

Camehl, Georg; Hahlweg, Kurt; Spieß, C. Katharina

Conference Paper

The Effects of a Parenting Program on Maternal Well-Being: Evidence from a Randomized Controlled Trial

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2018: Digitale Wirtschaft - Session: Health and Family Economics III, No. G11-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Camehl, Georg; Hahlweg, Kurt; Spieß, C. Katharina (2018) : The Effects of a Parenting Program on Maternal Well-Being: Evidence from a Randomized Controlled Trial, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2018: Digitale Wirtschaft - Session: Health and Family Economics III, No. G11-V1, ZBW - Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/181583>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Effects of a Parenting Program on Maternal Well-Being: Evidence from a Randomized Controlled Trial

Georg F. Camehl*, DIW Berlin and Freie Universität Berlin

Email: gcamehl@diw.de

Kurt Hahlweg, Technical University of Braunschweig, Institute for Psychology

Email: k.hahlweg@tu-bs.de

C. Katharina Spiess, DIW Berlin and Freie Universität Berlin

Email: kspiess@diw.de

Address: DIW Berlin

Mohrenstr. 58

10117 Berlin

Germany

**Corresponding author, Phone: +493089789268, Fax: +493089789115*

Acknowledgments: The authors thank Wolfgang Schulz, Marie LeMouel, Jan Marcus, and Sophia Schmitz for their valuable comments on previous versions of this paper. Moreover, we thank participants of the 3rd LEER workshop on Education Economics in Leuven, Belgium, for helpful comments. We also thank Adam Lederer for very helpful editorial assistance.

Abstract

This paper evaluates how the Triple P parenting program affects maternal well-being. We analyze data from a randomized controlled trial and a separate sample of mothers from a deprived neighborhood without a control group. For the latter, we generate a control group using SOEP survey data and evaluate the validity of this procedure. Overall, our results show a positive effect of Triple P on maternal well-being – with the largest effects appearing three years after treatment. Thus, we illustrate that maternal well-being is an additional channel through which parenting programs, as examples of early childhood interventions, benefit families.

Keywords: parenting program, family well-being, instrumental variables estimation, Triple P
JEL: I31, I26, J13, C21, C26

1 Introduction

Early childhood intervention programs can have both short and long-term impacts on child development. Economists extensively evaluate programs that focus directly on fostering cognitive skills or socio-emotional behavior of children. This literature covers the evaluation of targeted programs, such as the Perry Program or Head Start, as well as the effects of universal programs (see for example Barnett 2011, Heckman et al. 2010, Havnes and Mogstad 2011 or Peter et al. 2016).

However, many successful early intervention programs do not just address children but also work with parents – in particular mothers – to improve parenting skills. Some of these programs focus exclusively on parents while others combine this with components directly addressing the children. Comparatively less research exists on the effects of this program components on child development (for a summary of the few studies, see e.g. Heckman and Mosso 2014). This is particularly true for programs focusing solely on parents (that is parenting programs), albeit with the goal of improving child development.

Overall, there is little evidence how any of these programs affect parental well-being, in particular maternal subjective well-being. This is surprising, as maternal well-being is an outcome of interest for economists for several reasons: Well-being (seen by many researchers as a quantification of utility) and how it can be improved is of inherent interest as a core question in economics. It is also a central priority for policy makers in several countries (see for example Helliwell et al. 2016, OECD 2017). Moreover, maternal well-being is of interest as it is not only shown to improve child development (Berger and Spiess 2011, Dahlen 2016), but it also influences decisions on maternal labor supply and fertility (e.g. Sandner 2015). More broadly, well-being is associated with positive health outcomes (e.g. Diener and Chan 2011) and labor productivity (e.g. Oswald et al. 2015).

This paper analyzes the effects of an early childhood intervention on maternal well-being using data from both a randomized controlled trial (RCT) and an offering of the program free of charge to a group of more disadvantaged mothers for which no control group was recruited. More specifically, we evaluate how one particular parenting program – namely Triple P (Positive Parenting Program) – affects maternal well-being. Triple P is a “multilevel parenting and family support strategy designed to reduce the prevalence of behavioral and emotional

problems in preadolescent children.”¹ Thus, by providing parents with information on how to improve interactions with their children, it targets parents. While the program’s ultimate goal is to improve child outcomes, we argue that potential improvements in maternal well-being could be an additional benefit improving the cost-efficiency of the program.

We use the RCT data, the data from an additional offering of Triple P in a deprived neighborhood, and representative survey data. We apply a wide range of econometric techniques to analyze the effects of the program on maternal well-being. Specifically, we use instrumental variables regression (IV) with the data from the RCT to take into account selection out of the treatment (for those randomly assigned to the treatment group), thus improving over simple mean comparisons. Further, employing entropy balancing (Hainmueller 2011), we use data from the German Socio-Economic Panel Study (SOEP) to generate a control group for the case where we only have a treated group (‘non-experimental group’). This group is then used to evaluate the effect of taking part in Triple P for mothers living in a relatively deprived neighborhood to whom the program was offered free of charge.²

Our main contribution is the analysis of a substantial early childhood and family intervention that is not previously considered in the economics literature. This intervention addresses parents of young children in order to improve child outcomes, something rarely examined in the economics literature (see Heckman and Mosso 2014). In particular, we focus on maternal well-being, which is an important outcome for the mother, the child, the family, and the economy as a whole. We do this by focusing on subjective satisfaction as a measure for well-being, an approach widely used in the economic literature. Furthermore, we contribute to the literature by demonstrating that a parenting program has positive well-being effects for mothers in Germany, a country without a long tradition of parental programs. Germany provides a considerably different context from other countries where parenting programs exist and are analyzed more frequently, such as the US. Finally, we contribute to the body of research on methodology of program evaluation inspired by Lalonde (1986) by combining data from a small scale trial of the program with survey data to generate a control group. Subsequently, we validate this approach by showing that it yields similar results when applying it to the treatment group from the RCT.

¹ Sanders (1999, p 71); for more information also see Sanders (2012).

² The program was also offered free of charge to the treatment group of the experimental group.

The remainder of the paper proceeds as follows. Section 2 summarizes previous research on early childhood interventions and parental well-being with a focus on parenting programs. Section 3 presents a conceptual framework through which we analyze the effects of a parenting program on parental well-being. Section 4 presents the data, section 5 the methodology, and section 6 our main results. Section 7 provides additional evidence of the effects of a parenting program for the more disadvantaged group using an artificial control strategy. Section 8 concludes.

2 Previous research on early childhood interventions focusing on maternal well-being

The economic literature on parental and, in particular, maternal well-being as an outcome of early childhood interventions is relatively small.³ On the one hand, there are studies on the effects of center based child care programs on maternal well-being. For instance, Baker et al. (2008) show that a reform in Quebec, Canada, which led to an expansion of child care provision, adversely affected various child and family well-being outcomes, including parental life-satisfaction, paternal self-reported health, maternal depression, and work-family conflicts. For similar evidence see Brodeur and Connolly (2013), Kottelenberg and Lehrer (2013) and Herbst and Tekin (2014). Based on Australian data, Yamauchi (2010) shows a slightly positive correlation of local day care availability with maternal life-satisfaction. For Germany, studies by Schober and Stahl (2016) and Schober and Schmitt (2017) similarly point to a modest positive association between childcare provision and maternal well-being.⁴ Using a quasi-experimental approach, Schmitz (2017) shows that an expansion of child care provision for children aged three or older significantly increased maternal life-satisfaction in Germany.

In addition, other studies analyze day care programs with a parental component. Examples of these include evaluations of the Incredible Years Program (Webster-Stratton et al. 2004), Early Head Start (Love et al. 2005; for an overview of earlier research see Barnett 1995), the parental and home components of the Perry Preschool Program (Belfield et al. 2006 and Heckman et al. 2010), and the Abecedarian Program (Masse and Barnett 2002). Using data from the Head Start Impact Study, Ansari et al. (2016) and Ansari and Gershoff (2016) examine whether one year of Head Start differentially benefited parents as a function of their initial parenting behaviors. The outcomes they consider include maternal depressive symptoms, which only

³ If parental outcomes are included in studies outside economics, these most often refer to parenting practices or specific mental health aspects (Brooks-Gunn et al., 2000, give an overview of earlier research on the topic).

