University of Hohenheim

Institute for Health Care & Public Management

Chair for Household and Consumer Economics

Essays in Health Economics

by

Micha Kaiser

2018

A dissertation submitted in partial fulfilment of the requirements for the degree "Doctor of Economics" (Dr. oec) to the Faculty of Business, Economics and Social Sciences, University of Hohenheim.

Date of oral examination: 8th of November 2017

Examination Committee

Chairman:	Prof. Andreas Pyka
Supervisor:	Prof. Alfonso Sousa-Poza
Co-Supervisor:	Prof. Christian Ernst
Dean of Faculty:	Prof. Karsten Hadwich

Acknowledgments

I would like to thank my supervisor Prof. Alfonso Sousa-Poza for the great opportunity to write this thesis. Throughout the whole process of studying, learning, writing, and - of course – researching, he always had a sympathetic ear to my questions, thoughts, and concerns. Besides his direct support he also gave me the opportunity to decide freely about my research topics. In particular, the possibility to visit different summer-schools whenever I wanted was a big support since this helped a lot to improve my scientific understanding, and secondly to grow on a personal level. Moreover, he gave me the opportunity to visit the Oxford Institute of Population Ageing, which was a great experience for me. I am convinced that my work on this thesis would not have been possible without the trustful and close cooperation between him and me. I also want to thank Prof. Christian Ernst for being my co-supervisor.

Additionally, I want to thank all of my additional co-authors Jörg Schiller, Christopher Schreckenberger and Jan Bauer. Besides that, I want to thank Jan especially for his support during my first year at the institute, and Christopher for answering all the small and tedious questions which arose during the working process. Also, I thank Patricia Höss and Steffen Otterbach for their support of my work at the institute.

Moreover, I want to thank all of the colleagues at the Institute for Health Care & Public Management of the University of Hohenheim. I also wish to thank the staff of the Oxford Institute of Population Ageing for giving me the ability to use their facilities (and of course their scientific input) during my stay in Oxford.

I want to express thanks to my whole family, especially to my father Johann-Georg Kaiser, who always supported me during my whole scientific education.

Writing a doctoral thesis takes a lot of time. Time, which is normally dedicated to someone's family. I therefore deeply appreciate all the patience and endurance of my wife Bianca during the whole work on this thesis. I thank her and my daughter Marlene so much for their daily support, their kindness and their love.

Finally, I thank Jesus the Lord for loving me and having given me the opportunity and ability to manage this project.

Contents

List of Abbreviations

BUC	Blow-up and cluster estimator
CRELES	Costa Rican Longevity and Healthy Aging Study
ELSA	English Longitudinal Study of Ageing
FMM	Finite mixed model
GSOEP	German Socio-Economic Panel
HRS	Health and Retirement Study
KiGGS	German Health Interview and Examination Survey of Children and Adolescents
Kita	Kindertagesstätte
LASI	Longitudinal Aging Study in India
MC	Monte Carlo
NIA	National Institute on Aging
NUTS	Nomenclature of territorial units for statistics
SCREEN	Skin Cancer Research to Provide Evidence for Effectiveness of Screening in Northern Germany
SCS	Skin cancer screening
SHARE	Survey of Health, Ageing, and Retirement in Europe
SQD	Strength and difficulty
TILDA	The Irish Longitudinal Study on Ageing

List of Tables

List of Figures

List of Appendix Tables

1 Introduction

"Be careful about reading health books. You may die of a misprint."

- Mark Twain

In economic theory a lot of attention is given to the understanding and modelling of consumption decisions of individuals. Usually, most models assume that individuals consume different markets goods and maximize their utility with respect to certain constraints. These constraints can be of various kinds. Besides monetary constraints health related constraints are vitally important during the maximization process of individuals. In such a paradigm, individuals would therefore benefit indirectly from being in a good health state, since this would imply that they are less constrained and could therefore shift their individual utility to a higher level. Moreover, health can also be treated as a good itself. Such an approach would assign a direct effect of different health states to an individual`s utility rather than incorporating health states by including them as a source for binding constraints. Apart from the different strategies in modelling the consumption decisions, both ways of thinking have in common that the achievement as well as the maintenance of a good health state is – to some extent - a necessary condition to foster the utility maximization process.

Additionally, health outcomes of individuals are highly sensitive to economic circumstances and different policy interventions. For instance, a change in the individual`s income will lead to an adjustment of the optimal consumption decision and therefore also to an adjustment of the health outcome (either in a direct or indirect way). Therefore a profound understanding of the impact of changes in economic and political processes helps to assess their effects on the health outcomes of individuals. Hence, this thesis investigates the impact of different economic factors and policy interventions on health. In particular, the thesis contributes to the literature in the following way:

Chapter two uses 22 years of data from the German Socio-Economic Panel and information on plant closures to investigate the effects of unemployment on four indicators of unhealthy lifestyles: diet, alcohol consumption, smoking, and (a lack of) physical activity. The main goal is to assess possible causal effects of unemployment on risky behaviors. In fact, in contrast to much of the existing literature the empirical identification strategy used in this analysis, is able to clearly identify exogenous effect and therefore avoids endogeneity, which may result from reversed causality.

Chapter three evaluates the relation between preschool care and the well-being of children and adolescents in Germany by using data from the German Health Interview and Examination Survey of Children and Adolescents. Analyzing this relationship is important to provide conclusive knowledge for parents as well as policy-makers due to several reasons. While parents are interested in providing the best health outcomes for their children, policy-makers need to balance a possible trade-off between economic as well as social costs and benefits related to preschool care. Additionally, the chapter examines differences in outcomes based on child socioeconomic background by focusing on the heterogeneous effects for migrant children.

The fourth chapter analyzes how a nationwide population-based skin cancer screening program implemented in Germany in 2008 has impacted the number of hospital discharges following malignant skin neoplasm diagnosis and the malignant melanoma mortality rate per 100,000 inhabitants. Therefore, panel data from the Eurostat database, which covers subregions in 22 European countries is analyzed for the years 2000-2013. By using fixed-effects methods, the causal relationship between the skin cancer screening program and the change in diagnosis and mortality rates are identified and a policy implication is derived.

Additionally, chapter five analyzes how closely different income measures conform to Benford's law, a mathematical predictor of probable first digit distribution across many sets of numbers. Because Benford's law can be used to test data set reliability, a Benford analysis is used to assess the quality of six widely used health related survey data sets. This is of particularly importance for health economists, since the majority of empirical work in this field relies on information from survey data. Additionally, a simulation technique is implemented to check for the robustness of the preceding analysis and a "best practice" rule for the evaluation of the quality of data sets is given.

The thesis concludes with a short summary in chapter six.

2 Does unemployment lead to a less healthy lifestyle?[1]

Abstract:

In this paper, we use 22 years of data from the German Socio-Economic Panel and information on plant closures to investigate the effects of unemployment on four indicators of unhealthy lifestyles: diet, alcohol consumption, smoking, and (a lack of) physical activity. In contrast to much of the existing literature, which unlike our analysis is unable to assess causality, our results provide little evidence that unemployment gives rise to unhealthy lifestyles.

[1] This chapter is based on joint work with Alfonso Sousa-Poza from the University of Hohenheim and Jan Michael Bauer from the Copenhagen Business School. The candidate´s individual contribution focused mainly on the set-up of the empirical estimation strategy, the literature research and the writing. This is an Accepted Manuscript of an article published by Taylor & Francis in *Applied Economics Letters, 24(12), p. 815-819* on *2017*, available online: *http://www.tandfonline.com/doi/abs/10.1080/13504851.2016.1231888?journalCode=rael20*. The author wants to thank the anonymous reviewer for his/her valuable comments.

