched blocks.

In order to obtain a representative sample of blocks, we create block level sampling weights. These weights are chosen so as to match the distribution of gender, education groups, and income groups of eligible participants in the screening survey. We then draw a sample of 53 blocks from the 273 available blocks using these sampling weights, to obtain a representative subsample.

This results in 107 individuals assigned to treatment, 1377 assigned to the control group, and 212 individuals assigned to the "reserve," distributed evenly across 53 blocks.

At this point, the selected participants in the treatment group and control group were informed about their treatment status. 7 individuals in the control group wanted to be excluded from the study sample, 1 individual in the treatment group resigned his/her spot in the treatment group because of a job opportunity outside of Germany, and 1 individual in the treatment group could not be reached. For each of these missing individuals, we sampled one individual from the reserve group within the same block, to receive the corresponding treatment status.

[4]The Mahalanobis distance of two covariate vectors x_1 and x_2 that are realizations of a random vector X is given by $d(x_1, x_2) = \sqrt{(x_1 - x_2) \cdot Var(X)^{-1} \cdot (x_1 - x_2)}$.

Final study sample Our implementation partner decided to be willing to increase the size of the control group after the assignment. (Survey participation of control participants is costly and our implementation partner increased their funding for survey participation.) We henceforth added the remaining 203 from the reserve to the control group. Our final study sample hence includes 107 participants in the treatment group and 1,580 participants in the control group.

Table S1 summarizes the resulting study sample. The second and third columns show covariate averages for the 28 covariates used for blocking, for the treated and control group. Figure S1 shows the corresponding confidence intervals for mean differences across the treatment arms. The remaining columns show standard errors, confidence intervals, and p-values as discussed below. As can be seen from this table, we have achieved an extraordinary degree of balance between the treated and control group. Since inference here already corrects for blocking, it follows that in expectation one in 20 covariate differences should be significant at the 5% level.

S1.2 Treatment conditions and cash transfers

Treatment group Members of the treatment group received tax-free cash transfers of EUR 1,200, paid monthly, over the course of three years (*37*). There were no conditions attached to receiving the cash transfers, apart from completing six semi-annual online surveys. Note that the cash transfers were tax-free for the following reason: The overall sum of cash transfers, which amounts to EUR 43,200 (monthly payments of EUR 1,200 for three years), each participant of the treatment group received, consisted of multiple small gifts that our implementation partner redirects from different donors donating to them. Since each individual gift of these donors is too small to be relevant for taxes (ie is below EUR 20,000), the cash-transfers are overall tax-free.

Control group Members of the control group do not receive cash transfers, and were asked to complete the same six semi-annual surveys. For every completed survey, control participants received an incentive of EUR 10, plus an additional payment of EUR 30 if they completed all six surveys. We incentivized survey participation of our control group to limit attrition, which we discuss in detail below.

S1.3 Surveys

All surveys refer to online surveys. We designed the surveys. The surveys were distributed to participants by a professional German online survey company.

The baseline survey took place in February and March of 2021. The six semi-annual surveys took place in November 2021 (Wave 1), in Mai 2022 (Wave 2), in November 2022 (Wave 3), in May 2023 (Wave 4), in November 2023 (Wave 5), and in April 2024 (Wave 6).

We uploaded all surveys to OSF and are available upon request. In the following, we display and discuss in detail all questions that we used in the manuscript.

Well-being questions In these surveys, we repeatedly elicited standard measures of subjective and psychological well-being. We list the precise wording below. Note that general life satisfaction and depression are missing in wave 1, and perceived stress is missing in wave 2.

- Mental health, two scales:
  - WHO-5 depression scale[3]
    - Over the past 2 weeks...
      ...I have felt cheerful and in good spirits.
      ...I have felt calm and relaxed.
      ...I have felt active and vigorous.
      ...I woke up feeling fresh and rested.
      ...my daily life has been filled with things that interest me.
    - Responses: 0-4 Likert scale, at no point in time (=0), some of the time (=1), less than half the time (=2), more than half the time (=3), all of the time (=5).
    - Figure S2 shows a screenshot of the German version implemented in our survey.
  - Perceived stress scale[5], 6 items thereof:
    - In the last month, how often have you...
      ...felt difficulties were piling up so high that you could not overcome them?{−}
      ...felt that you were unable to control the important things in your life?{−}
      ...felt that things were going your way?
      ...felt confident about your ability to handle your personal problems?
      ...been able to control the way you spend your time?
      ...felt nervous and "stressed"? {−}
    - Responses are on a 0-4 Likert scale: Never (=0), almost never (=1), sometimes (=2), fairly often (=3), very often (=4).
    - To construct outcomes, we use inverted responses for first two items and the final one, which are marked by {−}.
    - Figure S3 shows a screenshot of the German version implemented in our surveys.
- Purpose in life:
  - Single item, as used in the German Socioeconomic Panel, SOEP,[6]:
    - Do you feel that what you do in life is meaningful and valuable?
    - Responses are on a 0-10 Likert scale: 0-10, endpoints not at all meaningful and valuable (=0), very much meaningful and valuable (=10).
    - Figure S4 shows a screenshot of the German version implemented in our surveys.

- Life satisfaction, several scales:
  - General life satisfaction, single item, as used in the German Socioeconomic Panel, SOEP,[6]:
    - How satisfied are you, overall, with your life
    - Responses are on a 0-10 Likert scale: 0-10, endpoints very much unsatisfied (=0), very much satisfied (=10).
    - Figure S5 shows a screenshot of the German version implemented in our surveys.
  - Domain satisfaction, single item for each domain, as used in the German Socioeconomic Panel, SOEP,[6]:
    - How satisfied are you, with your ...

      ... health

      ... sleep

      ... work

      ... income

      ... leisure

      ... social life
    - Responses are on a 0-10 Likert scale: 0-10, endpoints very much unsatisfied (=0), very much satisfied (=10).
    - Figure S6 shows a screenshot of the German version implemented in our surveys.

Other questions Beyond our outcome measures on well-being, we elicited additional information related to participants' demographic and socioeconomic background information (including, e.g., participants' gender, age, relationship status, household size, educational attainment) as well as information related to participants' labor outcomes, household finance outcomes, political behavior and economic (that is, risk, time, and social) preferences that will featured in future work.

For this project, we only made use of 4 other questions than our well-being outcomes. Based on these questions we conduct exploratory analyses on treatment effect heterogeneity and mediation that relate straightforwardly to the literature on well-being and money, as we discuss in the manuscript and below.

Our mediation analysis makes use of participants' perceived autonomy, which we surveyed in waves 1-5, but not in the baseline. Our heterogeneity analyses makes use of participants' income and wealth data that we elicited in all waves, including the baseline.

- Perceived autonomy
  - Single item, designed by ourselves.

    - How much do you agree to the following statement? I feel that I can freely determine my life.
    - Responses are on a 0-10 Likert scale: endpoints do not agree at all (=0), agree entirely (=10).
    - Figure S7 shows a screenshot of the German version implemented in our surveys.

- Household income
  - Single item
    - When you add up all your income: What is your current net income? Note: Please provide the monthly net amount, i.e., the money that remains after deductions for taxes and social security contributions. Please include regular payments such as housing benefits, child benefits, BAföG (federal educational assistance), alimony, or regular payments from family or friends. Please round the amount to a whole euro amount (without decimal places). Please estimate the monthly amount if you do not know the exact sum.
    - Figure S8 shows a screenshot of the German version implemented in our surveys.

- Financial security
  - Single item
    - If you suddenly found yourself in an unforeseen situation, could you support yourself for a year without working and without receiving social benefits? Note: For example, through savings or support from family and friends.
    - Figure S9 shows a screenshot of the German version implemented in our surveys.

- Assets
  - Single item
    - When you add up all your assets (cash and tangible assets including owner-occupied real estate), what do you estimate the total value to be? Note: Please do NOT subtract any debts, mortgages, credits, or loans. Remember, all your information will be stored anonymously and will not allow any conclusions to be drawn about your identity.
    - Figure S10 shows a screenshot of the German version implemented in our surveys.

- Debt
  - Single item
    - Do you currently have any debts or outstanding loans? Note: Please estimate the total amount of debt if you do not know the exact sum. Remember, all your information will be stored anonymously and will not allow any conclusions to be drawn about your identity.

    - Figure S11 shows a screenshot of the German version implemented in our surveys.

- Monthly savings
  - Single item
    - Since [Date 6 Months Ago], do you typically have a certain amount left each month that you save or set aside?
    - Figure S12 shows a screenshot of the German version implemented in our surveys.

- Donations
  - Single item
    - Since [Date 6 Months Ago], have you donated money for social, religious, cultural, or charitable purposes?
    - Figure S13 shows a screenshot of the German version implemented in our surveys.

- Financial support given to family and friends
  - Single item
    - Since [Date 6 Months Ago], have you have you financially supported relatives, your partner, friends, or acquaintances?
    - Figure S14 shows a screenshot of the German version implemented in our surveys.

- Consumption
  - Multiple items
    - You will now see some items on which one can, or must, spend money in everyday life. When you think about last month, approximately how many euros did you spend last month on the following items? Note: Please provide an amount in euros. If you did not spend any money on these items, simply enter a 0.
    - Figure S15 shows a screenshot of the German version implemented in our surveys.

- Time use
  - Multiple items

* What does your everyday life currently look like in a typical week (Mon–Sun)? Please indicate how many hours per week, on average, you spend on the following activities. Note: Sum up all activities within each category per week. Since activities may overlap, they do NOT need to add up to a total of 168 hours per week.
* Figure S16 shows a screenshot of the German version implemented in our surveys.

S2 Main Analysis

In this section, we discuss some methodological details of of our empirical analysis. We first describe our outcome variables. We then describe our approach to inference on treatment effects. Correct standard errors and p-values need to take into account the specifics of our experimental design, with matching and block-wise random assignment. We also discuss critical values for multiple hypothesis testing. Here we follow the corresponding literature.[2,7]

S2.1 Well-being outcomes

Based on all survey items on subjective well-being that we elicited from participants in the baseline survey and until wave 6 (our final survey wave), we construct our well-being outcomes. In the following, we define all outcomes, including the different levels of survey-question aggregation (across scales and waves), and the measurement of outcomes as changes relative to baseline values (for our main analysis) or realized levels (for our robustness checks).