⁴ The studies of Schober and coauthors apply fixed effect models and make use of the regional variation in day care attendance in Germany.

showed improvements among parents most at risk. The Perry Preschool Program had home visits once a week on average. The Abecedarian program did not have home visits, but interacted with parents at the Abecedarian day care center. The evidence generally supports positive (complementary) responses of parents to interventions. However, these evaluation studies concentrate on parental behavior and attitudes as outcome or changes in the home quality only and do not take maternal well-being explicitly into account.⁵

On the other hand, there is a strand of literature on the effects of home visitation programs and exclusive parenting programs. Notable examples of home visitation programs include the Infant Health and Development Program (IHDP; Brooks-Gunn et al. 1994, Klebanov et al. 2001) and the Nurse Family Partnership (NFP; Olds 2006). The IHDP offers home visitations and focuses on health outcomes. Klebanov et al. (2001) analyze a randomized controlled trial of the IHDP with respect to maternal well-being. They provide an overall measure for well-being that includes aspects of mental health as well as general life satisfaction. Using data from an RCT in the United States, they find that overall the program improved maternal well-being, especially for those mothers without a high school degree. The NFP program provided home visits to first time teenage mothers, advising them on proper nutrition and care of young children, including the importance of cognitive stimulation. The evaluation by Olds (2006), using a randomized controlled trial, finds improvements for mothers through reduced dependence on public assistance. However, he does not analyze well-being measures explicitly. Lindsay et al. (2011) evaluate implementations of three parenting programs (Incredible Years, Triple P, and “Strengthening families, strengthening communities”) in the United Kingdom with respect to well-being. In their measure for well-being, they include maternal life satisfaction, but they do not separately report the effects of the program on this measure. Their measure for well-being mixes life satisfaction with mental health components. Furthermore, their study is not a randomized controlled trial, rather it compares the different parenting programs. Their analyses show large improvements in their well-being measure from before to after participation in both Triple P and Incredible Years (0.81 and 0.88 of a standard deviation, respectively), but without comparisons to a control group. Sandner (2015) evaluates a German home-visitation program for disadvantaged mothers that was modelled after the NFP. Using a randomized controlled trial and administrative data, his study finds evidence of both improved

⁵ The Incredible Years Program (Webster-Stratton et al., 2004) focuses on positive parenting for parents of children younger than eight. It aims to reduce parent-child violence. Therefore, it is not a program that exclusively aims to improve child outcomes by providing information to parents.

maternal well-being and decreased labor market participation for treated mothers. In the economics literature, this study is closest to our own, as it also analyses the effects of a parenting program on maternal well-being in the German context.

The parental program we analyze is the Triple P program, which focuses on providing information that improves parenting skills with respect to handling child behavior and reducing parental stress. In this sense, it directly aims at improving parental well-being. Existing psychological studies analyze the effects of Triple P. Meta-analyses by Novak and Heinrichs (2000) and Sanders et al. (2014) reports positive effects of Triple P on outcomes such as parenting style, parenting efficacy and satisfaction, as well as parental adjustment and relationships. However, no study explicitly focuses on parental satisfaction.

3 Conceptual framework

To provide an economic perspective on the effects of the parental program on maternal well-being, we consider a simple framework of household production.⁶ The underlying assumption of this framework is that well-being can be regarded as a suitable quantification of utility. A person's utility function is given by $U(C, L)$, where C is the amount of goods and services consumed and L is leisure time. Individuals have an endowment of time (L_0) that they can use either for work (where L_B is time used on the labor market, L_H time used for home production) or for leisure:

$$L = L_0 - L_B - L_H$$

Components of C can be either bought (C_B) or produced within the household (C_H) and their size is measured in terms of monetary value:

$$C = C_B + C_H$$

The amount of goods that can be bought (C_B) will typically depend on the wage rate of the person. The component of home-produced goods and services (C_H) most relevant here is child care.⁷ A production function on C_H can be given by:

⁶ For general models of household production see Cahuc et al. (2014). There is also an economic literature on parenting style as an investment in child development (see Del Boca et al. 2014 and Cobb-Clark et al. 2016). However, parenting skill is a distinct concept from parenting style and more related to actual parenting practices; see Darling and Steinberg (1993). Therefore, we argue that we can model the effects of parenting skills more precisely than the previous literature on parenting style.

⁷ Child care, in a broad sense, is the most relevant aspect of home production but other aspects could also be influenced by parenting skills or child behavior.

$$C_H = q(p)L_H$$

We assume that the size of C_H can be influenced both by the amount of goods and services provided (for example the number of hours of child care) but also its quality (for example the quality of interactions with children to develop skills). Thus p is a measure for parenting skills and $q(\cdot)$ is a function determining the amount and quality of the produced services and, thus, its value. In effect, if the persons' parenting skills are higher, they get a higher amount of C_H per invested time unit. We make no further statement about the form of $q(\cdot)$ but note that it may cover situations where certain aspects of home production are independent of parenting skills while others are not. To sum up, in this simple model of labor supply including home production, an improvement in parenting skills has effects analogous to an increase in the wage rate for work on the labor market and, thus, theoretically increases – *ceteris paribus* – individuals' utility from assigning their time to different activities.

Additionally, children could also affect their parents' utility directly when L includes a component indicating the quality of the time parents spend with their children. If the children behave in a way that is agreeable to the parents, this may increase their valuation of the time spent together.⁸ We make no further claim about the allocation of time to different activities. However, the model serves as an illustration of the different channels through which improved parenting skills, due to participation in a parenting program, can lead to an increase in parental well-being in the framework of an economic model. This is because there is a direct channel through increased home productivity and an indirect channel through improved child behavior that leads to a higher valuation of leisure time.

4 Data and descriptive evidence

Datasets. We use data from a unique intervention study (the so-called *Projekt Zukunft Familie 1 – ZF1*,⁹ Heinrichs et al. 2006) of Triple P in Germany. In this study, a randomized control trial (RCT) was carried out in one German city in 2001 - the medium sized city of Braunschweig in the state of Lower Saxony.

Triple P can be administered in five different intensities, referred to as levels. The first two levels consist of broadly spreading information through videos and leaflets. Level three

⁸ This could also be an explanation for the negative employment effects of the home visitation program analyzed by Sandner (2015): Mothers who participated reported higher well-being but, at the same time reduced, employment. This could be due to a higher valuation of leisure time with their children.

⁹ This study was funded by the German Science Foundation (DFG).

includes group training, but with only a narrow focus on parenting skills, whereas level four broadens the focus to include general improvements to the home environment. Level five is an intensive family intervention resembling and going beyond the aforementioned home-visitation programs.

In the setting we analyze, parents participated in level four of Triple P. The participants were recruited from day care centers in a middle class neighborhood.¹⁰ The parent training took place in a group setting with an average of six parents, with telephone follow-ups. The Triple P training includes video lectures as well as role-playing to learn how to handle difficult situations under the supervision of a trained member of the project team. The treatment consisted of four sessions, each lasting two hours that were conducted in the rooms of the day care center that the child of the parents attended at that time and four optional telephone follow-ups. The data includes 280 families of children between 2.6 and 6 years of age. All children attend day care centers and all parents speak German. The field experiment was conducted with 186 families in the treatment and 94 families in the control group (ZF1 from here on). Single mothers are excluded from the analyses, as differential effects of parenting programs are expected for them (Heinrichs et al. 2009) and because the sample size of single mothers alone is too small for a separate analysis. Panel attrition of the remaining sample was low: of 219 mothers who participated in the study, still 206 answered the questionnaire three years later; an attrition rate of 6%. All subsequent analyses use the sample of mothers for whom information is available for the whole time period.

Outcomes. We use the satisfaction of the mother with her family life as our measure for maternal well-being.¹¹ More specifically, we analyze the differences between pre-treatment satisfaction with family life and measurements of it at four different points in time after the treatment. We use this satisfaction variable for two reasons: First, Triple P aims at improving parent-child-interactions and, thus, changes in family life satisfaction are reasonable. Second, this measure of satisfaction is also available in a representative survey dataset that we use for additional analyses.¹² Furthermore, Schober and Stahl (2016) show that satisfaction with family life is the well-being measure that is most affected by early childhood programs compared to satisfaction measures related to other areas of life. We follow the practice in the

¹⁰ Attendance of day care was 78.2% in the relevant age range in 2001 in Germany (Statistisches Bundesamt 2002, own calculations).

¹¹ For more information on the scale see Henrich and Herschbach (2000).

¹² See section 7 below.

well-being literature and assume the satisfaction variable to be on a cardinal scale.¹³ We use standardized measures of the well-being measure: Differences of well-being were Z-standardized for each point in time such that coefficients of regression models correspond to differences in terms of standard deviations.

In order to obtain a more comprehensive view of the influence of Triple P on maternal well-being and in line with the theoretical model presented above, we analyze two additional outcome measures: First, we consider a variable that measures parenting skills (see Arnold et al. 1993). The ZF1 study measures parenting skills as strategies and actions concerning the upbringing of children.¹⁴ Second, we consider a child related outcome measure, the Child Behavior Checklist (CBCL, Achenbach and Rescorla 2000, Döpfner et al. 2014). The CBCL is a widely used and validated scale that measures child behavior. Parents answer items on noticeable behavioral and emotional problems, indicating how frequently these occur. The CBCL can be grouped into two subscales, one on internalizing behavior (for example displayed through depressive symptoms) and one on externalizing behavior (for example aggressive behavior toward others). We use this scale as another outcome measure to learn more about a potential indirect channel of parenting via child behavior on maternal well-being.

Covariates. Our estimations control for various covariates. Table 1 includes pre-treatment descriptive statistics. Children are an average of four years old, the gender ratio is roughly balanced, and children have an average of one sibling. Households with net income above 1,500€ are considered as high income households.¹⁵ Additionally, more participants in the ZF1 study have earned a university entry degree ('Abitur').