2.1 Introduction

Poor lifestyle choices incur large social costs in terms of health care and individual well-being (Bouchery et al., 2011; Scarborough et al., 2011). Yet, although the effect of unemployment on unhealthy lifestyles is extensively discussed in the literature, drawing general conclusions remains difficult. Whereas most studies find that unemployment increases risky health behaviors (Ettner, 1997; Montgomery et al., 1998; Mossakowski, 2008; Dave and Kelly, 2012), others cannot confirm, or even contradict, these findings (Khan et al., 2002; Schmitz, 2011; Arcaya et al., 2014). In addition, earlier studies tend to suffer from the endogeneity inherent in standard regression models. To address this problem, a few studies estimate causal relations by exploiting natural experiments like plant closures (Debet al., 2011; Schmitz, 2011), in which job loss cannot be attributed to individual behavior and the shutdown is unlikely to be related to any one individual's lifestyle choices.

In this article, we use 22 years of data from the German Socio-Economic Panel (GSOEP) to estimate the effects of unemployment on four health behaviors: diet, alcohol consumption, smoking, and (lack of) physical activity. To highlight the importance of accounting for endogeneity, we use plant closure as an exogenous reason for unemployment. The analysis thus extends the literature by assessing the outcome of a range of lifestyle indicators and by using non-parametric estimation methods to account for unobserved heterogeneity and the ordinal nature of the data.

2.2 Methods and Data

The analysis is based on GSOEP data from 1991 to 2012 and restricted to individuals aged 25 to 60. Because we are interested in the effect of job loss, the sample contains only individuals who were employed at least once during the survey period. The basic model, which extends the approach taken by Schmitz (2011), is expressed by the following functional form:

$$Y_{it} = f(X_{it}^{pc}, X_{it}^{or}, X'^{AV}_{i(t-1)}, \mu_i, e_{it}) \tag{1}$$

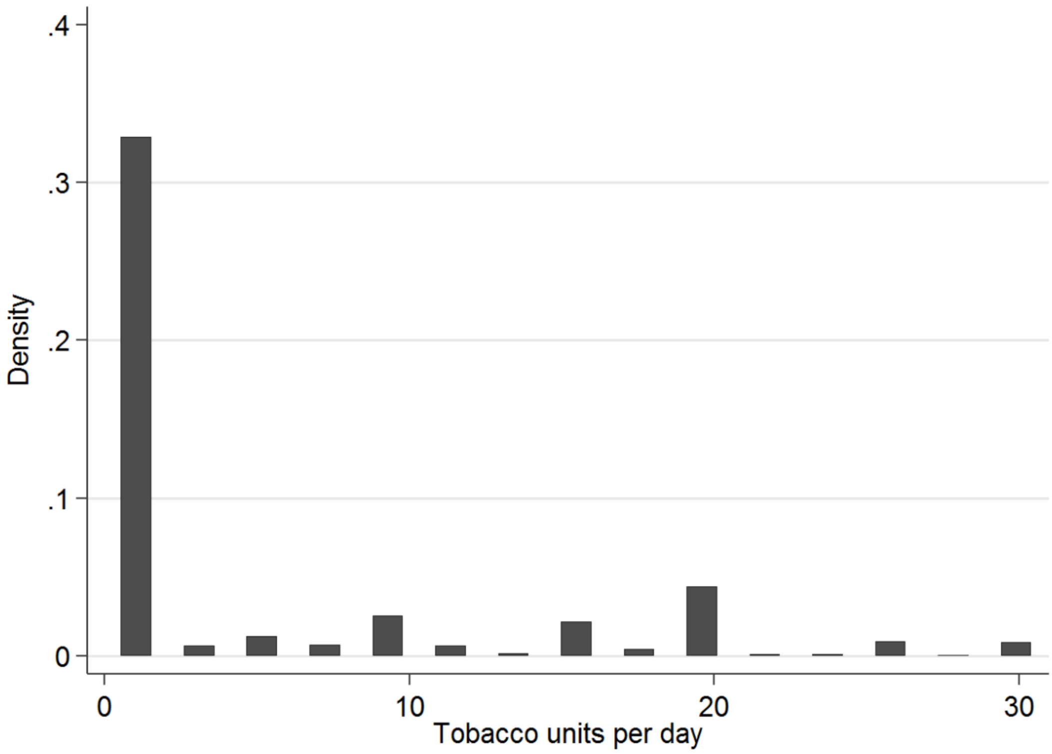
Here, the subscripts indicate individual i and time t, and the dependent variables are as described in Table 2.1. The main independent variables indicate job loss, with X_{it}^{pc} equalling 1 if a job loss occurred because of plant closure (exogenous layoff) and X_{it}^{or} equalling 1 if for some other reason, such as by mutual agreement (endogenous layoff). $X'^{AV}_{i(t-1)}$ captures lagged socio-demographic characteristics (i.e. age, number of children, household income, employment status, marital status, education, job type, and health insurance status) to control for pre-job loss differences without closing all possible channels for the effect on our outcome variable in period t.[2] The individual fixed effect μ_i (not used in the finite mixture model) captures unobserved characteristics, while e_{it} is the disturbance term. To estimate the ordinal measures, we apply a blow-up and cluster estimator (BUC), shown to be the most efficient for our research design (Baetschmann et al., 2015). Because smoking is measured in numbers of tobacco units per day and the data structure suggests more than one underlying density function (see Figure 2.1), we use a finite mixed model (FMM) approach for the estimation.

Table 2.1: DESCRIPTION OF DEPENDENT VARIABLES AND ESTIMATORS

	Variable			
	(1) Diet	(2) Alcohol	(3) Physical activity (sports)	(4) Smoking
Type of measure	4-point ordinal scale from very healthy (0) to not healthy (3)	4-point ordinal scale from never (0) to regular (3) for consumption of beer, wine, liquor, or mixed drinks	4-point ordinal scale from weekly (0) to never (3)	Metric: number of cigars, pipes, and cigarettes per day:
Survey year	Every second wave 2004–2012	Every second wave 2006–2010	Irregularly, with 14 years between 1992–2011	Every second wave 2002–2012
Estimator	BUC	BUC	BUC	FMM

[2] Thus, for example, a job loss can be assumed to have a direct (negative) income effect. If we controlled for income in period t rather than t-1, our income variable might be correlated with the job loss variable, making it impossible to assess the causal impact of (exogenous) unemployment on health behavior.

Figure 2.1: *Distribution of tobacco units per day*



2.3 Results

The results for the ordinally scaled variables (Table 2.2) indicate that, in general, the coefficients for the exogenous versus the endogenous layoffs differ substantially, although the fact that both regressors differ heterogeneously requires further explanation. For *diet* (i.e. the effect of unemployment on eating habits), we observe no significant effect when unemployment is treated endogenously. However, if individuals lose their jobs because of plant closure (exogenous layoff), their diets tend to become more health conscious, suggesting that unemployment may actually lead to better eating habits. Such an outcome may be attributable to the lower opportunity cost of the time needed to maintain a healthy diet. The lack of significance in the endogenous case suggests the existence of reverse causality, meaning that individuals with unhealthy eating habits (and their correlates such as obesity) are more likely to lose their jobs. A similar result is reported in Schmitz (2011), who shows that those in ill health select into unemployment.