1. **WHO-5 depression**:

   We average the five WHO-5 questions for each individual and for each wave to compute WHO-5 values for each individual and wave.[5] We calculate the standard deviation of the resulting WHO-5 values at baseline for all individuals, use it to normalize the WHO-5 values of each individual and wave, and obtain normalized WHO-5 values for each individual and wave.

   To obtain our first WHO-5 depression outcome, we calculate the difference in normalized WHO-5 values for each individual and wave > 0 and their normalized WHO-5 value at baseline, and average the resulting differences across waves 2-6. We refer to this outcome as WHO-5 depression in terms of changes, averaged across waves.

   To obtain our second WHO-5 depression outcome, we average the normalized WHO-5 values across waves 2-6. We refer to this outcome as WHO-5 depression in terms of levels, averaged across waves.

2. **PSS stress**:

[5] The WHO-5 questions were missing in wave 1, so WHO-5 values for wave 1 are missing.

We average the six PSS questions for each individual and for each wave to compute PSS values for each individual and wave.[6] We calculate the standard deviation of the resulting PSS values at baseline for all individuals, use it to normalize the PSS values of each individual and wave, and obtain normalized PSS values for each individual and wave.

To obtain our first PSS stress outcome, we calculate the difference in normalized PSS values for each individual and wave > 0 and their normalized PSS value at baseline, and average the resulting differences across waves 1, 3-6. We refer to this outcome as PSS stress in terms of changes, averaged across waves.

To obtain our second PSS stress outcome, we average the normalized PSS values across waves 1, 3-6. We refer to this outcome as PSS stress in terms of levels, averaged across waves.

3. **Domain satisfaction index**:

We average the six domain satisfaction questions for each individual and for each wave to compute the domain satisfaction index values for each individual and wave. We calculate the standard deviation of the resulting domain satisfaction index values at baseline for all individuals, use it to normalize the domain satisfaction index values of each individual and wave, and obtain normalized domain satisfaction index values for each individual and wave.

To obtain our first domain satisfaction index outcome, we calculate the difference in normalized domain satisfaction index values for each individual and wave > 0 and their normalized domain satisfaction index value at baseline, and average the resulting differences across waves 1-6. We refer to this outcome as the domain satisfaction index in terms of changes, averaged across waves.

To obtain our second DSI outcome, we average the normalized domain satisfaction index values across waves 1-6. We refer to this outcome as the domain satisfaction index in terms of levels, averaged across waves.

4. **General life satisfaction**:

We obtain the general life satisfaction values for each individual and for each wave from their respective response to the general life satisfaction question.[7] We calculate the standard deviation of the general life satisfaction values at baseline for all individuals, use it to normalize the general life satisfaction values of each individual and wave, and obtain normalized general life satisfaction values for each individual and wave.

[6] The PSS questions were missing in wave 2, so PSS values for wave 2 are missing.
[7] The general life satisfaction question was missing in wave 1, so its values for wave 1 are missing.

To obtain our first general life satisfaction outcome, we calculate the difference in normalized general life satisfaction values for each individual and wave > 0 and their normalized general life satisfaction value at baseline, and average the resulting differences across waves 2-6. We refer to this outcome as general life satisfaction in terms of changes, averaged across waves.

To obtain our second general life satisfaction outcome, we average the normalized general life satisfaction values across waves 2-6. We refer to this outcome as general life satisfaction in terms of levels, averaged across waves.

5. **Income satisfaction**:

   We obtain the income satisfaction values for each individual and for each wave from their respective response to the income satisfaction question. We calculate the standard deviation of the income satisfaction values at baseline for all individuals, use it to normalize the income satisfaction values of each individual and wave, and obtain normalized income satisfaction values for each individual and wave.

   To obtain our first income satisfaction outcome, we calculate the difference in normalized income satisfaction values for each individual and wave > 0 and their normalized income satisfaction value at baseline, and average the resulting differences across waves 1-6. We refer to this outcome as income satisfaction in terms of changes, averaged across waves.

   To obtain our second income satisfaction outcome, we average the normalized income satisfaction values across waves 1-6. We refer to this outcome as income satisfaction in terms of levels, averaged across waves.

6. **Leisure satisfaction**:

   We employ the same procedure discussed for the income satisfaction outcomes, but use the leisure satisfaction question to obtain the outcomes: leisure satisfaction in terms of changes, averaged across waves; and leisure satisfaction in terms of levels, averaged across waves.

7. **Social satisfaction**:

   We employ the same procedure discussed for the income satisfaction outcomes, but use the social satisfaction question to obtain the outcomes: social satisfaction in terms of changes, averaged across waves; and social satisfaction in terms of levels, averaged across waves.

8. **Health satisfaction**:

   We employ the same procedure discussed for the income satisfaction outcomes, but use the health satisfaction question to obtain the outcomes: health satisfaction in

terms of changes, averaged across waves; and health satisfaction in terms of levels, averaged across waves.

9. **Sleep satisfaction**:

   We employ the same procedure discussed for the income satisfaction outcomes, but use the sleep satisfaction question to obtain the outcomes: sleep satisfaction in terms of changes, averaged across waves; and sleep satisfaction in terms of levels, averaged across waves.

10. **Work satisfaction**:

    We employ the same procedure discussed for the income satisfaction outcomes, but use the work satisfaction question to obtain the outcomes: work satisfaction in terms of changes, averaged across waves; and work satisfaction in terms of levels, averaged across waves.

11. **Purpose in life**:

    We employ the same procedure discussed for the income satisfaction outcomes, but use the purpose in life question to obtain the outcomes: purpose in life in terms of changes, averaged across waves; and purpose in life in terms of levels, averaged across waves.

12. **Mental meath**:

    We re-calculate for each individual their normalized WHO-5 values and PSS values, both either in changes or in levels, by using only waves 3-6, and average them to obtain the aggregated mental health outcomes: mental health in terms of changes, averaged across waves; mental heath in terms of levels, averaged across waves.

13. **(Overall) Life satisfaction**:

    We re-calculate for each individual their normalized domain satisfaction index values, either in changes or in levels, by using only waves 2-6, and average general life satisfaction and the resulting domain satisfaction index to obtain the aggregated life satisfaction outcomes: life satisfaction in terms of changes, averaged across waves; life satisfaction in terms of levels, averaged across waves.

S2.2 Inference

Denote individual treatment assignment by D and outcomes by Y. Our primary object of interest is the sample average treatment effect

$$\delta = \sum_i (Y_i^1 - Y_i^0), \tag{1}$$

for various individual-level outcomes Y_i for individuals i and corresponding potential outcomes Y_i^0, Y_i^1. Our primary estimator is based on block-level differences in mean outcomes, averaged across blocks b:

$$\bar{Y}_b^1 = \frac{1}{n_b^1} \sum_{i:\ b_i=b} D_i Y_i \qquad \bar{Y}_b^0 = \frac{1}{n_b^0} \sum_{i:\ b_i=b} (1-D_i) Y_i$$
$$\hat{\delta}_b = \bar{Y}_b^1 - \bar{Y}_b^0 \qquad \hat{\delta} = \frac{1}{N} \sum \hat{\delta}_b, \tag{2}$$

where n_b^1 and n_b^0 are the number of treated and untreated individuals in block b, and N is the number of blocks. Inference is based on two alternative methods, both of which yield valid inference for the sample average treatment effect: Standard errors and confidence intervals based on a normal approximation, and randomization inference.

Standard errors To calculate a standard error for $\hat{\delta}$ as an estimator of δ, we calculate block-level standard-errors (allowing for arbitrary heteroskedasticity), and aggregate:

$$\hat{\sigma}_b^{2,1} = \frac{1}{n_b^1 - 1} \sum_{i:\ b_i=b} D_i \cdot (Y_i - \bar{Y}_b^d)^2 \qquad \hat{\sigma}_b^{2,0} = \frac{1}{n_b^0 - 1} \sum_{i:\ b_i=b} (1-D_i) \cdot (Y_i - \bar{Y}_b^d)^2$$
$$\hat{\sigma}_b^2 = \frac{1}{n_b^1} \hat{\sigma}_b^{2,1} + \frac{1}{n_b^0} \hat{\sigma}_b^{2,0} \qquad \hat{\sigma}^2 = \frac{1}{N} \sum_b \hat{\sigma}_b^2. \tag{3}$$

95% confidence intervals for δ are then calculated as

$$CI = [\hat{\delta} - 1.96 \cdot \hat{\sigma}^2, \hat{\delta} + 1.96 \cdot \hat{\sigma}^2]. \tag{4}$$

Neyman p-values (denoted p-val (N) in our tables) are similarly based on these standard errors and the normal approximation for the distribution of $\hat{\delta}$.

Randomization inference Our second approach toward inference is based on permutations of treatments, that is, based on randomization inference. This approach allows us to test the null hypothesis that the intervention had no effect of any kind, that is, $Y_i^1 = Y_i^0$ for all individuals i and potential outcomes Y_i^1, Y_i^0.

We re-assign treatment at random *within* each of the blocks b. For this counterfactual treatment assignment, we re-calculate any given test-statistic. Repeating this process many times, we calculate the share of re-assignments for which the test-statistic is bigger than the realized value of the test-statistic. This share is the Fisher p-value (denoted p-val (F) in our tables) for the null hypothesis of no effects.

Compound hypotheses In order to deal with the issue of multiple testing in a principled manner, we preregistered to use the Benjamini–Hochberg procedure, which allows us to control the false discovery rate, that is, the share of rejected hypotheses which in fact hold true. This

procedure works as follows. We sort the p-values for each of the m hypotheses tested by size, resulting in ordered values $P_{(j)}$. For a critical value α, we find the largest value k such that

$$P_{(k)} \leq \frac{k}{m}\alpha. \tag{5}$$

We reject the null hypothesis for all $i = 1, \ldots, k$.

We preregistered to apply this procedure separately for different groups of outcomes and stated that one group of outcomes refers to subjective well-being.

S3 Main Findings

S3.1 Treatment effects averaged across waves

To obtain summary estimates for each dimension of well-being during the study period that we cover, we focus on the average treatment effects on the aggregated outcomes (in terms of changes and levels) for mental health, purpose of life, and life satisfaction. We show the results for outcomes in terms of changes in the first three rows of Table 1 of the main manuscript. We show the results for outcomes in terms of levels in the first three rows in Table S2, here in the supplementary materials.