5 Empirical strategy

5.1 General strategy and evaluation of the RCT (ZF1)

¹³ Although life satisfaction is principally a latent variable, in many surveys respondents are asked to grade it on an ordinal scale (Schröder and Yitzhaki 2017 for a discussion.). Researchers then usually treat these answers as cardinal variables, which assumes that all respondents interpret the question in a similar way (that is, the distances between items in terms of the latent underlying variable are equal).

¹⁴ The German version (Naumann et al. 2010) of the parenting scale (for the English version, see Arnold et al. 1993) is administered to assess parenting skills. The scale is a 35-item questionnaire that measures dysfunctional discipline styles in parents. It yields a total score based on three factors: *laxness* (permissive discipline), *over-reactivity* (authoritarian discipline, displays of anger, meanness, and irritability) and *verbosity* (overly long reprimands or reliance on talking). Higher scores indicate dysfunctional parenting behavior. The total score has adequate internal consistency, good test-retest reliability ($r = 0.84$), and reliably discriminates between parents of clinic and non-clinic children.

¹⁵ The study includes income on an ordinal scale with seven steps of around 500€. However, data alignment to make additional analyses feasible (presented below) required us to generate a binary variable on income to have sufficient overlap with other groups.

Since the ZF1 data come from a RCT study, identification of causal effects is straightforward. The randomized assignment of individuals to either treatment- or control-group provides exogenous variation that makes it possible to interpret mean differences as causal. If randomization works perfectly, there is no systematic unobserved heterogeneity between the individuals in the treatment and control group. Therefore, additional econometric techniques are not necessary. However, we know from the implementation of the RCT that some mothers choose not to take part in the treatment even though they were assigned to the treatment group. If this selection is related to any characteristics that also influences the outcome, exogeneity of the treatment is violated and estimates of causal effects are likely to be biased. To learn more about control and treatment group differences, we verify that covariates are balanced between treatment and control group. We also do this comparison for those actually treated and for those not treated. Naturally, this only works for observable characteristics. Table 2 shows the mean differences of observable characteristics between groups after randomization and between those who subsequently took up the treatment and the remaining control group. The table shows that there are two variables for which the randomization did not seem to result in similar means for the treatment and control group: Mothers in the treatment group are statistically significantly more likely to have a college degree and less likely to only have a vocational degree. If we consider the actual treated and the actual control group, mothers in the treatment group are still less likely to have vocational training, but also more likely to not have a tertiary degree. This indicates that there was likely no significant selection into actual treatment participation based on observable characteristics apart from a higher likelihood of mothers without a tertiary degree to opt out of the treatment.

Nevertheless, this does not rule out selection into or out of the actual treatment based on unobservable characteristics. In other words, if mothers who choose not to take part in the treatment have other, unobserved, characteristics that also affect their well-being, mean differences between the treatment and control groups are biased estimates of the causal effect of the treatment. This problem can be mitigated using an instrumental variable (IV) approach. The assignment to the treatment group constitutes a valid instrument, as it is uncorrelated with unobserved characteristics of the mothers via the randomization. Therefore the exogeneity assumption likely holds for this instrument. At the same time, it is highly correlated with taking up the treatment and, thus, also a relevant instrument. Moreover, we use covariates in our estimations to further increase the precision of our estimates.

In addition to looking at the effects of Triple P on maternal well-being at different points in time, we follow the procedure laid out by Anderson (2008) and also consider a summation index of the outcome over all four points in time when the well-being was measured as a robustness check. The summation index is a weighted average generated with predictions from a generalized least squares (GLS) model including only a constant. In this way, the weights are set according to the covariance of the outcomes per individual in order to maximize the amount of information. The summation index has an advantage for interpretation as it gives an indication of *overall* effectiveness of the parenting program. This procedure reduces the number of tests concerning the overall effectiveness of the treatment to one, making the analyses additionally robust to overtesting.

6 Results

6.1 Main results

Table 3 shows the treatment effect of Triple P participation on maternal well-being. In principle, there are four different outcomes, each given by the difference between the well-being measurement at a given point in time and the pre-treatment measurement. First, we present the OLS results based on actual program participation for ZF1. We regress on the pre-post difference in the well-being including only the pre-treatment value of well-being (column 1) to control for potential level effects and, in a second step, all covariates (column 2). These estimations show medium sized effects (Cohen 1988), which are only statistically significant for the second and third year following treatment (column 2). The effect of the treatment on the summation index is also significant. As explained above, these estimates may be biased because of non-random take-up of the treatment. Using the instrumental variable approach (IV, again with control variables) for ZF1, standard errors increase, but effect sizes are not significantly different from the OLS results (column 3).¹⁶ The increasing standard errors are not surprising as IV only partly uses the variation between treatment and control groups to estimate the difference. The slightly decreasing effect sizes could be explained by the selection into the treatment: Likely explanations are heterogeneous treatment effects and selection based on these (some mothers who would have been less affected by the treatment could have chosen to not participate, thus biasing the difference in means upwards) or unobserved heterogeneity in their characteristics (something that makes them less likely to participate in the training and

¹⁶ T-tests for equality of the effects between the OLS results without covariates and the IV-model yielded p-values with a minimum value of 0.3701 for the effect on the difference after three years. Note that this corresponds to the most conservative way to test for differences between effects as all other differences are smaller.

makes them unhappier at the same time). The positive effect of the treatment seems to become most apparent after a few years instead of directly after the treatment.

Overall, the increases in well-being of between 20 and 30% of a standard deviation are quite sizeable when compared results in the literature. For example, Yamauchi (2010) finds an increase of satisfaction with free time by 16% of a standard deviation when day care is available in the area but no increase in satisfaction with family life. Schmitz (2017) finds an increase of general life satisfaction of 30% of a standard deviation if the child attends day care. However, day care is comparatively more costly than the Triple P intervention. Sandner (2015) finds a 15% increase of general life satisfaction after participating in the Pro-Kind program, a slightly lower effect.¹⁷

6.2 Robustness

In the main specifications above, we use changes in the well-being measures as outcomes in a linear model. Since the original measures come from a five-point ordinal scale it is *a priori* unclear if this is a valid approach. Thus, we also estimate ordered logit models as a robustness check.¹⁸ For this to be feasible, groups must be collapsed because large changes are rare in absolute terms: While, in principle, the maximum change in well-being is four points, the majority of observations have changes of -1, 0, or 1, with only a few changes of larger magnitude. Given these limitations and some loss in statistical power, we find that our results are not sensitive to our choice of using well-being as a metric variable (Table 4). The effects on the probability of positive and negative changes are very similar in magnitude, thus providing evidence of a linear relationship between the treatment and the outcome measure.

6.3 Channels

As described above, we analyze two potential channels for the change in maternal well-being. Table 5 gives the results of the IV-regressions on parenting skills and child behavior. Outcomes are recoded such that higher values correspond to better parenting skills and less problematic behavior by the child, respectively. Column 1 shows a strong effect of the treatment on

¹⁷ To put these results into perspective, these effects are similar in size to the effects of involuntary job loss, which is considered to be one of the most detrimental shocks with respect to well-being. Kassenboehmer and Haiken-DeNew (2009) report effects of between -0.230 and -0.531 points on an 11-point Likert-scale for women in Germany, which translate into effect sizes of -13% to -31% of their reported standard deviations. The effects for men in this study are between .0.702 and -0.724 (-41% to -42%).

¹⁸ The downside of a more complex model is, of course, a loss of statistical power, so given that both models provide similar estimates we prefer the simpler model.

parenting skills. This change in the home environment induced by changing parental skills seem to be permanent. Positive effects for child behavior are smaller and driven by the CBCL internalizing behavior subscale (columns 3). These are most pronounced and strongly statistically significant in the two-year follow-up. For the other points in time, effects on the internalizing subscale are smaller, but still significant in two out of three measurement points. Our results show that both improved parenting skills and less problematic child behavior could be channels through which the interventions improves maternal well-being; these are in line with our theoretical model of utility and home production. However, in this case, the improvements in parenting skill seem to be the more important factor.

7 Additional analyses for the non-experimental group of more disadvantaged families

7.1 Background and data

In addition to the randomized controlled trial evaluated above, Triple P was independently offered to mothers in a deprived neighborhood of Braunschweig. This is part of another intervention study, the so-called *Projekt Zukunft Familie 2*¹⁹ – ZF2 (Heinrichs and Jensen-Doss 2010). The analysis of this sample shows how the intervention affects potentially disadvantaged mothers. This group is of particular interest with respect to avoiding inequalities in parental behavior early on. This intervention study includes an additional 197 treated families. However, no control group exists for this study. In order to evaluate the effect of the treatment for the participants in ZF2, we use survey data from the German Socio-Economic Panel Study (SOEP) to generate a control group.²⁰ Thus, analyzing this sample is relevant from a methodological perspective as well.