For *alcohol use*, the endogenous regressor suggests a significant impact of unemployment on alcohol consumption. This relationship is, however, not evident if we look at individuals that lost their jobs due to plant closure. In contrast to other studies (Mossakowski, 2008; Ettner, 1997), we find no effect of unemployment on drinking behavior. For *physical activity*, measured here as engagement in sports, the endogenous and exogenous coefficients point in different directions. That is, unemployment seems to have a beneficial effect on physical activity by allowing individuals more free time to engage in it. On the other hand, the endogenous case suggests that less sporty individuals are more likely to be laid off.

Table 2.2: BUC ESTIMATES FOR DIET, ALCOHOL, AND PHYSICAL ACTIVITY

	Variables		
	(1) Diet	(2) Alcohol	(3) Physical
Unemployed because of plant closure (exogenous layoff)	-.2742*	-.2893	-.1730**
	(.1510)	(.2430)	(.0796)
Unemployed for other reasons (endogenous layoff)	.0005	-.3288***	.0997***
	(.0523)	(.0797)	(.0299)
N	29,913	12,146	168,005

Notes: Robust standard errors clustered on the individual level are in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01.

The results for the numerically scaled outcome variable *smoking* (Table 2.3) are estimated using two different finite mixture models, each made up of two components (see Figure 2.2). In the first, we estimate tobacco units per day assuming a negative binomial distribution (NEGBIN); in the second, we assume a normal distribution (NORMAL). This NORMAL model is restricted to individuals that smoked at least one tobacco unit per day during the sample period (i.e. were already smokers), meaning that marginal changes in tobacco consumption inferred from these estimates reflect the direct effect on smoking behavior. For the endogenous case, we obtain a significant coefficient in both models: NEGBIN yields a positive and significant relation for individuals who smoked very few tobacco units per day (component one), while NORMAL reveals a similar link for heavy smokers (component two).

Table 2.3: FMM AND OLS ESTIMATES FOR SMOKING

	Models		
	(1) FMM (NEGBIN)	(2) FMM (NORMAL)	(3) POOLED OLS
Component 1 (FMM)/Main (OLS)			
Unemployed because of plant closure (exogenous layoff)	-.3023	.0157	.6421
	(.4461)	(.5296)	(.4020)
Unemployed for other reasons (endogenous layoff)	.4082**	-.2208	.6036***
	(.1619)	(.2002)	(.1547)
Component mean	.8214	14.00	
Component 2 (FMM)			
Unemployed because of plant closure (exogenous layoff)	-.0221	-4.074	
	(.0329)	(2.874)	
Unemployed for other reasons (endogenous layoff)	.0011	3.251**	
	(.0151)	(1.484)	
Component mean	16.14779	26.66806	
N	55,038	18,721	55,038
adj. R^2			0.0830

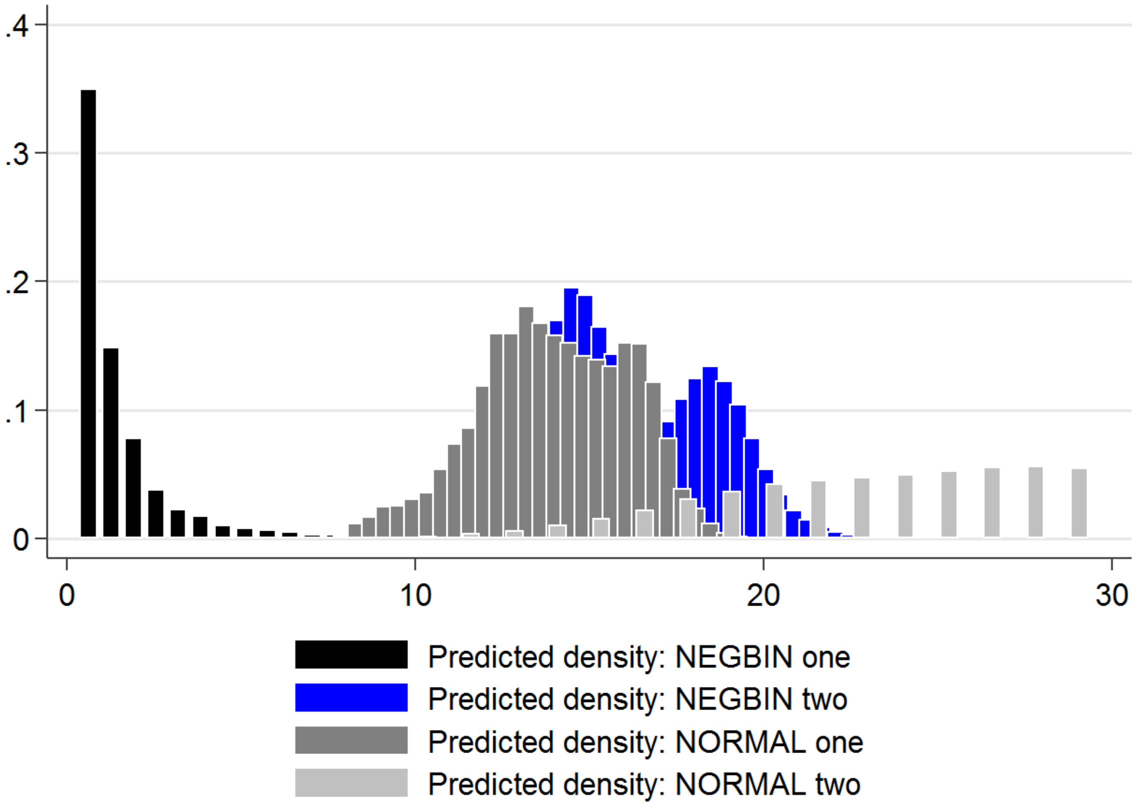
Notes: Column (1) reports the results for the two components estimated by an underlying negative binomial distribution. Column (2) shows the coefficients when tobacco consumption per day is normally distributed in both components. Because the NORMAL model is restricted to individuals who had already smoked in previous periods (i.e. excludes non-smokers), its results can be interpreted as the effect of unemployment on tobacco consumption. Column (3) reports the results for a pooled OLS regression, included as a robustness check. Robust standard errors clustered at the individual level are in parentheses. ** $p < 0.05$, *** $p < 0.01$.

Interestingly, neither model shows any effect for the group of regular smokers[3], who are covered by one component in each model. For the exogenous regressors,

[3] Regular smokers consumed around 15 tobacco units per day.

however, we obtain no significant coefficient in either model, thereby finding no evidence that unemployment has an impact on tobacco consumption. These findings are in line with the estimation results for the pooled OLS model, which is included as a robustness check. Again, the estimation results suggest reverse causality, i.e. individuals with a higher propensity to smoke are more likely to get laid off.

Figure 2.2: *Predicted densities for each FFM model and associated component*



However, the effect of unemployment on the lifestyle variables might be sensitive to the type of employment or personal conditions. We expect a 25 year-old single being laid off after a failed start-up to respond differently than a 45 year-old married mechanic. Furthermore, some unhealthy habits, like smoking, can be expensive and the reduction in income might limit the ability to change behavior. We test this by incorporating interaction terms for age, marriage, and household income with (exogenous/endogenous) layoff in each case. As job type is expected to show little within variation, particularly when interacted with layoff, we estimate two separate regressions for blue and white collar workers. We report the main findings in the Appendix, Table 8.1.

There is scant evidence for age-dependent heterogeneous effects related to diet and alcohol consumption. In the case of risky sport behavior, we find age to have a negative and statistically significant effect in the endogenous case. Including the age interaction for smoking yields the following pattern: in the endogenous case, we find an overall positive effect after the age of 29, while the overall effect for the exogenous case turns negative at the age of 48. This striking result suggests a very diverse response of smoking behavior to unemployed depending on age, which is partly offset in the average effect.