We find statistically significant improvements in mental health (0.347 standard deviations for the respective outcome in terms of changes, and 0.286 SD for the respective outcome in terms of levels; respective Newman and Fisher's exact p-values are below 0.05, see Table 1 and Table S2). Improvements in mental health are separately present for the WHO-5 depression scale (0.320 SD for in changes and 0.251 SD for in levels; respective Newman and Fisher's exact p-values are below 0.05, see Table 1 and Table S2) and the PSS stress scale (0.285 SD for in changes and 0.234 SD for in levels; respective Newman and Fisher's exact p-values are below 0.05, see Table 1 and Table S2). Effect sizes are slightly larger in the outcomes that refer to changes, reflecting negative treatment-control imbalances at baseline, which remained despite stratification, and are not adjusted for in the outcomes that refer to levels. Figure S17 depicts these treatment-control imbalances (with wave 0 denoting the baseline) in the panel on the left.

We find statistically significant improvements of purpose in life only when adjusting for treatment-control imbalances at baseline (estimated treatment effects are 0.250 SD with Newman and Fisher's exact p-values below 0.05 for in changes, and 0.122 SD with Newman and Fisher's exact p-values equal to 0.07 and 0.058, respectively, for in levels). This finding seems to reflect the negative treatment-control imbalance in purpose in life at baseline, which remained despite stratification. Figure S17 depicts this treatment-control imbalance at baseline in the panel in the middle.

We find statistically significant improvements in life satisfaction (0.417 SD for in changes, and 0.326 SD for in levels; respective Newman and Fisher's exact p-values are below 0.05, see Table 1 and Table S2). Improvements are separately present for general life satisfaction (0.351

SD for in changes and 0.277 SD for in levels; respective Newman and Fisher's exact p-values are below 0.05, see Table 1 and Table S2) and the domain satisfaction index (0.420 SD for in changes and 0.326 SD for in levels; respective Newman and Fisher's exact p-values are below 0.05, see Table 1 and Table S2). Effect sizes are again slightly larger in the outcomes that refer to changes, reflecting negative treatment-control imbalances at baseline, which remained despite stratification, and are not adjusted for in the outcomes that refer to final values. Figure S17 depicts these treatment-control imbalances at baseline (wave 0) in the panel on the right.

We find statistically significant improvements in satisfaction with income (0.551 SD for in changes, and 0.522 SD for in levels; respective Newman and Fisher's exact p-values are below 0.05, see Table 1 and Table S2), health (0.291 SD for in changes, and 0.171 SD for in levels; respective Newman and Fisher's exact p-values are below 0.05, see Table 1 and Table S2), and leisure (0.245 SD for in changes, and 0.203 SD for in levels; respective Newman and Fisher's exact p-values are below 0.05, see Table 1 and Table S2). We find statistically significant improvements in satisfaction with sleep (0.290 SD for in changes, and 0.138 SD for in levels; respective Newman and Fisher's exact p-values are below 0.05, see Table 1 and Table S2). Effect sizes are again slightly larger in the outcomes that refer to changes, reflecting negative treatment-control imbalances at baseline, which remained despite stratification, and are not adjusted for in the outcomes that refer to final values. Figure S18 depicts these treatment-control imbalances at baseline.

We find statistically significant improvements for social satisfaction only when we do not adjust for treatment-control imbalances at baseline (estimated treatment effects are 0.125 SD with Newman and Fischer's exact p-values equal to 0.082 in both cases for the outcome in terms of changes, and 0.190 SD with Newman and Fischer's exact p-values below 0.05 for the outcome in terms of levels). This reflects positive treatment-control imbalances at baseline, which remained despite stratification. Figure S18 depicts this treatment-control imbalance.

We do not find statistically significant improvements for work satisfaction (estimated treatment effects are 0.143 SD for the outcome in terms of changes, and 0.066 SD for the outcome in terms of levels; Newman and Fischer's exact p-values are larger than 0.05 for both cases). However, the treatment effects on work satisfaction are significant when we only consider the final three waves. The effect size is again slightly larger for changes relative to the baseline, reflecting negative treatment-control imbalances at baseline, which remained despite stratification. Figure S18 depicts this treatment-control imbalances.

S3.1.1 Longevity

We study the longevity of treatment effects six months after the final cash transfer and report the findings in Table S3. While treatment effects reduce slightly in size six months after the final cash transfers, they retain, on average, 84% of their effect size during the cash transfer program. More precisely, cash transfers continue to significantly improve mental health by 0.251 SD, purpose in life by 0.307 SD, and life satisfaction by 0.284 SD

S3.2 Treatment effect dynamics

To study the dynamics of the well-being improvements, we display treatment effects separately for all waves and outcomes in Figures 2 and 3 in the main manuscript for outcomes that refer to changes and in Figures S17 and S18 for outcomes that refer to final values. Treatment effects mostly remain constant over time, between waves 1-6. We observe increases in time for the purpose in life and work satisfaction. The treatment effect on income satisfaction is decreasing in time.

S4 Robustness

In this section we show that our findings are robust to multiple hypothesis testing adjustments, and that they do not appear to be affected by selective attrition.

S4.1 Multiple hypothesis testing

Following our preregistration, we apply the Benjamini–Hochberg (BH) procedure[8] to control for multiple hypothesis testing. As described in S2.2 of the Supplement, the BH procedure uses an adjustment of the cutoff for p-values that is used to determine statistical significance of an estimated treatment effect. To determine the BH cutoffs, we rank p-values across a set of tested hypotheses, and assign to each hypothesis a threshold of $\frac{k}{m}\alpha$, where k denotes the rank in p-values of each hypothesis, m denotes the total number of hypotheses tested, and $\alpha = 0.05$ denotes the desired upper bound on the false discovery rate. We reject all hypotheses up to the first value of k for which the p-value exceeds the cutoff $\frac{k}{m}\alpha$.

Our main results, that we report in Table 1 of the main manuscript, show substantial well-being improvements generated by regular, unconditional, and guaranteed cash transfers to recipients in Germany. These results are based on the $m = 13$ estimated effects of treatment on the change of outcomes. As shown in Table 1, 11 of these 13 estimates are significantly different from zero, i.e., the respective p-values are below the conventional critical value of $\alpha = 0.05$.

To determine whether these results are robust to multiple hypothesis testing according the BH procedure, we rank all estimated treatment effects according to their p-values, and compute for each one their respective BH cutoff, see Table S4 and Table S5. For Newman p-values, the p-value corresponding to the estimated treatment effect on leisure satisfaction, for example, is 0.008 and has rank 11, that is, it is the 11th smallest among all $m = 13$ Newman p-values shown in Table 1 of the main manuscript. The BH cutoff for the p-value for leisure satisfaction then is $\frac{k}{m}\alpha = 0.042$. The p-value of the estimated treatment effect on leisure satisfaction is therefore smaller than its BH cutoff. We find that all results stated in Table 1 are robust to BH procedure: the p-values of all statistically significant treatment effects are below their respective BH critical values, see Table S4 for results based on Newman p-values and Table S5 for results based on Fisher's exact p-values.

Our results remain significant after the adjustments of the BH procedure if we apply this procedure to the 26 estimated treatment effects that refer to our outcomes both in terms of changes and levels (rather than only the 13 outcomes measured in changes). Our initial findings are that 20 of the 26 estimated treatment effects are statistically significant (11 of the 13 treatment effects on outcomes that are referring to changes and 9 of the 13 treatment effects on outcomes that are referring to levels). For each estimated treatment effect, we calculate their specific BH critical value and compare them to their p-values. Tables S6 (based on Newman p-values) and S7 (based on Fisher's exact p-values) show that the p-values of all treatment effects that are below the critical value of 0.05 are also below their BH critical value. All our findings regarding the statistical significance of well-being improvements are hence robust to adjustments for multiple hypothesis testing.

S4.2 Attrition

We encountered no attrition in our treatment group and some attrition in our control group, where survey response rates declined slightly during waves 1 to 6. Nonetheless, 71% of our control group completed all 6 surveys, 79% completed at least 5 out of 6 waves, 88% at least 3 out of 6 waves, and 97% at least 1 out of 6 waves. Based on the following two analyses, we however conclude that attrition does not appear to be selective in a way that would impact our findings.

If attrition were selective, this might cause bias in our treatment effect estimates. We might find significant effects, even if in truth (without selection) there is no effect. As a placebo test for this possibility, we estimate treatment effects on our baseline covariates (where true treatment effects are by construction equal to 0), while artificially restricting our sample to only those individuals without missing data for any subsequent wave. The p-values corresponding to these estimates are shown in Figure S19. We find no significant effects for the restricted sample. This increases our confidence that our findings are not driven by selective attrition.

As a second test for selective attrition, we restrict attention to the control group ($D = 0$). Within this group, we compare the average of our main outcome variables (mental health, purpose in life, life satisfaction) at baseline (wave 0), across observations with different numbers of missing waves. If attrition were selective, we would expect that these means vary across the number of waves missing, see Table S8. We find no differences across number of waves missing.

Based on these two analyses, we conclude that our results are not likely to be affected by selective attrition.

S4.3 Experimenter Demand

The internal validity of experiments with human participants may, in principle, suffer from experimenter demand and Hawthorne effects.[9,10] If there are demand effects, treated participants might report greater well-being to lend support to basic income policies more generally. If our findings are due to Hawthorne effects, participants change their behavior only because they are

experimentally observed. The following observations suggests that experimenter demand and Hawthorne effects do not account for our findings.

First, we find no treatment effect on participants' political support for basic income policies. In the end of the final wave, we asked participants the following question: "Do you support the idea of a universal basic income for all citizens?"[8] Participants could response on a 1-4 scale, with 1 = yes, absolutely, 2 = rather yes, 3 = rather no, 4 = no, absolutely not. The average response of treated participants is 1.928 and of control participants is 1.964. The treatment difference is -0.035, which is small and not statistically significantly different from zero, with a one-sided Newman p-value of 0.33 and one-sided Fisher's exact p-value of 0.3.[9] This finding contradicts explanations based on experimenter demand, where recipients try to give answers that lend support to basic income policies.

Second, the dynamics over time of treatment effects on subjective well-being are not consistent with demand and Hawthorne effects, either. Demand and Hawthorne effects cannot explain the decreasing treatment effect on income satisfaction over time, nor the increasing treatment effects on purpose in life and work satisfaction. Instead, these patterns suggest that income satisfaction is subject to adaptation[11,12,13], while effects on purpose in life and work satisfaction are delayed because they require live changes that take time.