The SOEP is the largest household panel study in Germany (Wagner et al. 2007). We use SOEP waves from 2010 through 2014, as these are the first to include a question on family life satisfaction that is identical to that in the ZF2 data.²¹ Overall, 2,548 individuals with children under the age of six answered the personal questionnaire for all four waves in the SOEP. Furthermore, we restrict the sample to mothers whose child is between 2.6 and 6 years old and attended a day care center at the time of the interview. This reduces the sample size to 2,120

¹⁹ This study is funded by the Jacobs Foundation.

²⁰ Additional information on how this is provided in the empirical strategy section.

²¹ This is also the same measure as in the ZF1 data.

observations.²² In addition, we make use of data on the regional level, provided by *Federal Statistical Office* and *Statistics Braunschweig* (see section 7.2 for details).

An advantage of creating a control group using SOEP data is that in addition to the well-being measure we analyze, it includes a rich set of regional, socioeconomic, and sociodemographic characteristics. One challenge of the dataset is that the outcome is measured on slightly different scales in the various samples we use. While the ZF2 study uses a five point Likert-type scale, SOEP uses an eleven-point scale. Thus, we transform the eleven-point scale into a five point scale. Another challenge of the dataset is that the well-being information is measured ten years later than in the SOEP. This question might be crucial with respect to the outcome variable. Additionally, there might be differential trends stemming from unobservable shocks to one of the datasets. In general, we assume that the trajectory of satisfaction with family life in our control group is a good proxy for that in the treatment group given all observables. For evidence on the validity of this assumption and further discussion of both mentioned issues, see appendix A.

Pre-treatment satisfaction with family life is lower in the ZF2 study compared to the SOEP. (see table 6). With respect to covariates, a large difference between the ZF2 data and the other two datasets emerges: Less than a quarter of families fall into the high income group compared to around 60 percent for the two other subsamples. We expect this because the sampling of ZF2 families focuses on relatively deprived neighborhoods. The education variables also show a similar picture: participants of the ZF2 study have, on average, lower school degrees, with a smaller share of mothers having earned a college degree.

7.2 Method: Matching with a generated control group

All individuals in the ZF2 subsample were treated. To estimate the treatment effect with this data, additional data are needed. Therefore, we create a control group using SOEP data. We then use this control group to estimate the effects of the treatment via entropy balancing (Hainmüller 2012). Entropy balancing is a weighting method to preprocess the data such that the means and potentially higher moments of the generated control group matches the treatment data. It can be regarded as a matching method²³ and, thus, the conditional independence

²² Sufficient knowledge of German was an additional prerequisite for taking part in ZF2, but there are no SOEP observations that had to be excluded due to insufficient knowledge of German.

²³ For a discussion of the differences of regression and matching estimators with special regard to entropy balancing, see the empirical strategy section of Anger et al. (2017).

assumption (selection on observables) is necessary for identification of causal effects. The corresponding balancing graphs are shown in Appendix B. To further control for differences between the treatment and the generated control group, we add regional administrative data. In principle, we would like to use all the background information that is available to us to generate a control group. However, this can make the optimization algorithm used for entropy balancing infeasible for two reasons (Hainmüller 2012): (1) Too many covariates compared to the number of observations, and (2) too large differences between the moments of the two groups for specific characteristics. To solve these problems, we conduct the following additional procedure: Based on the regional information, we generate subgroups of the control group: Each subgroup covers a certain share of SOEP individuals. Which individuals are in each subgroup depends on a measure of similarity between the individuals' home municipality and Braunschweig. Given the set of regional characteristics for Braunschweig, we calculate a similarity index for the regions where the SOEP observations come from. As a measure of similarity, we calculate the Mahalanobis distance²⁴ based on regional indicators. For each region, the measure is given by

$$d_i(x, y) = \sqrt{\sum_{j=1}^k \left(\frac{x_j - y_{ij}}{sd(y_j)} \right)^2}$$

where i is the indicator for the region and j denotes the different characteristics, as shown in the data section. Observations are then ordered according to this similarity measure.

The regional information is on the county (*Kreise* in German) or the municipality level (*Gemeinde* in German).²⁵ However, the two sample areas of the ZF2 data set are lower regional levels than a county or municipality. For this lower level, *Statistics Braunschweig* provides us with the same information given on the upper levels. The selected indicators are the following: The unemployment rate, the percentage of families on welfare, the percentage of foreigners, and the average living space per dwelling in square meters. In addition, we include the socioeconomic background of families. Table 7 shows these regional indicators. Overall,

²⁴ The Mahalanobis distance is a generalization of the Euclidean distance including a z-standardization (subtracting the mean and dividing by the standard deviation) of the distance for each characteristic.

²⁵ We used data from the most recent German census (2011), accessed from <https://ergebnisse.zensus2011.de/> (download November 2017).

means of the participants in ZF1 are relatively similar to those in our SOEP sample. By definition, regional characteristics for ZF2 show a comparatively disadvantaged background.

As previously noted, each observations is assigned a similarity measure based on regional characteristics. We then create different control groups based on the quantiles in which an observation lies on the similarity distribution. We define ten cutoffs, starting with the first decile of the distribution, continuing in ten-percentage-point steps. In the main specification, we take the 20% cutoff (for the reason of this see next paragraph). Results for other specifications are presented in Appendix B.

A way to validate this entire procedure is to apply it to the ZF1 sample. Thus, we create another artificial control group with the same procedure as described above. When we rerun the models using the ZF1 treatment group and the new artificial control group we find that the specification that comes closest to the original estimation is the one using the 20% cutoff, providing the reason for choosing the 20% cutoff for our main specification with the ZF2 treatment group. Results for other cutoffs are presented in Appendix B.

7.3 Results

Table 8 first shows the comparison of the experimental group (ZF1) and the generated control group (column 2). It also repeats the main results from section 6 as a reference (column 1). As mentioned above, we calculate effects using the experimental group in order to validate our control group creation approach.²⁶ Effects using the generated control group are, in general, somewhat smaller than those from the evaluation of the RCT, except for one case, the measurements after one year. However, they show a similar pattern with the largest effects emerging after three years. Next, we consider the effect of the treatment in a relatively disadvantaged neighborhood (ZF2) using the generated control group (column 3). We also find positive effects of the treatment for people from the disadvantaged neighborhood.

We check the assumption necessary to treat well-being as a metric variable and test for a linear relationship between treatment and outcome as a robustness check. The test indicates that linearity seems to hold.²⁷

²⁶ In the appendix, the full results for different subgroups of the artificial control groups are shown. Here we include only the 20% subset to show that, in this case, the results of this approach come closest to the results derived from the analyses of the RCT.

²⁷ This is analogous to the robustness check in section 6.2. The table of results is included in appendix C.

8 Conclusion

This study analyzes the effect of a parenting program on maternal well-being using both experimental and non-experimental methods. In general, there is little literature on the effects of interventions aimed at improving child outcomes on parental well-being. The economic literature on the evaluation of parenting programs almost exclusively focuses on child outcomes. To our knowledge, this is the first paper evaluating the Triple P program with respect to maternal well-being in general and the first paper on Triple P from an economic perspective. Unlike most parenting program evaluations, this paper uses data from a context, namely Germany, with a universal day care system and an underdeveloped infrastructure of parenting programs.

Using two different implementations of the program, as well as additional administrative and survey data, we find that the program has a positive medium- to long-run effect on maternal well-being. We measure a small effect one year after the treatment that could be attributed to changes in parenting style. This is in line with empirical evidence for this channel in the literature on the effects of parenting programs on parental skills. In the long-run, effects on maternal well-being are larger and increasing. A potential mechanism for this is the improvement in child outcomes, which takes time to materialize. We find the improvement of maternal well-being noteworthy as Triple P is a low cost intervention yielding comparatively large effects on maternal well-being. Moreover, our analysis shows that evaluations of parenting programs that only take child outcomes into account may underestimate the benefits of the program. Improved maternal well-being has positive consequences for both the family and society. For policy makers who are interested in increasing the well-being of mothers with young children, early childhood interventions addressing parents are an effective tool.