Marriage seems to mediate the effect of unemployment differently for sport and smoking. While only singles seem to improve their physical activity after an exogenous layoff, married people smoke significantly less compared to their single counterparts.

In contrast to the aggregated regression, we observe no significant effect of plant closure on diets for both *job type*s. Interestingly, the negative effect on risky alcohol, physical activity, and smoking behavior of exogenous layoffs is mainly driven by blue collar workers, while the white collar sample even shows more risky smoking behavior in response to exogenous unemployment.

For physical activity and risky alcohol consumption we find a positive interaction effect with *household income*. These findings support the idea that budget constrains might be a limiting factor for pursuing some types of risky behavior.

2.4 Discussion

In this article, we estimate the impact of unemployment on healthy lifestyles by using plant closures as an exogenous event that, in contrast to regular layoffs, is unlikely to be subject to a reverse causality bias.[4] Specifically, by comparing both exogenous and endogenous regressors, we are able to assess the magnitude of this bias. Our estimates, unlike those of earlier studies, reveal no negative impact of unemployment on any of the four lifestyle indicators assessed (diet, alcohol consumption, physical activity, and smoking). In fact, they imply a positive effect of unemployment on diet and physical activity, which, given the time intensity of ensuring a healthy diet and sufficient physical activity, is probably related to the lower opportunity costs of time during unemployment. This conclusion is also supported by research that shows a pro-cyclical behavior of unhealthy lifestyles (e.g. Freeman, 1999; Ettner, 1997). We additionally test for long-run effects of job loss by considering two periods of unemployment (results available on request). For smoking and sport, only endogenous layoffs have a persistent effect over two periods.

Overall, our findings stand in contrast to those of most previous studies (Montgomery et al., 1998; Dave and Kelly, 2012; Arcaya et al., 2014), which too often fail to account for the endogeneity bias. Virtually throughout all our estimations, the endogenous coefficients, unlike the exogenous coefficients (plant closure), are biased in a positive direction. We therefore conclude that individuals who make unhealthier lifestyle decisions are more likely to be laid off. This conclusion, together with our failure to observe any negative causal impact of unemployment, suggests that such policy interventions as the German "Equity in Health" cooperation network (*Kooperationsverbund Gesundheitliche Chancengleichheit*[5]), which is aimed specifically at influencing the behavior of unemployed individuals, may come too late. Policy-makers should thus shift their efforts to preventive measures that target unhealthy lifestyles in the whole population, which would not only improve health in general but might also mitigate the risk of individuals with unhealthy lifestyles becoming unemployed.

[4] To test for the possible anticipation of a plant closure, we match people experiencing a plant closure in t+1 with employees that remain employed in t+1 based on a set of observables other than lifestyle. We find no significant difference in the means of the four lifestyle variables in period t.

[5] See http://www.gesundheitliche-chancengleichheit.de

3 Preschool child care and child well-being in Germany: Does the migrant experience differ?[6]

Abstract

Because the value of preschool child care is under intensive debate among both policy-makers and society in general, this paper analyzes the relation between preschool care and the well-being of children and adolescents in Germany. It also examines differences in outcomes based on child socioeconomic background by focusing on the heterogeneous effects for migrant children. Our findings, based on data from the German Health Interview and Examination Survey of Children and Adolescents, suggest that children who have experienced child care have a slightly lower well-being overall. For migrant children, however, the outcomes indicate a positive relation.

[6] This chapter is based on joint work with Jan Michael Bauer from the Copenhagen Business School. The candidate´s individual contribution focused mainly on the literature research, the empirical work and the writing. The underlying manuscript is currently unpublished.

3.1 Introduction

In Germany, the use of preschool child care ranks high on the political agenda and is the subject of an ongoing public debate about its implications for child and family outcomes. Whereas advocates highlight the importance of sufficient public child care to promote female employment and provide equal educational opportunity across social strata, opponents consider the familial environment most beneficial for child development. Because certain political groups are currently promoting legislation to make participation in preschool child care mandatory rather than voluntary, a better understanding of its outcomes is essential.

Although the share of parents who use child care has risen during past decades, discussion of the short and long term effects of its use in the early years is ongoing in both academic and general discourse. Evaluating these effects is important because parents need to know the risks and benefits of early care in order to provide their children with the best opportunities possible, whereas policy-makers need to evaluate its economic and social costs in order to make subsidization decisions that benefit society.

The child care discussion is also related to the recent debate on migrant integration into Germany as increased migration and ethnic segregation raise questions about how to foster migrant children's chances for economic and personal success. Because these children tend to come from families with low socioeconomic status and limited German language skills, they are often disadvantaged. Hence, participation in preschool child care is often promoted as a tool to improve integration into and acculturation to both the public school environment and German society as a whole (Crosnoe, 2007; Dhuey, 2011; Spiess, Büchel, & Wagner, 2003). In this paper, therefore, we investigate the relation between experiencing preschool child care in Germany and well-being among children and adolescents, especially those from migrant families. The overall aim is to determine whether and to what extent children benefit from the early child care experience. We begin the discussion by outlining the institutional background in section 3.2, reviewing the pertinent literature in section 3.3, and describing our methodology and data in section

3.4. We report the results of our estimations in section 3.5 and conclude with a discussion of policy implications in section 3.6.

3.2 Institutional Background and Preschool Child Care in Germany

Preschool child care in Germany is organized into two phases based on child age. Between six months and three years, children can go to nursery school (Kindertagesstätte, henceforth Kita), after which they usually transfer to kindergarten before going to elementary school at around age six. Whereas regular schooling is compulsory, preschool child care is voluntary and allows families to choose a range of options from infrequent morning care to full day care during the work week. Although some families rely solely on home-based preschool education, most parents send their children at least to kindergarten (Statistisches Bundesamt, 2012).

Hence, in 1996, the government passed a law that grants lawful entitlement to kindergarten access for all children from age three until elementary school. However, the provision of child care has traditionally bene a local responsibility of the federal states (Evers, Lewis, & Riedel, 2005). Preschool child care is usually higher in Germany's eastern region because of its close relation with the history of female employment. Because kindergarten costs are regulated on the community level, they differ substantially, often based on number of children and family income. This redistributional approach is designed to promote the use of preschool child care by low-income families and those needing social assistance, who are entitled to additional public support from youth welfare offices to cover additional expenses like child subsistence costs.

In recent years, however, despite Germany's long kindergarten tradition, the core tasks of preschool child care have shifted away from social and pedagogical care toward early child education. Following the recommendations of the PISA studies and the rising demand for increased female labor force participation, the German government has intensified its efforts to improve and modernize the supply of preschool child care (Hemmerling, 2008). For example, a 2008 law focused on extending Kita placement promised a stepwise expansion of early child day care that would ensure universal coverage by 2013. This legislation, however, failed to meet expectations, and the coverage of available Kita places remains limited. Another new law engendered by regional political pressure provided alternative financial compensation for families that chose to raise their children at home in a more traditional family model rather than exercising their lawful right to send them to Kita. This financial compensation, however, was criticized as a misdirected incentive because not only did it not benefit families on social security (Arbeitslosengeld II) but experts worried it would strengthen traditional gender roles and reduce the use of preschool child care by low-income households (Spieß, 2012). Nevertheless, even though the regulation was found unconstitutional and repealed in mid-2015, preschool child care remains a highly debated topic in Germany, with advocates frequently stressing its importance for child development and the ability of early interaction with other children to improve social competencies later in life. The increasing share of migrant families in Germany, particularly, are encouraged to take advantage of publicly offered day care as a means to foster social integration and improve language abilities. The scientific evidence for preschool child care's ability to achieve these goals, however, remains ambiguous.