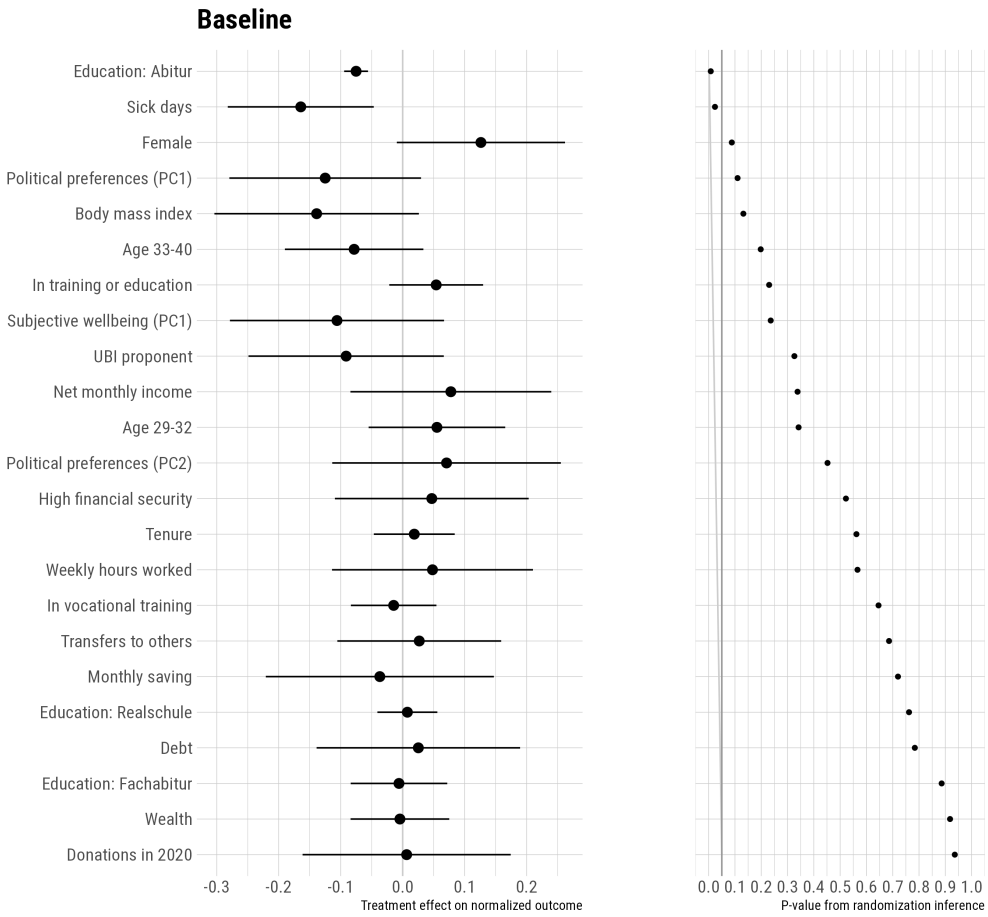
Third, our experimental design limits the potential roles of demand and Hawthorne effects. To limit demand effects, we explicitly asked participants to respond accurately to factual questions, stated that there are no right or wrong answers to subjective questions, and used a third-party survey company to implement our surveys. Regarding Hawthorne effects, treated and control participants are observed to the same extent. This makes explanations of our findings based on Hawthorne effects, where outcomes are moved by observation rather than treatment, implausible.

[8] In German, we asked: "Befürworten Sie die Idee eines bedingungslosen Grundeinkommens für alle Bürger/innen?"

[9] Since demand effects predict greater support for treated participants, we conducted one-sided tests, in favor of finding significant demand effects.

S5 Additional Figures and Tables

Figure S1: Confidence intervals for baseline differences between treatment and control



Notes: This figure shows mean differences of baseline covariates for the treated and control group in our study sample. Confidence intervals and p-values are adjusted for blocked random assignment.

Die folgenden Aussagen betreffen Ihr Wohlbefinden in den letzten zwei Wochen.
Bitte geben Sie jeweils an, welche Aussage am besten beschreibt, wie Sie sich in den letzten zwei Wochen gefühlt haben.

In den letzten zwei Wochen ...

	zu keinem Zeitpunkt	ab und zu	etwas weniger als die Hälfte der Zeit	etwas mehr als die Hälfte der Zeit	die ganze Zeit	keine Angabe
... war ich froh und guter Laune.						
... habe ich mich ruhig und entspannt gefühlt.						
... habe ich mich energisch und aktiv gefühlt.						
... habe ich mich beim Aufwachen frisch und ausgeruht gefühlt.						
... war mein Alltag voller Dinger, die mich interessieren.						

Figure S2: Implemented German version of the WHO-5 depression scale. English translation in Section S1.3

Table S1: Balance of baseline covariates in the study sample

Outcome	Treated	Control	ATE	SE	t-stat	p-val (N)	p-val (F)	n treated	n control
Education: Abitur	0.038	0.055	-0.017	0.002	-7.653	0.000	0.008	107	1580
Sick days	7.676	10.941	-3.265	1.194	-2.736	0.006	0.024	107	1580
Female	0.472	0.409	0.062	0.034	1.824	0.068	0.088	107	1580
Political preferences (PC1)	-0.019	0.158	-0.177	0.112	-1.584	0.113	0.110	107	1580
Body mass index	24.707	25.457	-0.750	0.455	-1.649	0.099	0.132	107	1580
Age 33-40	0.340	0.378	-0.038	0.028	-1.373	0.170	0.198	107	1580
In training or education	0.170	0.150	0.019	0.014	1.401	0.161	0.230	107	1580
Subjective wellbeing (PC1)	-0.336	-0.138	-0.197	0.164	-1.202	0.229	0.236	107	1580
UBI proponent	0.500	0.545	-0.045	0.040	-1.135	0.257	0.326	107	1580
Net monthly income	1963.654	1933.191	30.463	32.388	0.941	0.347	0.338	107	1580
Age 29-32	0.355	0.329	0.026	0.026	0.983	0.326	0.342	107	1580
Political preferences (PC2)	0.130	0.043	0.086	0.115	0.753	0.452	0.452	107	1580
High financial security	0.327	0.305	0.022	0.037	0.589	0.556	0.522	107	1580
Tenure	0.774	0.766	0.008	0.014	0.563	0.574	0.562	107	1580
Weekly hours worked	38.042	37.382	0.659	1.133	0.582	0.561	0.566	107	1580
In vocational training	0.425	0.432	-0.007	0.017	-0.413	0.680	0.646	107	1580
Transfers to others	368.396	344.550	23.847	60.055	0.397	0.691	0.686	107	1580
Monthly saving	284.553	294.116	-9.563	24.356	-0.393	0.695	0.720	107	1580
Education: Realschule	0.217	0.214	0.003	0.010	0.308	0.758	0.762	107	1580
Debt	10072.006	9401.943	670.064	2216.113	0.302	0.762	0.784	107	1580
Education: Fachabitur	0.236	0.238	-0.003	0.017	-0.150	0.881	0.886	107	1580
Wealth	24795.597	25010.975	-215.378	1973.942	-0.109	0.913	0.918	107	1580
Donations in 2020	99.148	97.592	1.556	20.739	0.075	0.940	0.936	107	1580

Notes: This table shows averages of baseline covariates for the treated and control group in our study sample, as well as their difference. As this table shows, we were able to achieve a very high degree of balance for almost all variables.

Figure S3: Implemented German version of the perceived stress scale. English translation in Section S1.3

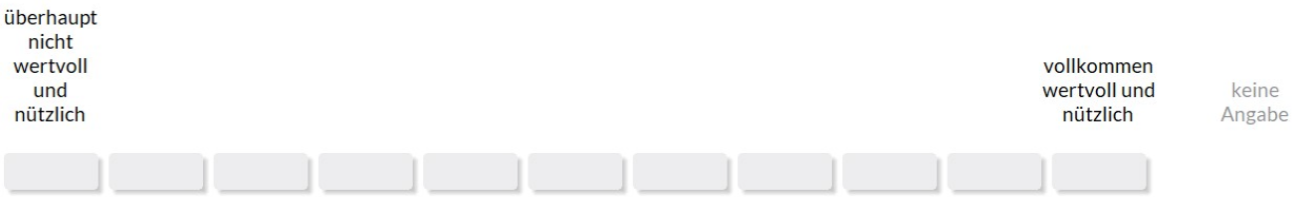


Figure S4: Implemented German version of the single item purpose in life measure. English translation in Section S1.3

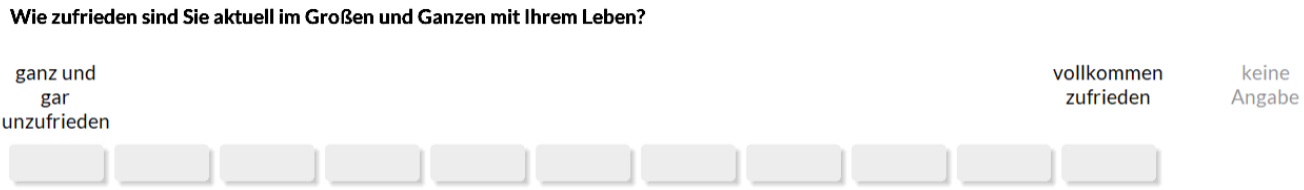


Figure S5: Implemented German version of the single item general life satisfaction question. English translation in Section S1.3

Figure S6: Implemented German version of the domain satisfaction questions. English translation in Section S1.3

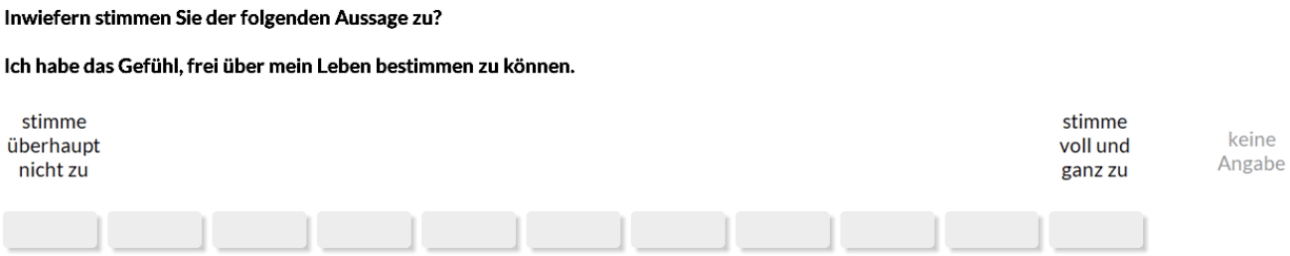


Figure S7: Implemented German version of perceived autonomy. English translation in Section S1.3