References

- Achenbach, T. M. and L. A. Rescorla (2000): *Manual for the ASEBA preschool forms & profiles: An integrated system of multi-informant assessment; Child behavior checklist for ages 1 1/2-5; Language development survey; Caregiver-teacher report form*. University of Vermont.
- Anger, S. G. Camehl and F. Peter (2017): Involuntary Job Loss and Changes in Personality Traits, *Journal of Economic Psychology* 60: 71-91.
- Ansari, A., K.M. Purtell and E. T. Gershoff (2016): Parenting Gains in Head Start as a Function of Initial Parenting Skill, *Journal of Marriage and the Family* 78 (5): 1195–1207.
- Ansari, A. and E. T. Gershoff (2016): Parent Involvement in Head Start and Children's Development: Indirect Effects Through Parenting, *Journal of Marriage and Family* 78(2): 562–579.
- Anderson, M.L. (2008): Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects, *Journal of the American Statistical Association* 103 (484): 1481-1495.
- Arnold, D. S., O'Leary, S. G., Wolff, L. S., & Acker, M. M. (1993): The Parenting Scale: A measure of dysfunctional parenting in discipline situations. *Psychological Assessment* 5 (2): 131-144.
- Achenbach, T.M., and Rescorla, L.A. (2000): *Manual for the ASEBA Preschool Form & Profiles*. Burlington: University of Vermont, Department of Psychiatry.
- Baker, M., J. Gruber and K. Milligan (2008): Universal child care, maternal labor supply, and family well-being, *Journal of Political Economy* 116 (4): 709-745.
- Barnett, W.S. (1995): Long-term effects of early childhood programs on cognitive and school outcomes, *The Future of Children*, 5 (3): 25-50.
- Barnett, W. S. (2011): Effectiveness of Early Educational Intervention, *Science* 333: 975-978.
- Belfield, C. R., M. Nores, S. Barnett und L. Schweinhart (2006): The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Follow up, *Journal of Human Resources* 41 (1): 162–190.
- Berger, E. M. and C.K. Spiess (2011): Maternal Life Satisfaction and Child Outcomes: Are they related?, *Journal of Economic Psychology* 32 (1): 142-158.
- Brodeur, A. and M. Connolly (2013): Do higher child care subsidies improve parental well-being? Evidence from Quebec's family policies, *Journal of Economic Behavior & Organization* 93: 1-16.
- Brooks-Gunn, J., L.J. Berlin and A.S. Fuligni (2000): Early childhood intervention programs: What about the family? In: J. P. Shonkoff and S. J. Meisels (eds.): *Handbook of Early Childhood Intervention*. Cambridge University Press.
- Brooks-Gunn, J., M.C. McCormick, S. Shapiro, A.A. Benasich and G.W. Black (1994): The Effects of Early Education Intervention on Maternal Employment, Public Assistance, and Health

- Insurance: The Infant Health and Development Program, *American Journal of Public Health*, 84 (6): 924-931.
- Cahuc, P., S. Carcillo and A. Zylberberg (2014): Labor Economics. 2nd edition. MIT Press.
- Cobb-Clark, D., Salamanca, N. and A. Zhu (2016): Parenting Style as an Investment in Human Development, *IZA Discussion Paper No. 9686*.
- Cohen, J. (1988): *Statistical power analysis for the behavioral sciences*. Hillsdale. NJ: Lawrence Earlbaum Associates, 2.
- Cunha, F., J. J. Heckman (2009): The Economics and Psychology of Inequality and Human Development, *Journal of the European Economic Association* 7 (2-3): 320–364.
- Dahlen, H. M. (2016): The impact of maternal depression on child academic and socioemotional outcomes, *Economics of Education Review* 52: 77-90.
- Darling, N. and L. Steinberg (1993): Parenting style as context: An integrative model. *Psychological Bulletin* 113 (3): 487-496.
- Del Boca, D., C. Flinn and M. Wiswall (2014): Household choice and child development. *The Review of Economic Studies* 81 (1): 137-185.
- Diener, E. and M.Y. Chan (2011): Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity, *Applied Psychology: Health and Well-Being* 3(1): 1-43.
- Döpfner, M., Plück, J., Kinnen, C., Arbeitsgruppe Deutsche Child Behavior Checklist (2014): *CBCL/6-18R, TRF/6-18R, YSR/11-18R. Deutsche Schulalter-Formen der Child Behavior Checklist von Thomas M. Achenbach. Elternfragebogen über das Verhalten von Kindern und Jugendlichen (CBCL/6-18R), Lehrerfragebogen über das Verhalten von Kindern und Jugendlichen (TRF/6-18R), Fragebogen für Jugendliche (YSR/11-18R)*. Göttingen: Hogrefe.
- Ferrer-i-Carbonell, A. and P. Frijters (2004): How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114(497): 641-659.
- Hainmueller, J. (2012): Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies, *Political Analysis* 20: 25-46.
- Havnes, T. and M. Mogstad (2011): “No Child Left Behind: Subsidized Child Care and Children’s Long-Run Outcomes”, *American Economic Journal: Economic Policy*, 3(2), 97-129.
- Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev and A. Yavitz (2010): "The rate of return to the HighScope Perry Preschool Program", *Journal of Public Economics*, 94: 128.
- Heckman JJ, Mosso S. (2014): The economics of human development and social mobility. *Annual Review of Economics* 6: 689–733.
- Heinrichs, N., K. Hahlweg, H. Bertram, A. Kuschel, S. Naumann and S. Harstick (2006): Die langfristige Wirksamkeit eines Elterntrainings zur universellen Prävention kindlicher Verhaltensstörungen, *Zeitschrift für Klinische Psychologie und Psychotherapie* 35 (2): 82-96.
- Heinrichs, N. and Jensen-Doss, A. (2010): The effects of incentives on families' long-term outcome in a parenting program. *Journal of Clinical Child & Adolescent Psychology*, 39 (5): 705-712.

- Helliwell, J.F., R. Layard and J.D. Sachs (2016): *World Happiness Report 2016, Update (Vol. I)*. New York: Sustainable Development Solutions Network.
- Henrich, G., and P. Herschbach (2000): Fragen zur Lebenszufriedenheit (FLZ-M). In U. Ravens-Sieberer, A. Cieza, M. Bullinger, N. von Steinbuechel & E. Poepfel (Hrsg.), *Lebensqualität und Gesundheitsökonomie in der Medizin. Konzepte, Methoden, Anwendung*. Landsberg: ecomed.
- Herbst, C. M., and E. Tekin, (2014): Child Care Subsidies, Maternal Health, and Child–Parent Interactions: Evidence from Three Nationally Representative Datasets, *Health Economics* 23 (8): 894-916.
- Klebanov, P. K., J. Brooks-Gunn, and M. C. McCormick (2001): Maternal coping strategies and emotional distress: Results of an early intervention program for low birth weight young children, *Developmental Psychology* 37 (5): 654.
- Kassenboehmer, S. C. and J. P. Haisken-De New (2009): Your're fired! The negative effect of entry unemployment on life satisfaction, *The Economic Journal* 119 (546): 448-462.
- Kottelenberg, M. J. and S. F. Lehrer (2013): New evidence on the impacts of access to and attending universal child-care in Canada, *Canadian Public Policy* 39 (2): 263-286.
- LaLonde, R. (1986): Evaluating the Econometric Evaluations of Training Programs, *American Economic Review* 76(4): 604-620.
- Lindsay, G., S. Strand and H. Davis (2011): A comparison of the effectiveness of three parenting programmes in improving parenting skills, parent mental-well being and children's behavior when implemented on a large scale in community settings in 18 English local authorities: the parenting early intervention pathfinder (PEIP), *BMC Public Health* 11: 962-975.
- Love, J. M., E. E. Kisker, C. Ross, H. Raikes, J. Constantine, K. Boller, J. Brooks-Gunn, R. Chazan-Cohen, L. B. Tarullo, C. Brady-Smith, A. S. Fuligni, P. Z. Schochet, D. Paulsell and C. Vogel (2005): The effectiveness of early head start for 3-year-old children and their parents: lessons for policy and programs, *Developmental Psychology* 41 (6): 885.
- Masse, L.N. and W. St. Barnett (2007): Comparative benefit-cost analysis of the Abecedarian program and its policy implications, *Economics of Education Review* 26: 113-125.
- Naumann, S., Bertram, H., Kuschel, A., Heinrichs, N., Hahlweg, K. & Döpfner, M. (2010): Der Erziehungsfragebogen (EFB). Ein Fragebogen zur Erfassung elterlicher Verhaltenstendenzen in schwierigen Erziehungssituationen. *Diagnostica*, 56: 144-157.
- Nowak, C. and N. Heinrichs (2008): A Comprehensive Meta-Analysis of Triple P-Positive Parenting Program Using Hierarchical Linear Modeling: Effectiveness and Moderating Variables, *Clinical Child and Family Psychology Review* 11: 114-144.
- OECD (2017): *How's Life? 2017: Measuring Well-being*. Paris: OECD Publishing.
- Olds, D.L. (2006): The Nurse-Family Partnership: An Evidence-Based Preventive Intervention, *Infant Mental Health Journal* 27 (1): 5-25.