3.3 Related Literature

Although psychological evaluations of preschool care are numerous, most are U.S. studies on the relation with language and math skills and/or child behavior (i.e., problem externalization) whose findings are ambiguous. For instance, Burger's (2010) drew a generally positive conclusion about the link between early child care and cognitive development, pointing to an overall beneficial impact on children's start in life, with short term effects usually exceeding long term. Likewise, dependent on the quality of the preschool program, several studies provided evidence for a positive relation between experiencing early care and later cognitive development (Belsky, 2006; Cryan, Sheehan, Wiechel, & Bandy-Hedden, 1992; Votruba-Drzal, Li-Grining, & Maldonado-Carreo, 2008).

Other evidence, however, suggested that children who experience care often have more problems with social compliance (Belsky, 2006; NICHD, 2002; NICHD, 2004;), although these findings are highly sensitive to care duration and type, as well as to socioeconomic factors like family income and/or social background (NICHD, 2001). Belsky (2006), in fact, identified both risky and beneficial effects of early center-based care for U.S. children, with center quality positively related to child linguistic and cognitive skills, but the overall duration of care associated with a higher probability of social noncompliance and riskier behavior (Magnuson, Ruhm, & Waldfogel, 2007; NICHD, 2003). The age at which the child begins preschool care also seems to be important for various outcomes: Loeb, Bridges, Bassok, Fuller, & Rumberger (2007) associated a starting age between two and three years with the best academic outcomes but link a longer duration and higher intensity of care with a higher risk of social noncompliance. This negative effect was also identified by Magnuson et al. (2007), who showed that detrimental outcomes tend to persist longer and be more imperishable than any positive effect on math and language skills.

An analysis of U.S. academic data also provided strong evidence for the positive impact of full-day kindergarten on behavioral and schooling outcomes (Cryan et al., 1992), a finding in line with later verification of a positive but diminishing effect on reading and math skills up to the age of 12 (Votruba-Drzal et al., 2008). Moreover, although studies of early preschool child care's long term effects are few, they did provide some evidence of a positive association between early child care and, for instance, labor market participation (Havnes & Mogstad, 2011) or cognitive achievements in upper social strata (Peisner-Feinberg et al., 2001).

One important factor for the long term development of a child's cognitive ability as well as his language skills is the quality of the early child care environment, including the child-teacher relationship and preschool class size (Peisner-Feinberg et al., 2001). Nonetheless, although several studies emphasized the importance of program quality for later school performance (Belsky, 2006; Care & Development, 2002; Vandell, Belsky, Burchinal, Steinberg, & Vandergrift, 2010), other studies found no support for this link (e.g., Driessen, 2004; NICHD, 2001). Blau (1999), for instance, using National

Longitudinal Survey of Youth (NLSY) data to estimate the effects on child development of several child care quality measures (e.g., group size, staff-child ratio), concluded that despite some evidence for a relation between development and child subgroup, on average there is little or no evidence for a causal impact of child care quality. The persistence of early child care effects becomes even more ambiguous when the analysis considers demographic characteristics: once controls are included for a rich set of covariates (e.g., social class, environmental circumstances, occupational status, or migrant background), participation in preschool child care programs is not significantly associated with higher cognitive and non-cognitive competency outcomes (NICHD 2001).

Dustmann, Raute, & Schönberg (2016) analyzed the impact of preschool (kindergarten) on school readiness and health. They exploited a political reform that subsidizes kindergarten as an exogenous variation to estimate the heterogeneous effects of different subpopulations. They found kindergarten to work as an equalizer for children with differences in observed and unobserved characteristics. That is, their findings suggested that disadvantaged children are less likely to attain kindergarten, even though they tend to benefit the most. On the contrary, children who are most likely to attend kindergarten gain little from preschool children in terms of overall school readiness. Those findings contrast the idea that selection into child care is the main explanation for observed differences in post kindergarten outcomes. Using similar methodology but data from a different German federal state, Felfe & Lalive (2014) analyzed the effect of child care before the age of three on school readiness and related outcomes. In line with Dustmann et al. (2016) they found that gains from child care are strongest among low SES children and support the notion that the effect of child care is heterogeneous. Focusing on non-cognitive outcomes, Datta Gupta & Simonsen (2010) found no effects in terms of behavioral measures (measured with the strength and difficulty (SDQ) index) from attending preschool in Denmark. However, they found family day care, as an alternative, to negatively affect boys from low SES households. Additionally, children may benefit through positive effects channeled through their parents, as , for instance, the German expansion of early childcare lead to some increases in parental well-being (Schober & Schmitt, 2017).

The literature evaluating the impact of preschool child care on migrants, particularly, by focusing virtually exclusively on school performance, supports the idea that preschool child care is beneficial for migrant children (Crosnoe, 2007; Dhuey, 2011). For instance, Magnuson et al. (2006) found that the English proficiency of U.S. migrants improves through child care, thereby increasing their "school readiness." Likewise, Schlack, Hölling, & Kurth (2007), relying on the German Health Interview and Examination Survey of Children and Adolescents (KiGGS) data, showed that the preschool daycare participation rate of migrant children in Germany is significantly lower than that of nonmigrant children. They also demonstrated that the share of migrants whose children ever experience preschool child care is significantly lower between the ages of two and three and higher between the ages of five and six but that the risk of mental problems is twice as high for migrant children as for nonmigrant children. They identified no negative risk of attending preschool child care on mental problems for their full child sample. Positive effects related to the preschool child care experience of migrant children are also identified by Spiess et al. (2003), who showed that migrant children in Germany who experience such care are less likely to be enrolled in lower track secondary education. The authors were unable, however, to detect any positive and significant effect for native children in the same study.

Given the above findings, the frequent statistical indication of lower child care participation among migrant children is surprising. One assumption is that, as shown by Obeng (2007) for migrants from Africa, it may be linked to a parental desire to instill the native cultural identity. In fact, Turney and Kao (2009), in an analysis of pre-kindergarten child care effects on child behavior, documented clear effect differences based on country and ethnicity of origin. They identified no effect, however, on children's feelings of sadness and loneliness, indicators of emotional well-being, a subject that has, to the best of our knowledge, yet to be analyzed in depth in the context of preschool child care outcomes.

Overall, however, empirical evidence on the relationship between early child care and overall well-being of children and adolescence is limited, which motivates our present attempt to glean new insights into the relation between preschool child care and psychometric measures for schoolchildren. In particular, we analyze the heterogeneous relationship for native German and migrant children, which is important when preschool child care is considered as a public instrument for the integration of migrant children.

3.4 Data and Methodology

Our empirical analysis is based on data from the first wave of KiGGS, collected between 2009–2012 by the Robert Koch Institute (http://www.kiggs-studie.de/english/home.html). Designed primarily to gather information on the health status of Germany’s youth, this survey offers 17,000 observations of 0- to 17-year-olds obtained through differently administered questionnaires (e.g., filled out by parents, physicians, or the children themselves; Kurth et al., 2008). To permit a more detailed subsample analysis, however, it also oversamples East German and migrant children, an unequal selection probability that we adjust for by using survey weights throughout the analysis.