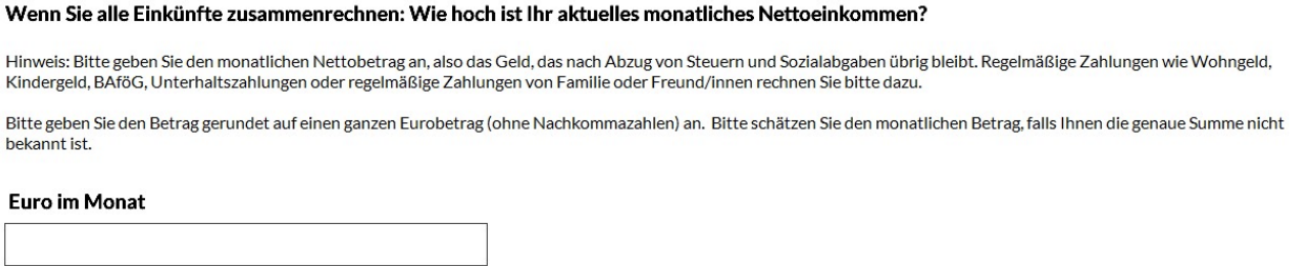


Figure S8: Implemented German version of net household income. English translation in Section S1.3

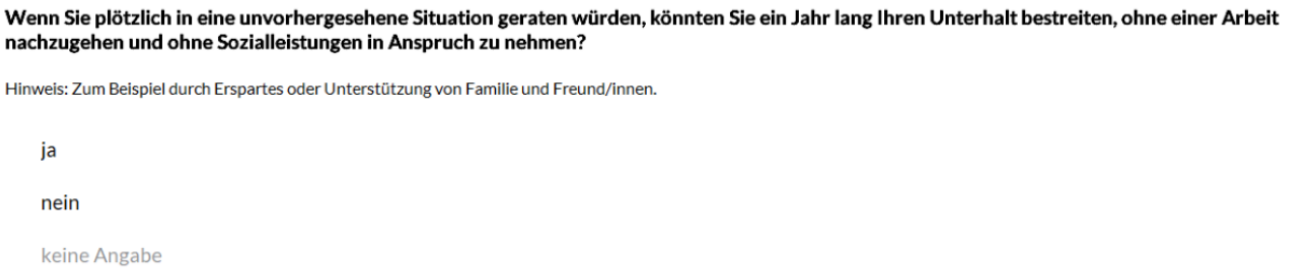


Figure S9: Implemented German version of financial security question. English translation in Section S1.3

Wenn Sie Ihr gesamtes Vermögen zusammenrechnen (Geld- und Sachvermögen einschließlich selbstgenutztem Wohneigentum), wie hoch schätzen Sie den Gesamtwert?

Hinweis: Bitte ziehen Sie eventuelle Schulden, aufgenommene Hypotheken, Kredite oder Darlehen **NICHT** ab.

Zur Erinnerung: Alle Ihre Angaben werden anonymisiert abgespeichert und erlauben keine Rückschlüsse auf Ihre Person.

keinerlei Vermögen

1 bis unter 5.000 Euro

5.000 bis unter 10.000 Euro

10.000 bis unter 20.000 Euro

20.000 bis unter 50.000 Euro

50.0000 bis unter 100.000 Euro

100.000 bis unter 500.000 Euro

500.000 Euro und mehr

keine Angabe

Figure S10: Implemented German version of the total assets question. English translation in Section S1.3

Haben Sie derzeit Schulden oder laufende Kredite?

Hinweis: Bitte schätzen Sie den gesamten Schuldenbetrag, falls Ihnen die genaue Summe nicht bekannt ist.

Zur Erinnerung: Alle Ihre Angaben werden pseudonymisiert abgespeichert und erlauben keine Rückschlüsse auf Ihre Person.

ja, und zwar insgesamt etwa: Euro

nein

keine Angabe

Figure S11: Implemented German version of the debt question. English translation in Section S1.3

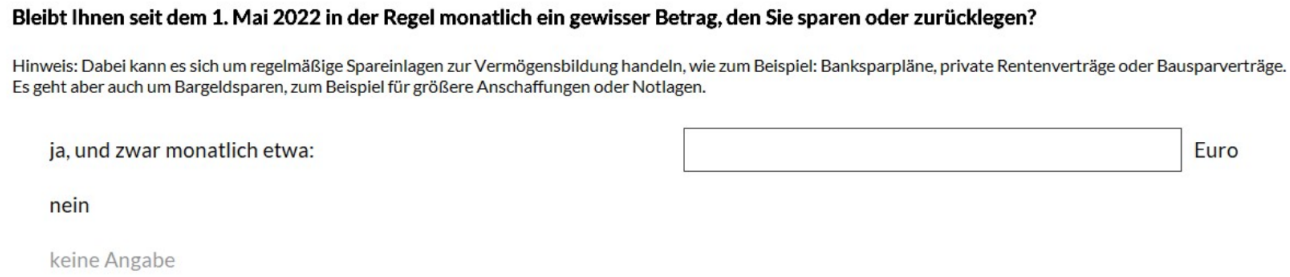


Figure S12: Implemented German version of the savings question. English translation in Section S1.3

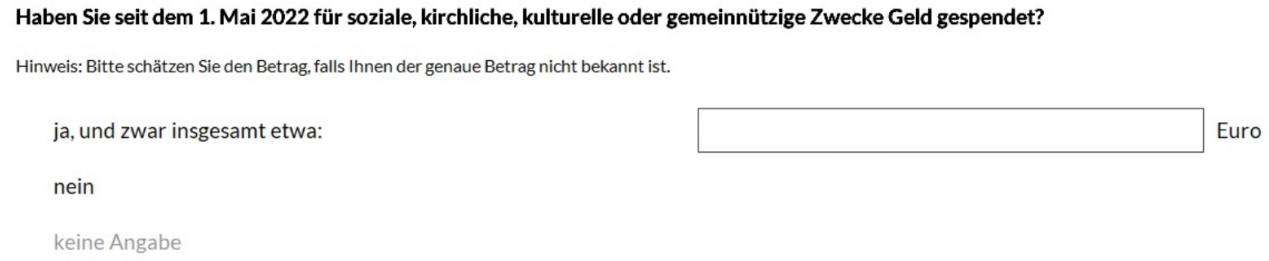


Figure S13: Implemented German version of the donation question. English translation in Section S1.3

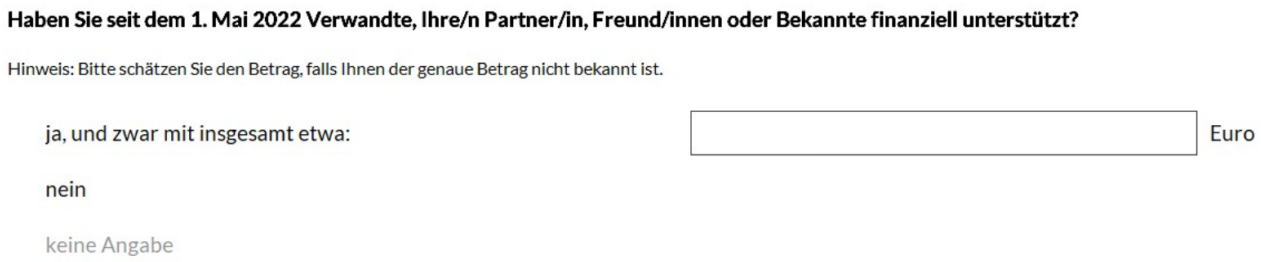


Figure S14: Implemented German version of the financial support of others question. English translation in Section S1.3

Sie sehen nun einige Dinge für die man im Alltag Geld ausgeben kann, oder muss.
Wenn Sie einmal an den letzten Monat denken: Wie viel Euro haben Sie im letzten Monat in etwa für die folgenden Dinge ausgegeben?

Hinweis: Bitte geben Sie einen Betrag in Euro an. Wenn Sie nichts für die Dinge ausgegeben haben, tragen Sie einfach eine 0 ein.

	Betrag in Euro
Größere Anschaffungen (z.B. Auto, Möbel, elektrische Geräte usw.)	
Artikel des täglichen Bedarfs (z.B. Lebens- und Genussmittel, Hygieneartikel, Reinigungsmittel oder Ähnliches)	
Bekleidung und Schuhe	
Freizeitaktivitäten (z.B. Restaurantbesuche, Kulturveranstaltungen, Fitnessstudio usw.)	
Mobilität (z.B. Kraftstoff, Fahrzeugkredite und laufende Kosten, Bus und Bahntickets)	
Reisen, Urlaub	
Wohnkosten (z. B. Miete, Wohngeld, Hypothekenrückzahlungen, sonstige Nebenkosten - ohne Energiekosten)	
Energie und Heizkosten	

Figure S15: Implemented German version of the consumption questions. English translation in Section S1.3

Wie sieht gegenwärtig Ihr Alltag in einer typischen Woche (Mo-So) aus?
Bitte geben Sie an, wie viele Stunden pro Woche Sie im Durchschnitt mit den folgenden Aktivitäten verbringen.

Hinweis: Addieren Sie dabei alle Aktivitäten einer Kategorie pro Woche auf. Da Aktivitäten sich überschneiden können, müssen sie sich **NICHT** auf den Gesamtwert von 168 Wochenstunden aufaddieren.