- Oswald, A.J., E. Proto and D. Sgroi (2015): Happiness and Productivity, *Journal of Labor Economics*, 33 (4): 789-822.
- Peter, F., P.S. Schober and C. K. Spiess (2016): Early Birds in Day Care: The Social Gradient in Starting Day Care and Children's Non-Cognitive Skills, *CESifo Economic Studies* 62 (4): 725–751.
- Sanders, M.R. (1999): Triple P-Positive Parenting Program: Towards an Empirically Validated Multilevel Parenting and Family Support Strategy for the Prevention of Behavior and Emotional Problems in Children, *Clinical Child and Family Psychology Review* 2 (2): 71-90.
- Sanders, M. R. (2012): Development, Evaluation, and Multinational Dissemination of the Triple P-Positive Parenting Program. *Annual Review of Clinical Psychology* 8: 1-35.
- Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014): The Triple P-Positive Parenting Program: A systematic review and meta-analysis of a multi-level system of parenting support. *Clinical Psychology Review* 34 (4): 337-357.
- Sandner, M. (2015): Effects of Early Childhood Intervention on Fertility and Maternal Employment: Evidence from a Randomized Controlled Trial, *SOEPpapers* 799.
- Schmitz, Sophia (2017): The impact of publicly funded childcare on parental well-being -Evidence from cut-off rules. Mimeo.
- Schober, P. and C. Schmitt (2017): Day-care availability, maternal employment and satisfaction of parents: evidence from cultural and policy variations in Germany, *Journal of European Social Policy* 27 (5): 433-446.
- Schober, P. and J. Stahl (2016): Expansion of full-day childcare and subjective well-being of mothers: Interdependencies with culture and resources, *European Sociological Review* 32 (5): 593-606.
- Schroeder, C. and S. Yitzhaki (2017): Revisiting the evidence for cardinal treatment of ordinal variables, *European Economic Review* 92: 337-358.
- Statistisches Bundesamt (2002): *Statistisches Jahrbuch 2002 für die Bundesrepublik Deutschland*. Wiesbaden: Metzler-Poeschl.
- Wagner, G.G., J.R. Frick and J. Schupp (2007): The German Socio-Economic Panel Study (SOEP) – Scope, evolution, and enhancements. *Schmollers Jahrbuch* 127: 139-169.
- Webster-Stratton, C., M.J. Reid and M. Hammond (2004): Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training, *Journal of Clinical Child and Adolescent Psychology* 33 (1): 105-124.
- Yamauchi, C. (2010): The availability of child care centers, perceived search costs and parental life satisfaction. *Review of Economics of the Household* 8(2): 231-253.

Table 1: Descriptive statistics (pretreatment)

	ZF1
Child age (in years)	4.06 (1.02)
Child female (%)	48.57 (50.07)
Number of siblings	1.09 (0.94)
Mother single (%)	10.50 (30.73)
HH income above 1500€(%)	65.00 (47.78)
Mother's age (in years)	35.23 (4.95)
Mother German (%)	96.43 (18.59)
Mother low school degree (%)	11.07 (31.43)
Mother medium school degree (%)	34.64 (47.67)
Mother high school degree (%)	54.29 (49.91)
Mother no tertiary degree (%)	12.86 (33.53)
Mother vocational training (%)	53.57 (49.96)
Mother college degree (%)	33.57 (47.31)
Mother working (%)	76.79 (42.30)
Pre-treatment satisfaction with family life (5-point scale)	4.21 (0.81)
N	280

Standard deviations in parentheses
Source: Projekt Zukunft Familie 1.

Table 2: Observable characteristics for ZF1 treatment and control groups as assigned through randomization (columns one to three) and for ZF1 participants who were eventually treated or not treated (columns four to five)

	Control group (1)	Treatment group (2)	Difference (1) – (2) (3)	Not Treated (4)	Treated (5)	Difference (4) – (5) (6)
Child age (in years)	4.11 (0.98)	3.97 (0.97)	0.14	4.03 (1.04)	4.00 (0.93)	0.03
Child female (%)	43.55 (49.99)	47.13 (50.08)	-3.59	45.56 (50.08)	46.51 (50.07)	-0.96
Number of siblings	1.15 (0.96)	1.08 (0.79)	0.07	1.08 (0.91)	1.11 (0.78)	-0.03
Income above median (%)	74.19 (44.11)	79.62 (40.41)	-5.42	77.78 (41.81)	78.29 (41.38)	-0.52
Mother's age (in years)	36.00 (4.79)	35.47 (4.43)	0.53	35.41 (5.19)	35.77 (4.02)	-0.36
Mother German (%)	98.39 (12.70)	96.82 (17.62)	1.57	97.78 (14.82)	96.90 (17.40)	0.88
Mother low school degree (%)	9.68 (29.81)	6.37 (24.50)	3.31	7.78 (26.93)	6.98 (25.57)	0.80
Mother medium school degree (%)	27.42 (44.97)	36.94 (48.42)	-9.52	32.22 (46.99)	35.66 (48.09)	-3.44
Mother high school degree (%)	62.90 (48.70)	56.69 (49.71)	6.22	60.00 (49.26)	57.36 (49.65)	2.64
Mother no tertiary degree (%)	12.90 (33.80)	7.01 (25.61)	5.90	14.44 (35.35)	4.65 (21.14)	9.79**
Mother vocational training (%)	38.71 (49.11)	59.87 (49.17)	-21.16***	42.22 (49.67)	62.02 (48.72)	-19.79***
Mother college degree (%)	48.39 (50.38)	33.12 (47.22)	15.27**	43.33 (49.83)	33.33 (47.32)	10.00
Mother working (%)	80.65 (39.83)	76.43 (42.58)	4.21	75.56 (43.22)	79.07 (40.84)	-3.51
Pre-treatment satisfaction with family life (5-point scale)	4.28 (0.66)	4.25 (0.79)	0.03	4.30 (0.71)	4.23 (0.79)	0.08
N			219			219

Notes: The table can be read as follows: Column one shows means of characteristics for mothers who were assigned to the control group, column two those who were assigned to the treatment group. Column three shows differences between means. Column four shows means of characteristics of mothers who did not receive the treatment (regardless of the outcome of the randomization), column five those of mothers who were treated. Column six again shows differences in means between the last two groups. Calculations exclude single mothers.

Significance levels: ** $p \leq 0.05$, *** $p \leq 0.01$ of a t-test between groups.

Source: Projekt Zukunft Familie 1.

Table 3: Effects of Triple P on difference in maternal well-being from pre-treatment to later measurements of well-being (ZF1)

Difference in well-being between pre-treatment and measurement...	(1) OLS (without covariates)	(2) OLS (controlling for baseline well-being)	(3) OLS (including covariates)	(4) IV
... directly after the treatment	0.23* (0.14)	0.21 (0.13)	0.18 (0.14)	0.14 (0.18)
... after 1 year	0.12 (0.12)	0.10 (0.12)	0.06 (0.12)	0.05 (0.16)
... after 2 years	0.32** (0.14)	0.29** (0.12)	0.22* (0.12)	0.19 (0.16)
... after 3 years	0.33** (0.14)	0.31** (0.12)	0.30** (0.13)	0.28* (0.17)
Summation index	0.30** (0.13)	0.27** (0.11)	0.23** (0.11)	0.20 (0.15)
N	206	206	206	206

Notes: Each cell shows effect sizes from one model including covariates as described in the data section above. Robust standard errors in parentheses. *p≤0.1, ** p≤0.05

Source: Projekt Zukunft Familie 1.

Table 4: Changes in well-being on ordinal scale – Predicted probabilities for change from ordinal logit model

Difference in well-being between pre-treatment and measurement...	OLS		
	Change: -1	No change	Change: +1
... directly after the treatment	-0.07 (0.05)	-0.00 (0.01)	0.07 (0.05)
... after 1 year	-0.03 (0.04)	0.00 (0.00)	0.03 (0.04)
... after 2 years	-0.06 (0.04)	0.00 (0.01)	0.06 (0.04)
... after 3 years	-0.11** (0.05)	0.02 (0.01)	0.10** (0.04)
Summation index	-0.10** (0.04)	0.01 (0.01)	0.08** (0.04)
N	206	206	206

Notes: Each cell shows the probability for a discrete change from the pretreatment level of well-being. Groups are collapsed to three as described in section 6.2. Robust standard errors in parentheses. ** p≤0.05.

Source: Projekt Zukunft Familie 1.

Table 5: Effects of Triple P on parenting skills and child behavior

Difference in well-being between pre-treatment and measurement...	(1) Parenting skills	(2) Child behavior (CBCL)	(3) Internalizing behavior (CBCL)	(4) Externalizing behavior (CBCL)
... directly after the treatment	0.53*** (0.10)	0.19 (0.16)	0.26* (0.16)	0.18 (0.18)
... after 1 year	0.35*** (0.10)	0.32* (0.17)	0.44*** (0.17)	0.20 (0.18)
... after 2 years	0.51*** (0.11)	0.24 (0.17)	0.24 (0.17)	0.10 (0.17)
... after 3 years	0.36*** (0.11)	0.24 (0.17)	0.30* (0.16)	0.18 (0.18)
Summation index	0.67*** (0.16)	0.27* (0.16)	0.36** (0.16)	0.18 (0.18)
N	259	250	250	250

Notes: Each cell shows effect sizes from one IV-regression model including covariates on pre-treatment socio-economic characteristics as described in the data section; robust standard errors in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Source: Projekt Zukunft Familie 1.