Our main outcome of interest is preschool child care’s effect on child well-being, which we approximate by the KiGGS’ quality of life sum score (y in equation 1) derived from 24 Likert-scale items in six different dimensional scales (emotional well-being, physical well-being, self-esteem, family, friends, and school) in the parental version of KINDL (Bullinger, Brütt, Erhart, & Ravens-Sieberer, 2008). Those items are combined and transformed to a one-item sum score range ranging from 0 to 100. The reliability and validity of this score, one of the few German-language measures of child quality of life, has been verified using several tests (Ravens-Sieberer & Bullinger, 2000). In the KiGGS data set, parental information on the sum score is available for ages 3 to 17, which reduces the sample size to fewer than 15,000 observations, with self-assessed values collected only from children aged 10–17. We rely mostly on the values from this parental evaluation because of its larger sample size and demonstrated reliability (Erhart, Ellert, Kurth, & Ravens-Sieberer, 2009). Nevertheless, we later split the sample by different age groups and then take a detailed look at individual subscales of the sum score to identify the impact on different life domains.

As an explanatory variable, we focus primarily on the experience of child care. For the main analysis, we use a *child care* dummy equal to 1 ($D = 1$) if a child has experienced any type of preschool child care (e.g., Kita and/or kindergarten) and 0 ($D = 0$) if the child has been raised exclusively in the family household. To analyze the different effects for migrant children, we adopt the KiGGS definition of migrant (M) as either (1) a child born in a foreign country with at least one non-German parent ($M = 1$) or (2) a child with two non-German parents. Children born in Germany with only one non-German parent are not considered as migrants ($M = 0$).

To identify the relation between preschool child care and child well-being, we rely on the following population model estimated by ordinary least squares (OLS):

$$y = x\beta + \gamma D + \delta M + \rho(D \times M) + \varepsilon \quad (1)$$

To allow for different effects of preschool child care on migrant children, we include an interaction term $(D \times M)$ in some of the regressions. We also address the question of preschool child care starting age by differentiating children who attended Kita from those who began on the kindergarten level (we attribute a starting age under 3 to Kita and one between 3 and 6 to kindergarten). In line with survey administrator suggestions, we cluster the standard errors (ε) on the sample point level. To estimate a causal relationship of child care on well-being one needs to account for all characteristics that might correlate with the uptake of child care and also well-being. Selection into child care might differ between different socio-economic status and values. For instance, despite controlling for the need of child care, lower earnings still have a negative impact on uptake in Bulgaria (Meurs, 2006). To account for selection into child care, we rely on a rich set of child, parental, and household covariates that capture differences in socio-economic status and residency, which are captured by the $1 \times K$ vector x in the population model (see Table 3.1), with K equal to the number of covariates included in a particular model plus a constant. We cannot rule out the possibility that selection is based on unobserved characteristics, as our cross-section data do not provide a clearly exogenous variation of child care uptake. However, recent work studying this phenomenon show that such selection into child care does not constitute the main determinant for different outcomes between children with and without preschool child care experience (Dustmann et al., 2016; Felfe & Lalive, 2014).

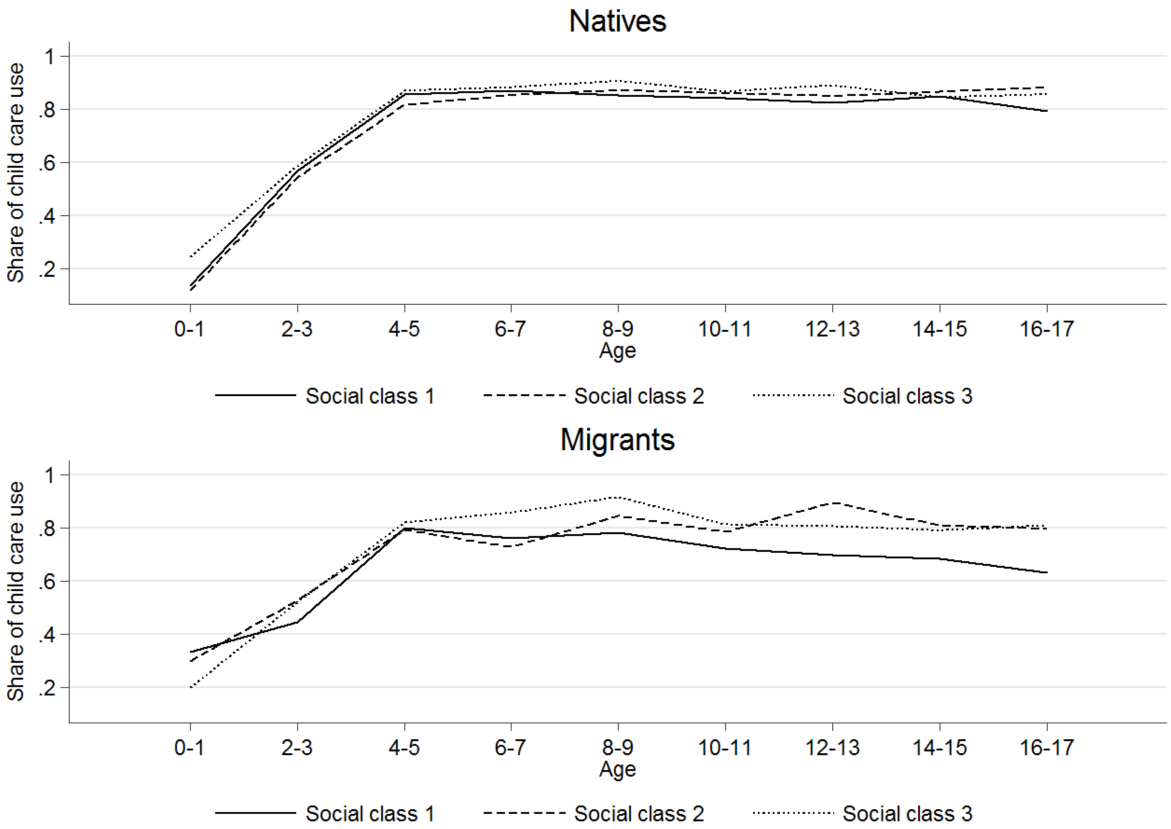
The separation between children with and without experience of child care (Table 3.1) indicates that the former tend to come from families with a higher social status and higher employment levels. The statistics also show that the share of migrant children with child care experience is lower than the share of native children. In Figure 3.1, which separates the share of children formerly or currently in preschool care by migration and social status, all groups show an increasing rate of experience up to the age of kindergarten entry (the socioeconomic categories are based on the Winkler index (Winkler & Stolzenberg, 1999), which divides society into “classes” based on parental education, occupation, employment, and income). Among German natives, however, the share does not differ by social strata and remains fairly constant for older cohorts. Among migrants, we observe two notable differences: (1) the overall share is lower than for German children and (2) preschool child care experience is lower for older cohorts in the lowest socioeconomic strata. These observations stem from a past, albeit declining, selections of low status migrants out of preschool child care. The descriptive statistics in Table 3.1 show not only that migrant children tend to come from lower socioeconomic backgrounds, but that they tend to live more frequently in large cities (city size is measured as categorical variable according to the number of inhabitants: rural area less than 5,000; small city 5,000 to 20,000; medium sized city 20,000 to 100,000; large city more than 100,000). On average, migrants also tend to have more behavioral problems (SDQ sum score) and lower well-being, signaled by differences in group means derived through multivariate regression analysis.