	Stunden pro Woche (Mo-So)
Erwerbstätigkeit (einschl. möglicher Überstunden und Wegezeiten)	
Aus- und Weiterbildung, Lernen (z. B. Kurse, Schule, Studium, Promotion)	
Sport (z. B. Vereinssport, Fitnessstudio, Yoga)	
Zeit für Freunde (z. B. gemeinsame Aktivitäten, Kommunikation)	
Zeit für Partnerschaft (z. B. gemeinsame Aktivitäten, Kommunikation)	
Zeit für Familie (z. B. gemeinsame Aktivitäten, Kommunikation)	
Ehrenamtliches, soziales oder politisches Engagement (z. B. Vereinsarbeit)	
Schlaf	
Information und Unterhaltung (z. B. Fernsehen, Kino, Theater, Spiele, Social Media)	
Besorgungen, Hausarbeit, Wohnung, Vewaltungsaufgaben (z. B. Rechnungen überweisen)	

Figure S16: Implemented German version of the time use questions. English translation in Section S1.3

Outcome	Treated	Control	ATE	SE	t-stat	p-val (N)	p-val (F)	n treated	n control
Aggregates									
Mental Health	4.305	4.019	0.286	0.078	3.667	0.000	0.000	107	1418
Purpose in life	3.003	2.881	0.122	0.067	1.814	0.070	0.070	107	1476
Life Satisfaction	4.511	4.185	0.326	0.068	4.801	0.000	0.000	107	1442
Aggregate components									
WHO-5 Depression	3.221	2.971	0.251	0.078	3.214	0.001	0.002	107	1445
PSS Stress	4.548	4.314	0.234	0.068	3.448	0.001	0.000	107	1470
Domain Satisfaction Index	4.441	4.115	0.326	0.064	5.117	0.000	0.000	107	1475
General Life Satisfaction	3.783	3.506	0.277	0.065	4.246	0.000	0.000	107	1445
Domain satisfactions									
Health satisfaction	3.106	2.935	0.171	0.065	2.632	0.008	0.012	107	1477
Sleep satisfaction	2.637	2.499	0.138	0.056	2.441	0.015	0.012	107	1477
Work satisfaction	2.638	2.572	0.066	0.056	1.182	0.237	0.264	107	1475
Income satisfaction	3.222	2.700	0.522	0.066	7.944	0.000	0.000	107	1477
Leisure satisfaction	2.777	2.574	0.203	0.053	3.859	0.000	0.000	107	1477
Social satisfaction	2.798	2.608	0.190	0.056	3.384	0.001	0.004	107	1477

Table S2: We report average treatment effects (ATE) in standard deviations for all outcomes in levels: mental health, purpose in life, life satisfaction, WHO-5 depression, PSS stress, the domain satisfaction index, general life satisfaction, and the six domain satisfactions separately. Inference is based on robust standard errors (SE), and Neyman (N) and Fisher's exact (F) p-values. We reject the null of no effect for all outcomes but work satisfaction and sleep satisfaction.

Outcome	Treated	Control	ATE	SE	t-stat	p-val (N)	p-val (F)	n treated	n control
Aggregates									
Mental Health	0.273	0.038	0.236	0.113	2.086	0.037	0.048	103	1104
Purpose in life	0.132	-0.175	0.307	0.113	2.717	0.007	0.000	104	1105
Life Satisfaction	0.187	-0.096	0.283	0.117	2.419	0.016	0.022	104	1084
Aggregate components									
WHO-5 Depression	0.349	0.049	0.300	0.119	2.522	0.012	0.016	103	1107
PSS Stress	0.138	0.017	0.121	0.107	1.136	0.256	0.250	104	1104
Domain Satisfaction Index	0.173	-0.128	0.301	0.119	2.540	0.011	0.020	104	1084
General Life Satisfaction	0.168	-0.041	0.209	0.116	1.795	0.073	0.086	104	1106
Domain satisfactions									
Health satisfaction	-0.069	-0.372	0.303	0.116	2.614	0.009	0.006	104	1107
Sleep satisfaction	0.097	-0.136	0.233	0.116	2.009	0.045	0.056	104	1107
Work satisfaction	-0.078	-0.193	0.115	0.121	0.950	0.342	0.382	104	1087
Income satisfaction	0.207	0.064	0.143	0.122	1.168	0.243	0.258	104	1104
Leisure satisfaction	0.373	0.160	0.213	0.118	1.810	0.070	0.070	104	1106
Social satisfaction	0.132	-0.021	0.153	0.100	1.528	0.127	0.128	104	1106

Table S3: We report average treatment effects (ATE) in standard deviations for all outcomes in changes for wave 7: mental health, purpose in life, life satisfaction, WHO-5 depression, PSS stress, the domain satisfaction index, general life satisfaction, and the six domain satisfactions separately. Inference is based on robust standard errors (SE), and Neyman (N) and Fisher's exact (F) p-values.

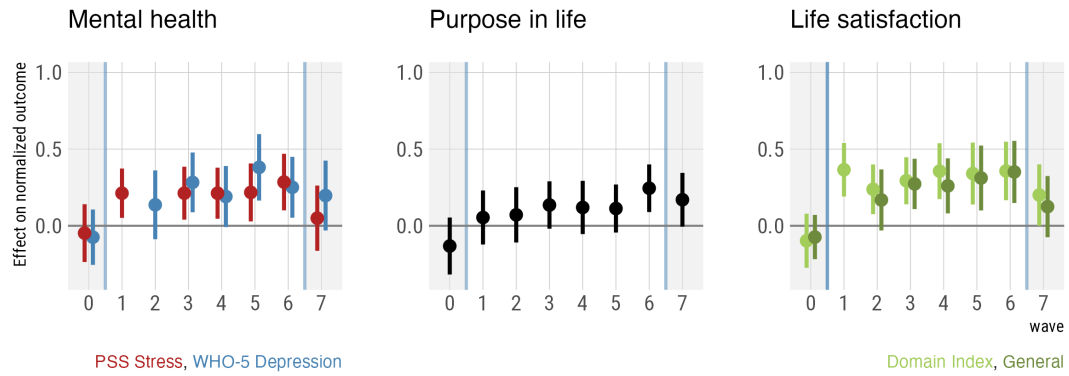


Figure S17: Unadjusted treatment effects (in standard deviations) and 95% confidence intervals for all outcomes and waves. Differences in mean normalized outcomes between treatment and control for each stratum, averaged across strata.

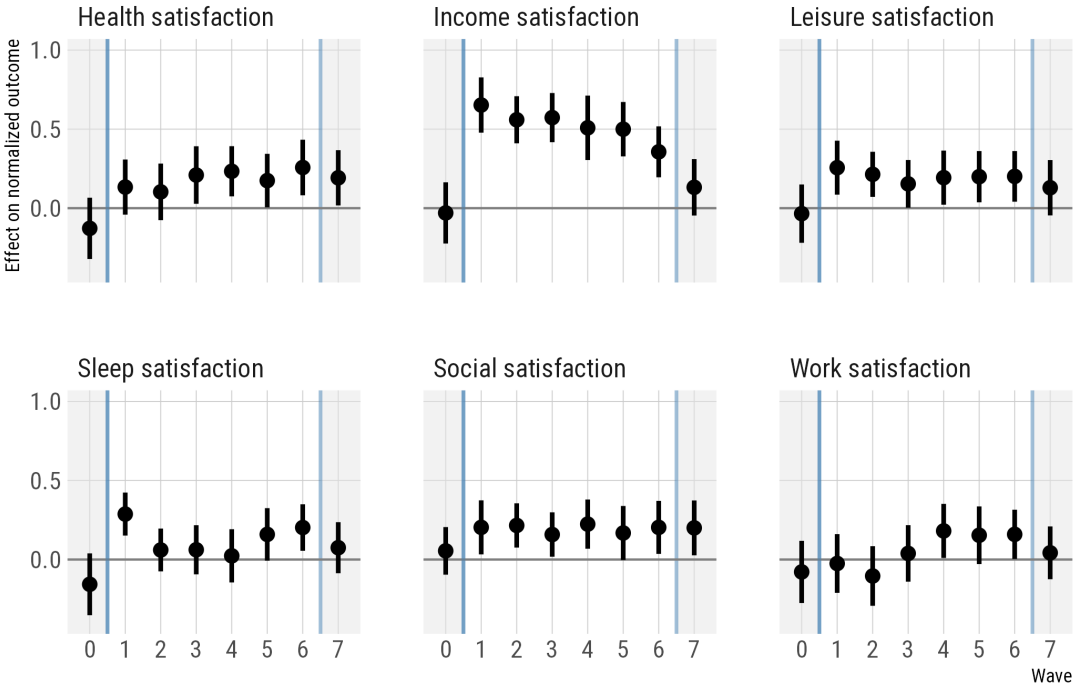


Figure S18: Unadjusted treatment effects (in standard deviations) and 95% confidence intervals for each wave and domain of life satisfaction. Differences in mean normalized outcomes between treatment and control for each stratum, averaged across strata.

Outcome	p-val (N)	p-val rank	BH cv	p-val (N) < BH cv	result robust to BH
Aggregates					
Mental Health	0.000	1st	0.004	yes	yes
Purpose in life	0.004	10th	0.038	yes	yes
Life Satisfaction	0.000	1st	0.004	yes	yes
Aggregate components					
WHO-5 Depression	0.000	1st	0.004	yes	yes
PSS Stress	0.000	7th	0.004	yes	yes
Domain Satisfaction Index	0.000	1st	0.004	yes	yes
General Life Satisfaction	0.000	1st	0.004	yes	yes
Domain satisfactions					
Health satisfaction	0.001	8th	0.031	yes	yes
Sleep satisfaction	0.001	8th	0.031	yes	yes
Work satisfaction	0.138	13th	0.050	no	yes
Income satisfaction	0.000	1st	0.004	yes	yes
Leisure satisfaction	0.008	11th	0.042	yes	yes
Social satisfaction	0.082	12th	0.046	no	yes

Table S4: We report the results of controlling for multiple hypothesis testing, according to Benjamini–Hochberg (BH) procedure and Newman p-values and $m = 13$, on our main results on positive treatment effects for the following outcomes in changes: mental health, purpose in life, life satisfaction, WHO-5 depression, PSS stress, the domain satisfaction index, general life satisfaction, income satisfaction, leisure satisfaction, health satisfaction, and sleep satisfaction. The Newman p-values of the respective estimated treatment effects are below their respective BH critical value (cv). Overall, we find that conclusions about statistical significance are hence robust for all results.