Table 6: Descriptive statistics (pre-treatment)

	ZF2	SOEP
Child age (in years)	4.42 (1.03)	4.39 (1.06)
Child female (%)	44.16 (49.78)	49.01 (50.00)
Number of siblings	0.94 (0.90)	1.27 (0.98)
Mother single (%)	12.30 (32.97)	10.09 (30.13)
HH income above 1500€(%)	22.84 (42.09)	60.08 (48.98)
Mother's age (in years)	33.49 (5.54)	34.55 (5.59)
Mother German (%)	85.79 (35.01)	90.19 (29.75)
Mother low school degree (%)	31.98 (46.76)	17.85 (38.30)
Mother medium school degree (%)	45.69 (49.94)	45.81 (49.83)
Mother high school degree (%)	22.34 (41.76)	36.34 (48.11)
Mother no tertiary degree (%)	20.81 (40.70)	20.17 (40.13)
Mother vocational training (%)	68.53 (46.56)	56.21 (49.62)
Mother college degree (%)	10.66 (30.94)	23.62 (42.48)
Mother working (%)	65.99 (47.50)	63.45 (48.17)
Pre-treatment satisfaction with family life (5-point scale)	3.99 (0.94)	4.30 (0.74)
N	197	2375

Source: Projekt Zukunft Familie 2, SOEP v31, German Census and Statistics Braunschweig.

Table 7: Descriptive statistics of regional characteristics (one observation equals one regional unit)

	ZF1 (1)	ZF2 (2)	SOEP control group (3)
Share of foreigners (%)	7.30 (0.00)	11.98 (0.00)	7.30 (5.45)
Living space (m ² per person)	43.80 (0.00)	37.96 (0.00)	43.24 (3.89)
Population density (people per km ²)	1262.20 (0.00)	3464.61 (0.00)	937.70 (1052.10)
Mean age (in years)	43.00 (0.00)	44.43 (0.00)	43.27 (1.98)
Strain (young + old / working age population)	55.80 (0.00)	74.42 (0.00)	59.12 (6.10)
Share of Catholics (%)	14.00 (0.00)	13.66 (0.00)	29.75 (24.21)
Share of Protestants (%)	39.60 (0.00)	42.12 (0.00)	28.33 (16.86)
Share receiving welfare (%)	18.70 (0.00)	39.31 (0.00)	15.74 (8.68)
Migration background (%)	21.40 (0.00)	49.95 (0.00)	18.49 (9.88)
Share with high school degree (%)	53.70 (0.00)	n.a. n.a.	42.24 (12.44)
Unemployment rate (%)	6.10 (0.00)	9.41 (0.00)	4.87 (2.08)
Share of women in employment (%)	74.10 (0.00)	69.51 (0.00)	75.01 (3.63)
N	1	1	874

Notes: Numbers on *Kreis*- or *Gemeinde*-level (county or municipality) depending on characteristics.

Sources: Projekt Zukunft Familie 1 & 2, SOEP v31, German Census and Statistics Braunschweig.

Table 8: Effects of Triple P on maternal well-being using the SOEP based control group

Difference in well-being between pre-treatment and measurement...	ZF1: IV (using experimental control group)	ZF1: Entropy balancing with SOEP control group	ZF2: Entropy balancing with SOEP control group
... directly after the treatment	0.14 (0.18)	0.10 (0.14)	0.22** (0.11)
... after 1 year	0.05 (0.16)	0.10 (0.12)	0.06 (0.11)
... after 2 years	0.19 (0.16)	0.14 (0.11)	0.12 (0.13)
... after 3 years	0.28* (0.17)	0.20* (0.11)	0.21* (0.11)
Summation index	0.20 (0.15)	0.17* (0.10)	0.19* (0.10)
N	206	320	316

Notes: Each cell shows effect sizes from one IV-regression model including covariates on pre-treatment socio-economic characteristics as described in the data section; robust standard errors in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$.

Sources: Projekt Zukunft Familie 2, SOEP v31, German Census and Statistics Braunschweig.

Appendix A: Data alignment of well-being measures between different data sources

Maternal well-being measures. Both the ZF1- and ZF2-studies included a five-point scale on satisfaction with family life. The categories of the scale were *unsatisfied – rather unsatisfied – rather satisfied – fairly satisfied* to *very satisfied*. The SOEP uses an eleven-point scale is used (with only two labels at the extreme points: *completely unsatisfied* and *completely satisfied*). A transformation is needed in order to compare scores. Multiplying the SOEP scores by 0.4 and adding 1 results in a linear transformation. We assume that the scales represent the underlying latent variable accurately enough to use these models. Any measurement error in the scores from the ZF1- and ZF2-studies due to the reduced amount of information of the scale will be captured by the error term in the regression. Table A1 shows the relations between the well-being scales from the SOEP and the ZF1- and ZF2-studies as well as the transformed scale.

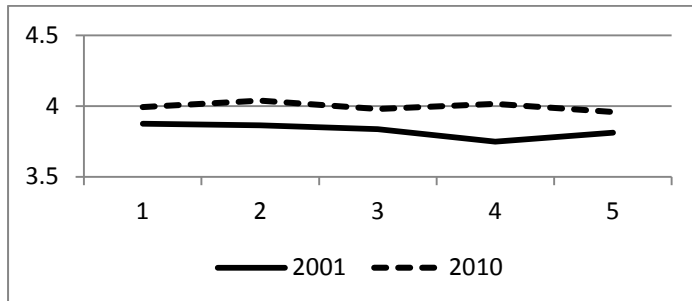
Table A1: Relations between the different maternal well-being scales

SOEP	Transformed scale	ZF1 and ZF2
0	1	1
1	1.4	
2	1.8	
-	-	2
3	2.2	
4	2.6	
5	3	3
6	3.4	
7	3.8	
-	-	4
8	4.2	
9	4.6	
10	5	5

Due to data availability, SOEP data are from different years than the data from ZF1 and ZF2. This could bias estimates if systematic differences affect study participants differentially between the two studies. Figure A1 shows trajectories of general life satisfaction (as a substitute for satisfaction with family life) for the two time periods using SOEP data. This cannot be shown with our satisfaction variable, which is not available for all years. The solid line shows the trajectory for the years 2001 to 2005. This corresponds to the years in which ZF1 and ZF2 were implemented. The dashed line shows the trajectory for the years 2010 to 2014, this corresponds to the years from which the generated SOEP control group was drawn. The lines run in parallel, indicating that well-being evolves similarly over time in both time periods. As

we include pre-treatment measures of the outcome variable in our models, we take the different levels into account.

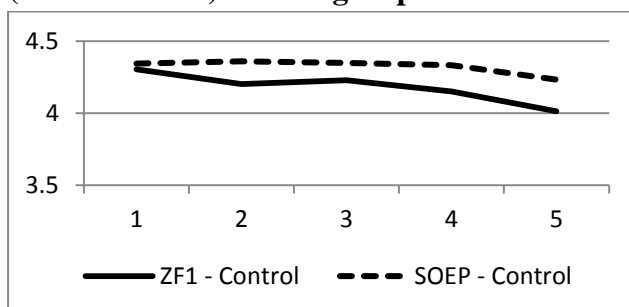
Figure A1: Five-year trajectories of satisfaction with life starting in 2001 and in 2010 (SOEP)



Source: SOEP v31.

Additionally, we check if trends for satisfaction with family life are similar in the ZF1 control group and the SOEP generated control group. Figure A2 shows that there is a decrease in satisfaction with family life for the ZF1 control group. This trend is slightly smaller for the SOEP control group. Apparently, satisfaction with family life decreased more over time for families in Braunschweig in 2001 than it did in Germany as a whole nine years later. Using the SOEP data without further adjustment likely leads to biased estimates of the effect of program participation. As our main results show, it is possible to recover effects of the program using the SOEP generated control group when using data from regions that are similar in terms of our distance measure to the ones where the original study was conducted. Thus, we can at least partly control for factors leading to the different trajectories in figure A2. While not completely establishing external validity of the study for the whole of Germany, our results indicate that Triple P can be shown to have a similar effects other populations in Germany when taking into account the variables we did.

Figure A2: Trajectories of satisfaction with family life ZF1 (2001 onwards) and SOEP (2010 onwards) control groups



Sources: Projekt Zukunft Familie 2, SOEP v31.