Table 3.1: DESCRIPTIVE STATISTICS

	Full	Child care	Home care	Nonmigrant	Migrant
Child age (years)	10.203	10.249	9.954**	10.195	10.258
Child male	0.513	0.516	0.494	0.513	0.515
Parents single	0.043	0.044	0.040	0.043	0.042
More than 4 Persons in HH	0.302	0.294	0.342***	0.287	0.408***
Sibling in HH	0.823	0.815	0.862***	0.817	0.861***
Social category	12.213	12.338	11.540***	12.589	9.607***
Father					
Vocational training	0.269	0.267	0.282	0.264	0.305***
University degree	0.255	0.261	0.224***	0.269	0.162***
Part-time job	0.027	0.028	0.020**	0.025	0.038**
Full-time job	0.895	0.896	0.888	0.910	0.788***
Self-employed	0.141	0.142	0.137	0.148	0.093***
Age (group)	5.003	5.000	5.017	5.050	4.677***
Mother					
Vocational training	0.278	0.281	0.263	0.281	0.263
University degree	0.164	0.175	0.107***	0.165	0.159
Part-time job	0.497	0.499	0.485	0.520	0.332***
Full-time job	0.181	0.195	0.100***	0.176	0.211**
Self-employed	0.064	0.068	0.043***	0.067	0.044***
Age (group)	4.440	4.444	4.420	4.502	4.010***
East Germany	0.161	0.181	0.057***	0.175	0.068***
Rural area	0.195	0.193	0.200	0.215	0.055***
Small city	0.286	0.285	0.294	0.298	0.208***
Medium sized city	0.294	0.291	0.313	0.288	0.342***
Large city	0.225	0.230	0.193***	0.200	0.395***
Net HH income (grouped)	8.984	9.031	8.731***	9.197	7.507***
SDQ sum score	7.837	7.889	7.553**	7.666	9.021
KINDL sum score	77.271	77.176	77.785**	77.335	76.830
Migrant	0.126	0.118	0.170**	0.000	1.000
Child care	0.844	1.000	0.000	0.851	0.790***
Observations	10,814	9,358	1,456	9,650	1,164

Notes: Parental age grouped in 5 year brackets starting below 25 to above 55. Household income groups are measured as increasing brackets ranging from 250 to 1000 €. Differences in means between the two subgroups are indicated by * p <0 .1, ** p < 0.05, *** p <0.01.

Figure 3.1: *Preschool child care participation of natives and migrants by social status*



3.5 Results

The relation between child care and well-being is outlined in Table 3.2, whose first column reports the outcomes for the model without interaction effects. These results, although they do not attain statistical significance, suggest an overall negative impact of preschool care on child well-being that is especially high for migrant children. Over all the models, the KINDL sum score diminishes as age rises and is substantially lower for children in nontraditional families, with a notable reduction in well-being when the household includes a new partner. On average, children in East Germany and those from richer households show higher levels of well-being.

Column 2 then introduces the interaction term between migrants and preschool child care, which once its different effects on migrants are considered, changes the coefficients and yields a significantly lower level of well-being for migrant children. More specifically, the size of the coefficient is now comparable with that for children from single female parent (vs. two parent) homes. The overall child care coefficient also becomes negative, suggesting that children who experience preschool child care have lower levels of well-being. For migrant children, on the other hand, the opposite seems true: they appear to benefit from this experience. In fact, the highly significant interaction term indicates that migrant children who experience preschool care score nearly 2 sum points higher than their counterparts who remain at home.

When we further divide child care based on age of entry (column 3), the results suggest that in general, children who attend Kita (i.e., experience preschool child care before the age of 3) have lower levels of well-being. The subsequent introduction of the interaction term (column 4) yields results similar to those from the previous model: the experiences of nonmigrants and migrants differ substantially, but the overall effect is significantly negative, with an early start in preschool child care seeming to produce an especially strong reduction in well-being. For migrant children, however, the positive interaction term suggests an overall beneficial relation, with higher levels of well-being among children who attend kindergarten only.

Given the substantial experiential difference between migrants and nonmigrants, in columns 5 and 6, we further investigate the intensive margin. In Sample II, we attempt to determine whether the year that child care is begun affects well-being by excluding all of the children with no experience of preschool child care. In line with previous results, a higher starting age appears to be associated with higher levels of well-being, implying that receiving preschool child care too early has detrimental effects that do not differ significantly between nonmigrant and migrant children. To identify the different effects on elementary versus secondary school children, Table 3.3 divides the sample by age group (since our age data use 2-year increments, we include 10-year-olds in the secondary school sample). Because the data are cross-sectional, however, we cannot rule out a possible bias through cohort effects, which cannot be distinguished from child age. For children aged 6 to 9, the results in column 1 reveal no differences in well-being based on either child care or being a migrant, but, as in the full sample, those in column 2 indicate

Table 3.2: OLS ESTIMATES FOR THE KINDL CHILD WELL-BEING MEASURE FOR CHILDREN 3–17

	OLS estimates for child well-being measured by the KINDL sum score					
	Sample I				Sample II	
	(1)	(2)	(3)	(4)	(5)	(6)
	Child stayed at home as reference					
Child care	-0.5198	-0.9855***				
	(0.338)	(0.368)				
Kindergarten			-0.4296	-0.8760**		
			(0.344)	(0.375)		
Kita			-0.8723**	-1.2689***		
			(0.398)	(0.425)		
Starting age of care					0.2361*	0.2042
					(0.125)	(0.130)
Child care*Migrant		2.8998***				
		(0.806)				
Kindergarten*Migrant				2.8614***		
				(0.834)		
Kita*Migrant				2.5951**		
				(1.093)		
Starting age*Migrant						0.2729
						(0.380)
Migrant	0.5592	-1.7922**	0.6389*	-1.5954**	1.3471***	0.5022
	(0.364)	(0.760)	(0.369)	(0.770)	(0.398)	(1.206)
Age	-1.3119***	-1.3101***	-1.3133***	-1.3104***	-1.3535***	-1.3549***
	(0.058)	(0.058)	(0.058)	(0.058)	(0.070)	(0.071)
Male	-0.3056	-0.2926	-0.3229	-0.3119	-0.3676	-0.3688
	(0.201)	(0.201)	(0.203)	(0.203)	(0.226)	(0.226)
Sibling in HH	-0.0859	-0.0912	-0.1920	-0.2005	0.8027	0.8027
	(0.735)	(0.744)	(0.761)	(0.771)	(0.719)	(0.719)
Net income HH	0.2791***	0.2760***	0.2880***	0.2836***	0.2905***	0.2899***
	(0.054)	(0.054)	(0.055)	(0.055)	(0.062)	(0.062)
Parental situation	*Parents married and living together as a reference*					
Mother with partner	-1.7939***	-1.8062***	-1.7512***	-1.7708***	-1.7074***	-1.7106***
	(0.520)	(0.521)	(0.518)	(0.519)	(0.541)	(0.541)
Father with partner	-5.2867***	-5.2579***	-5.4407***	-5.4115***	-5.4071***	-5.4002***
	(1.796)	(1.794)	(1.807)	(1.806)	(1.966)	(1.966)
Single mother	-1.3797*	-1.4382*	-1.4544*	-1.5144*	-2.3650**	-2.3613**
	(0.827)	(0.828)	(0.854)	(0.856)	(0.936)	(0.936)
Single father	-0.2611	-0.3108	-0.0664	-0.1087	-0.7054	-0.6974
	(1.381)	(1.390)	(1.419)	(1.427)	(1.682)	(1.682)
Other	-3.8431***	-3.8903***	-4.2007***	-4.2504***	-3.9998***	-4.0173***
	(1.251)	(1.255)	(1.276)	(1.282)	(1.356)	(1.357)
Type of region	*Rural area as a reference*					
Small city	-0.4288	-0.4214	-0.3771	-0.3704	-0.2096	-0.2115
	(0.322)	(0.319)	(0.320)	(0.317)	(0.393)	(0.394)
Medium city	-0.0762	-0.0542	-0.0722	-0.0538	-0.0657	-0.0718
	(0.342)	(0.339)	(0.339)	(0.335)	(0.399)	(0.400)
Large city	-0.1450	-0.1526	-0.1221	-0.1315	-0.2564	-0.2638
	(0.355)	(0.351)	(0.363)	(0.359)	(0.408)	(0.408)
East Germany	1.5645***	1.6266***	1.8035***	1.8419***	1.8930***	1.8674***
	(0.300)	(0.300)	(0.327)	(0.328)	(0.350)	(0.353)
Constant	81.5713***	82.1541***	81.4647***	82.0426***	80.5753***	80.6338***
	(1.271)	(1.299)	(1.343)	(1.375)	(1.565)	(1.567)
N	10835	10835	10536	10536	9088	9088
Adj. R^2	0.083	0.085	0.084	0.085	0.086	0.086