Outcome	p-val (F)	p-val rank	BH cv	p-val (F) < BH cv	result robust to BH
Aggregates					
Mental Health	0.000	1st	0.004	yes	yes
Purpose in life	0.008	10th	0.038	yes	yes
Life Satisfaction	0.000	1st	0.004	yes	yes
Aggregate components					
WHO-5 Depression	0.000	1st	0.004	yes	yes
PSS Stress	0.000	1st	0.027	yes	yes
Domain Satisfaction Index	0.000	1st	0.004	yes	yes
General Life Satisfaction	0.000	1st	0.004	yes	yes
Domain satisfactions					
Health satisfaction	0.002	8th	0.031	yes	yes
Sleep satisfaction	0.004	9th	0.035	yes	yes
Work satisfaction	0.168	13th	0.050	no	yes
Income satisfaction	0.000	1st	0.004	yes	yes
Leisure satisfaction	0.008	10th	0.038	yes	yes
Social satisfaction	0.082	12th	0.046	no	yes

Table S5: We report the results of controlling for multiple hypothesis testing, according to Benjamini–Hochberg (BH) procedure and Fisher's exact p-values and $m = 13$, on our main results on positive treatment effects for the following outcomes in changes: mental health, purpose in life, life satisfaction, WHO-5 depression, PSS stress, the domain satisfaction index, general life satisfaction, income satisfaction, leisure satisfaction, health satisfaction, and sleep satisfaction. Fisher's exact p-values of the respective estimated treatment effects are below their respective BH critical value (cv). Overall, we find that conclusions about statistical significance are hence robust for all results.

Outcome	p-val (N)	p-val rank	BH cv	p-val (N) < BH cv	result robust to BH
Aggregates					
Mental Health, changes	0.000	1st	0.002	yes	yes
Purpose in life, changes	0.004	19th	0.037	yes	yes
Life Satisfaction, changes	0.000	1st	0.002	yes	yes
Mental Health, levels	0.000	1st	0.021	yes	yes
Purpose in life, levels	0.070	23rd	0.044	no	yes
Life Satisfaction, levels	0.000	1st	0.002	yes	yes
Aggregate components					
WHO-5 Depression, changes	0.000	1st	0.002	yes	yes
PSS Stress, changes	0.000	1st	0.021	yes	yes
Domain Satisfaction Index, changes	0.000	1st	0.002	yes	yes
General Life Satisfaction, changes	0.000	1st	0.002	yes	yes
WHO-5 Depression, levels	0.001	14th	0.027	yes	yes
PSS Stress, levels	0.001	14th	0.027	yes	yes
Domain Satisfaction Index, levels	0.000	1st	0.002	yes	yes
General Life Satisfaction, levels	0.000	1st	0.002	yes	yes
Domain satisfactions					
Health satisfaction, changes	0.001	14th	0.027	yes	yes
Sleep satisfaction, changes	0.001	14th	0.027	yes	yes
Work satisfaction, changes	0.139	25th	0.048	no	yes
Income satisfaction, changes	0.000	1st	0.002	yes	yes
Leisure satisfaction, changes	0.008	20th	0.038	yes	yes
Social satisfaction, changes	0.082	24th	0.046	no	yes
Health satisfaction, levels	0.008	20st	0.038	yes	yes
Sleep satisfaction, levels	0.015	22nd	0.042	yes	yes
Work satisfaction, levels	0.237	26th	0.050	no	yes
Income satisfaction, levels	0.000	1st	0.002	yes	yes
Leisure satisfaction, levels	0.000	1st	0.002	yes	yes
Social satisfaction, levels	0.001	14th	0.027	yes	yes

Table S6: We report the results of controlling for multiple hypothesis testing, according to Benjamini–Hochberg (BH) procedure and Newman p-values and $m = 26$, in our main results on positive treatment effects for the following outcomes in changes and levels: mental health, purpose in life, life satisfaction, WHO-5 depression, PSS stress, the domain satisfaction index, general life satisfaction, and the domain satisfactions separately. The Newman p-values of the respective estimated treatment effects are below their respective BH critical value (cv). Overall, we find that conclusions about statistical significance are hence robust for all results.

Outcome	p-val (F)	p-val rank	BH cv	p-val (F) < BH cv	result robust to BH
Aggregates					
Mental Health, changes	0.000	1st	0.002	yes	yes
Purpose in life, changes	0.008	19th	0.036	yes	yes
Life Satisfaction, changes	0.000	1st	0.002	yes	yes
Mental Health, levels	0.000	1st	0.002	yes	yes
Purpose in life, levels	0.058	23rd	0.044	no	yes
Life Satisfaction, levels	0.000	1st	0.002	yes	yes
Aggregate components					
WHO-5 Depression, changes	0.000	1st	0.002	yes	yes
PSS Stress, changes	0.000	1st	0.002	yes	yes
Domain Satisfaction Index, changes	0.000	1st	0.002	yes	yes
General Life Satisfaction, changes	0.000	1st	0.002	yes	yes
WHO-5 Depression, levels	0.000	1st	0.002	yes	yes
PSS Stress, levels	0.000	1st	0.002	yes	yes
Domain Satisfaction Index, levels	0.000	1st	0.002	yes	yes
General Life Satisfaction, levels	0.000	1st	0.002	yes	yes
Domain satisfactions					
Health satisfaction, changes	0.002	17th	0.032	yes	yes
Sleep satisfaction, changes	0.004	18th	0.034	yes	yes
Work satisfaction, changes	0.168	25th	0.048	no	yes
Income satisfaction, changes	0.000	1st	0.002	yes	yes
Leisure satisfaction, changes	0.008	19th	0.036	yes	yes
Social satisfaction, changes	0.082	24th	0.046	no	yes
Health satisfaction, levels	0.008	19th	0.036	yes	yes
Sleep satisfaction, levels	0.022	22nd	0.042	yes	yes
Work satisfaction, levels	0.258	26th	0.050	no	yes
Income satisfaction, levels	0.000	1st	0.002	yes	yes
Leisure satisfaction, levels	0.000	1st	0.002	yes	yes
Social satisfaction, levels	0.000	1st	0.002	yes	yes

Table S7: We report the results of controlling for multiple hypothesis testing, according to Benjamini–Hochberg (BH) procedure and Fisher's exact p-values and $m = 26$, in our main results on positive treatment effects for the following outcomes in changes and levels: mental health, purpose in life, life satisfaction, WHO-5 depression, PSS stress, the domain satisfaction index, general life satisfaction, and the domain satisfactions separately. The Fisher's exact p-values of the respective estimated treatment effects are below their respective BH critical value (cv). Overall, we find that conclusions about statistical significance are hence robust for all results.

No. of missing waves	Mental Health	Purpose in life	Life Satisfaction	n control
0	2.049	3.010	4.298	1,124
1	2.032	2.970	4.183	127
2	2.100	3.153	4.436	80
3	2.070	3.020	4.286	70
4	2.134	3.138	4.359	55
5	2.001	2.974	4.286	73
6	2.225	3.075	4.449	51

Table S8: Average (non-normalized) responses to (i) all mental heath questions, (ii) the purpose in life question, and (iii) all questions regarding life satisfaction at baseline by control participants' number of missing waves later on in the RCT.

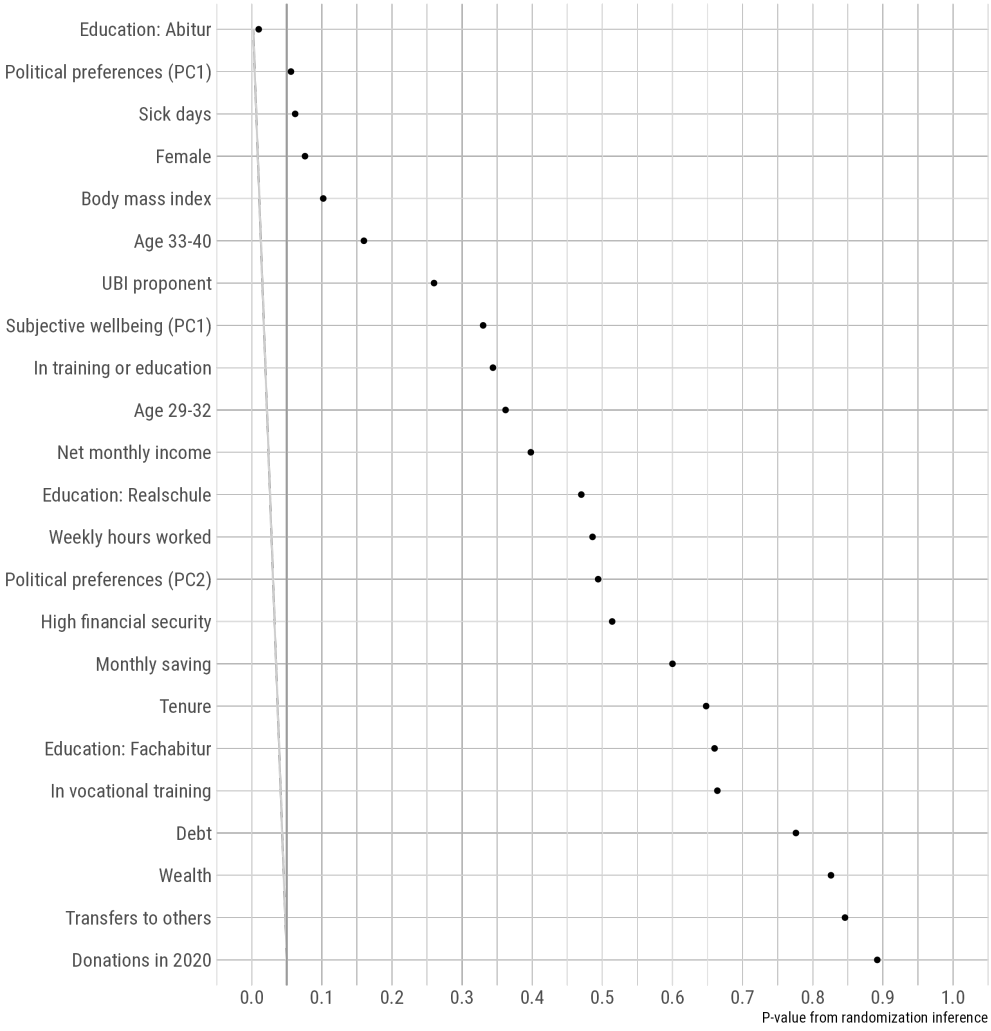


Figure S19: P-values for differences between treatment and control of baseline covariates, after restricting the sample to observations without any missing waves.