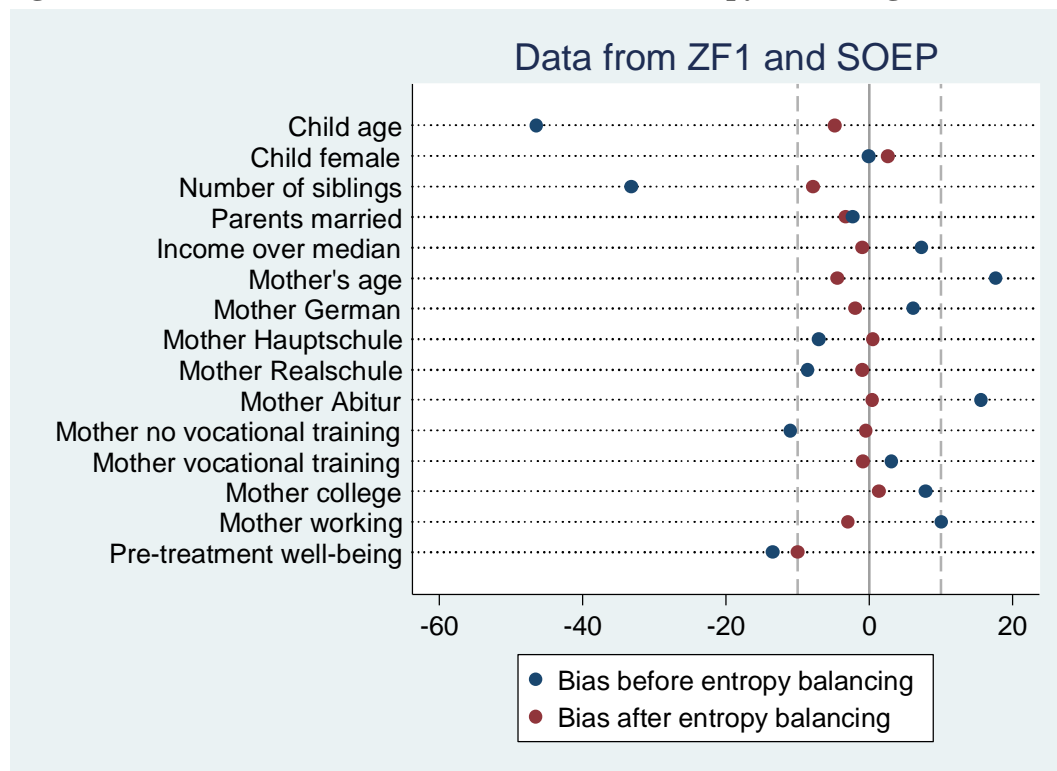
Appendix B: Entropy balancing and selection of observations based on regional data

Table B1: Using ZF1 and SOEP, including different percentiles of SOEP participants depending on distance measure

Difference in well-being between pre-treatment and measurement...	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
... directly after the treatment	0.03 (0.12)	0.10 (0.14)	0.07 (0.13)	0.04 (0.11)	0.01 (0.10)	0.08 (0.10)	0.07 (0.10)	0.06 (0.10)	0.08 (0.10)	0.08 (0.09)
... after 1 year	0.20 (0.17)	0.10 (0.12)	0.07 (0.11)	0.08 (0.11)	0.04 (0.10)	0.07 (0.10)	0.08 (0.10)	0.05 (0.10)	0.04 (0.10)	0.03 (0.09)
... after 2 years	0.11 (0.13)	0.14 (0.11)	0.10 (0.10)	0.07 (0.10)	0.06 (0.09)	0.04 (0.09)	0.05 (0.09)	0.07 (0.09)	0.07 (0.09)	0.08 (0.09)
... after 3 years	0.24* (0.13)	0.20* (0.11)	0.14 (0.10)	0.12 (0.10)	0.13 (0.10)	0.11 (0.10)	0.12 (0.09)	0.13 (0.09)	0.14 (0.09)	0.12 (0.09)
Summation index	0.18 (0.12)	0.17* (0.10)	0.12 (0.09)	0.10 (0.09)	0.07 (0.09)	0.09 (0.09)	0.10 (0.09)	0.10 (0.09)	0.10 (0.09)	0.09 (0.08)
N	222	320	419	517	614	714	810	908	1006	1103

Notes: Each cell shows effect sizes from one model including covariates as described in the data section. Robust standard errors in parentheses. * $p \leq 0.1$.
Sources: Projekt Zukunft Familie 1, SOEP v31, German Census and Statistics Braunschweig.

Figure B1: Covariate balances before and after entropy balancing – ZF1 (using 20% of SOEP observations)



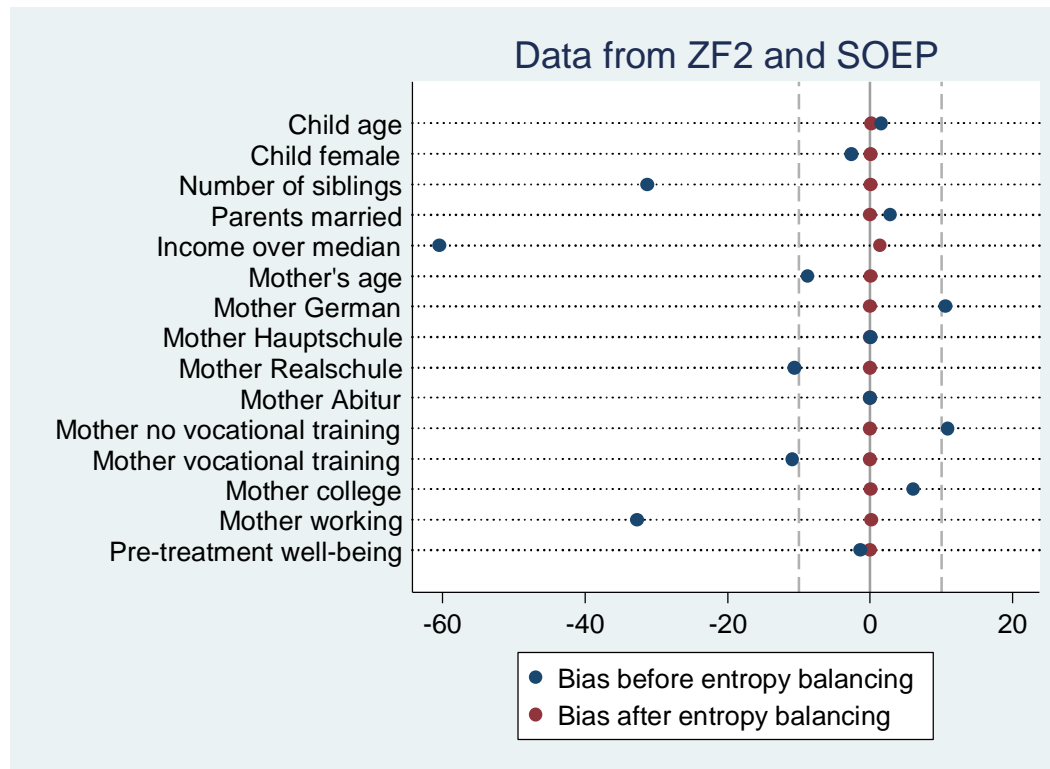
Sources: Projekt Zukunft Familie 1, SOEP v31, German Census and Statistics Braunschweig.

Table B2: Using ZF2 and SOEP, including different percentiles of SOEP participants, depending on distance measure

Difference in well-being between pre-treatment and measurement...	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
... directly after the treatment	0.12 (0.12)	0.22** (0.11)	0.20* (0.11)	0.20* (0.11)	0.22** (0.10)	0.20** (0.10)	0.19* (0.10)	0.20** (0.09)	0.21** (0.09)	0.19** (0.09)
... after 1 year	-0.02 (0.12)	0.06 (0.11)	0.04 (0.10)	0.01 (0.10)	0.03 (0.10)	0.01 (0.10)	0.03 (0.10)	0.06 (0.10)	0.06 (0.10)	0.04 (0.09)
... after 2 years	0.40** (0.17)	0.12 (0.13)	0.16 (0.12)	0.09 (0.11)	0.07 (0.11)	0.05 (0.11)	0.07 (0.11)	0.04 (0.11)	0.05 (0.10)	0.03 (0.10)
... after 3 years	0.11 (0.12)	0.21* (0.11)	0.21* (0.11)	0.17* (0.10)	0.13 (0.10)	0.12 (0.10)	0.14 (0.10)	0.12 (0.10)	0.11 (0.10)	0.10 (0.09)
Summation index	0.19 (0.12)	0.19* (0.10)	0.18* (0.10)	0.14 (0.09)	0.14 (0.09)	0.12 (0.09)	0.13 (0.09)	0.13 (0.09)	0.13 (0.09)	0.11 (0.09)
N	218	316	414	513	627	707	806	906	1002	1099

Notes: Each cell shows effect sizes from one model including covariates as described in the data section. Robust standard errors in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$
Sources: Projekt Zukunft Familie 2, SOEP v31, German Census and Statistics Braunschweig.

Figure B2: Covariate balances before and after entropy balancing – ZF2 (using 20% of SOEP observations)



Sources: Projekt Zukunft Familie 2, SOEP v31, German Census and Statistics Braunschweig.

Appendix C: Additional table

Table C1: Predicted probabilities for changes in well-being from ordered logit estimations

Difference in well-being between pre-treatment and measurement...	ZF2: Entropy balancing with SOEP control group		
	Change: -1	No change	Change: +1
... directly after the treatment	-0.07*** (0.03)	-0.04* (0.02)	0.11** (0.04)
... after 1 year	-0.02 (0.03)	-0.00 (0.01)	0.02 (0.03)
... after 2 years	-0.03 (0.03)	-0.01 (0.01)	0.04 (0.04)
... after 3 years	-0.08** (0.03)	-0.00 (0.01)	0.08** (0.03)
Summation index	-0.05 (0.03)	-0.00 (0.00)	0.05 (0.03)
N	316	316	316

Notes: Each cell shows the probability for a discrete change from the pretreatment level of well-being. Groups are collapsed to three, as described in section 6.2. Robust standard errors in parentheses. *p≤0.1, ** p≤0.05, *** p≤0.01

Sources: Projekt Zukunft Familie 2, SOEP v31, German Census and Statistics Braunschweig.