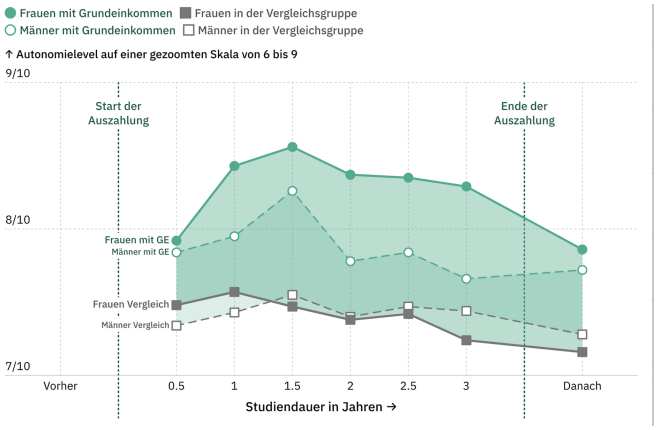


Figure S20: Purpose in life (non-standardized) over time, by treatment condition and gender. Text in plot in German. Translation available upon request.

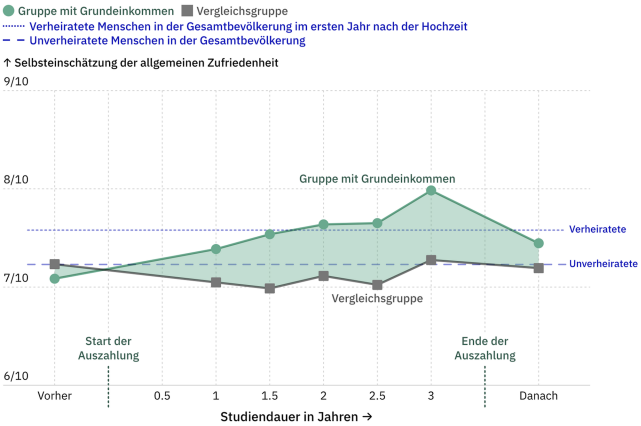


Figure S21: General life satisdaction (non-standardized) over time, by treatment condition. Text in plot in German. Translation available upon request.

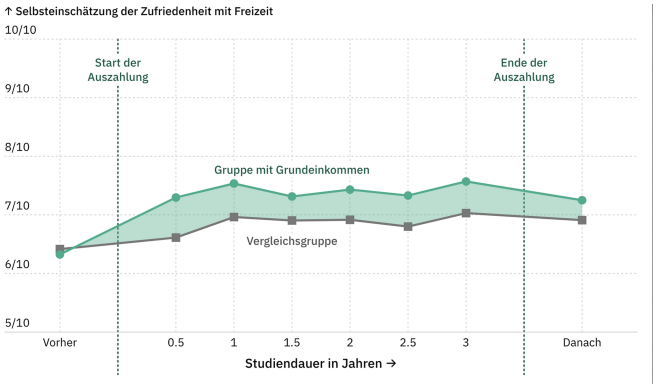


Figure S22: Leisure satisdaction (non-standardized) over time, by treatment condition. Text in plot in German. Translation available upon request.

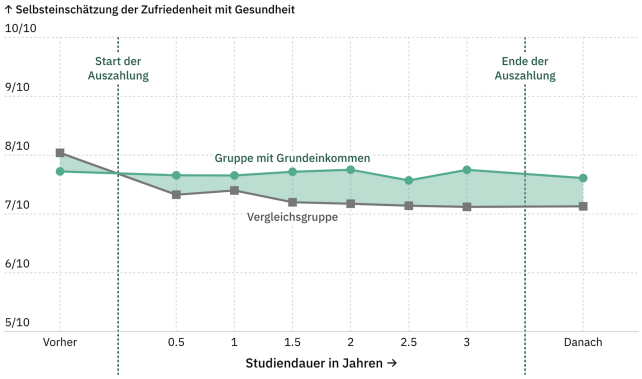


Figure S23: Health satisdaction (non-standardized) over time, by treatment condition. Text in plot in German. Translation available upon request.

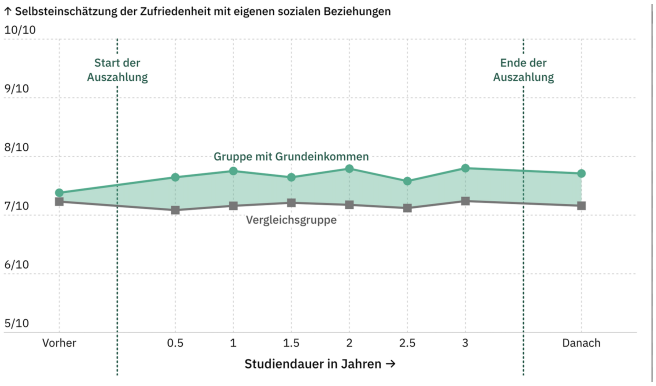


Figure S24: Social satisdaction (non-standardized) over time, by treatment condition. Text in plot in German. Translation available upon request.

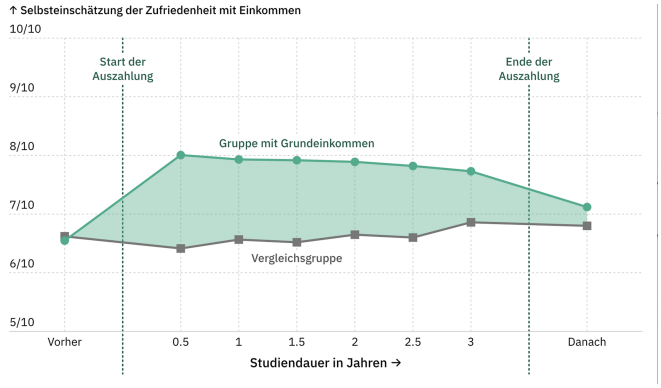


Figure S25: Income satisdaction (non-standardized) over time, by treatment condition. Text in plot in German. Translation available upon request.

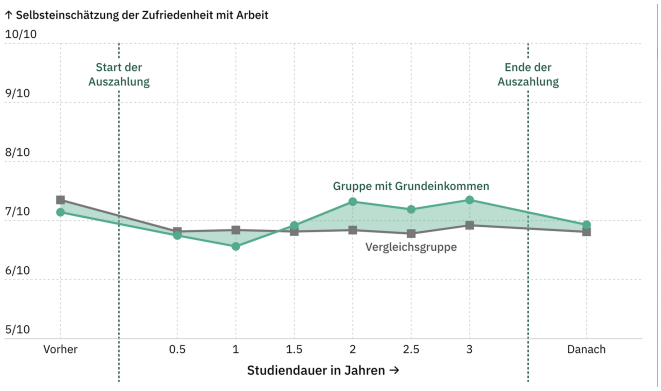


Figure S26: Work satisdaction (non-standardized) over time, by treatment condition. Text in plot in German. Translation available upon request.

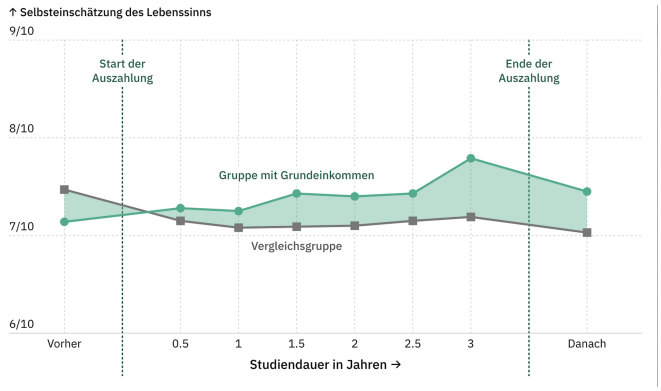


Figure S27: Purpose in life (non-standardized) over time, by treatment condition. Text in plot in German. Translation available upon request.

References

[1] Moore, R. T. and Schnakenberg, K. blocktools: Blocking, assignment, and diagnosing interference in randomized experiments. r package version 0.6-3, (2021).

[2] Athey, S. and Imbens, G. W. The econometrics of randomized experiments. In *Handbook of economic field experiments*, volume 1, 73–140. Elsevier (2017).

[3] Topp, C. W., Østergaard, S. D., Søndergaard, S., and Bech, P. The who-5 well-being index: a systematic review of the literature. *Psychotherapy and psychosomatics* **84**(3), 167–176 (2015).

[4] Likert, R. A technique for the measurement of attitudes. *Archives of psychology* (1932).

[5] Cohen, S., Kamarck, T., and Mermelstein, R. A global measure of perceived stress. *Journal of health and social behavior* , 385–396 (1983).

[6] Richter, D., Rohrer, J., Metzing, M., Nestler, W., Weinhardt, M., and Schupp, J. Soep scales manual (updated for soep-core v32. 1). Technical report, SOEP survey papers, (2017).

[7] Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. Sampling-based versus design-based uncertainty in regression analysis. *Econometrica* **88**(1), 265–296 (2020).

[8] Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**(1), 289–300 (1995).

[9] De Quidt, J., Haushofer, J., and Roth, C. Measuring and bounding experimenter demand. *American Economic Review* **108**(11), 3266–3302 (2018).

[10] Levitt, S. D. and List, J. A. Was there really a hawthorne effect at the hawthorne plant? an analysis of the original illumination experiments. *American Economic Journal: Applied Economics* **3**(1), 224–238 (2011).

[11] Easterlin, R. A. and Sawangfa, O. Happiness and economic growth: Does the cross section predict time trends? evidence from developing countries. *International differences in well-being* , 166–216 (2010).

[12] Kahneman, D. and Deaton, A. High income improves evaluation of life but not emotional well-being. *Proceedings of the national academy of sciences* **107**(38), 16489–16493 (2010).

[13] Vendrik, M. C. Adaptation, anticipation and social interaction in happiness: An integrated error-correction approach. *Journal of Public Economics* **105**, 131–149 (2013).

[14] VanderWeele, T. J. A unification of mediation and interaction: a four-way decomposition. *Epidemiology (Cambridge, Mass.)* **25**(5), 749 (2014).

[15] Haaland, I. K., Roth, C., Stantcheva, S., and Wohlfart, J. Measuring what is top of mind. Technical report, National Bureau of Economic Research, (2024).

[16] VanderWeele, T. J. *Explanation in causal inference: methods for mediation and interaction.* Oxford University Press, (2015).

[17] VanderWeele, T. J. Mediation analysis: a practitioner's guide. *Annual review of public health* **37**, 17–32 (2016).