Notes: All reported estimates are weighted nonstandardized regression coefficients. Sample I includes all children: Sample II includes only children reported to have experienced some preschool child care. All models include controls for number of individuals in the household, parental age, parental education, parental employment, and parental occupation. Robust standard errors clustered on the sampling point level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

that well-being among migrant children in elementary school (Sample I) varies greatly depending on early child care experience. Whereas the well-being of migrant children is nearly 5 sum score points lower overall than that of nonmigrant children, this difference is nullified when they have experienced preschool child care. Breaking the samples down by additional variables in columns 3 and 4 suggests that experiencing kindergarten only is slightly more valuable for well-being than experiencing early child care. For the older children, we find less heterogeneity between nonmigrant and migrant children but still observe a significantly negative coefficient for child care overall; particularly, for kindergarten only.

In Table 3.4, we attempt to identify which well-being domains are most affected by the child care experience based on the six different subscales that make up the KINDL sum score. We again control for possible heterogeneity among age groups by splitting the sample into elementary and secondary school children. For physical well-being (column 1), the coefficients are only significant for elementary school children (panel B). In this domain, migrant children score over 5 points less than nonmigrants, although this difference is more than offset for those who have experienced child care. Based on the KINDL sum score, psychological well-being is lower overall for the full sample of migrant and nonmigrant (panel A), but a more pronounced difference emerges for migrants in the sample of elementary school children (panel B). Self-esteem, on the other hand, shows a long term effect in migrant children with experience of care, with a strong positive interaction term among secondary school children (panel C). For the family domain, child care experience seems slightly detrimental for nonmigrant children but differs between panels for migrant children. That is, whereas the full sample yields positive point estimates for the migrant dummy and interaction term, the elementary school age subsample has a 4.3 point higher sum score related to child care, and the interaction term is insignificant for secondary school children, among whom migrants generally score 3.5 points higher independent of preschool experience. In the friends domain, the coefficients again show the general pattern of child care experience making a strong difference, especially among younger migrants.

Table 3.3: CHILD WELL-BEING SPLIT BY AGE GROUP

	OLS estimates for child well-being measured by the KINDL sum score											
	Children aged 6 to 9 (elementary school)						Children aged 10 to 17 (secondary school)					
	Sample I				Sample II		Sample I				Sample II	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Child stayed at home as reference						*Child stayed at home as reference*					
Child care	-0.0845	-0.8359					-0.7251	-1.0096**				
	(0.497)	(0.544)					(0.486)	(0.487)				
Kindergarten			0.0653	-0.6941					-0.8395*	-1.1237**		
			(0.504)	(0.550)					(0.500)	(0.503)		
Kita			-0.4220	-1.0649					-0.0725	-0.3014		
			(0.633)	(0.648)					(0.567)	(0.582)		
Starting age of care					0.1776	0.1654					-0.0719	-0.1143
					(0.210)	(0.207)					(0.164)	(0.175)
Child care*Migrant		5.4195***						1.6836				
		(1.662)						(1.220)				
Kindergarten*Migrant				5.5800***						1.7416		
				(1.704)						(1.303)		
Kita*Migrant				4.8656**						1.3897		
				(2.318)						(1.607)		
Starting age*Migrant						0.1327						0.3671
						(0.624)						(0.546)
Migrant	-0.2154	-4.8099***	-0.1294	-4.6907**	0.8122	0.4120	0.4996	-0.8665	0.7792	-0.5524	1.2208**	0.0263
	(0.682)	(1.837)	(0.719)	(1.878)	(0.624)	(2.006)	(0.479)	(1.135)	(0.486)	(1.147)	(0.534)	(1.743)
N	3084	3084	3009	3009	2655	2655	5583	5583	5422	5422	4679	4679
Adj. R^2	0.021	0.026	0.021	0.026	0.024	0.024	0.032	0.032	0.033	0.034	0.030	0.030

Notes: All reported estimates are weighted nonstandardized regression coefficients. Sample I includes all children; Sample II includes only children reported to have experienced some preschool child care. All models include controls for child age, gender, a dummy for having at least one sibling, number of individuals in the household, household net income, parental situation (married, single, living with new partners),parental age, parental education, parental employment, parental occupation, and type of region (rural area, small/medium/large city), East Germany. Robust standard errors clustered on the sampling point level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.4: SEPARATE SUBSCALES FOR THE KINDL SUM SCORE

	OLS estimates for the child well-being subscales of the KINDL sum score					
	Physical well-being	Psychological well-being	Self-esteem	Family	Friends	School
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
Child care	-0.3352	-1.3324***	-0.9580*	-0.9832**	-1.3587***	-0.7504
	(0.580)	(0.457)	(0.525)	(0.468)	(0.405)	(0.565)
Child care*Migrant	1.9367	3.4113***	3.9725***	2.3283*	4.1206***	1.5678
	(1.313)	(1.156)	(1.381)	(1.281)	(1.016)	(1.400)
Migrant	-1.5275	-2.6815***	-3.1415**	1.6208	-2.3533***	-5.1189***
	(1.197)	(1.028)	(1.411)	(1.167)	(0.892)	(1.282)
N	10745	10795	10796	10847	10833	10125
Adj. R^2	0.044	0.030	0.039	0.034	0.020	0.175
Panel B: Children aged 6 to 9 (elementary school)						
Child care	-0.5597	-0.5650	-1.5455*	-0.7003	-1.7262**	-0.6551
	(0.926)	(0.694)	(0.914)	(0.813)	(0.775)	(0.866)
Child care*Migrant	7.6915***	5.4490**	3.0743	4.3561**	8.4081***	2.7381
	(2.559)	(2.561)	(2.813)	(2.036)	(2.202)	(2.969)
Migrant	-5.1396**	-5.4637**	-3.3488	-0.7443	-7.2887***	-6.8709**
	(2.412)	(2.689)	(3.020)	(2.263)	(2.189)	(2.945)
N	3058	3077	3079	3090	3089	2712
Adj. R^2	0.010	0.013	0.019	0.028	0.024	0.070
Panel C: Children aged 10 to 17 (secondary school)						
Child care	0.0473	-1.8325***	-0.9440	-0.7485	-1.4952**	-1.0125
	(0.789)	(0.634)	(0.683)	(0.669)	(0.595)	(0.695)
Child care*Migrant	-1.1754	2.8044*	4.3556**	0.5868	3.5110**	1.5607
	(1.866)	(1.626)	(1.962)	(1.947)	(1.522)	(1.973)
Migrant	0.9371	-1.8416	-2.8318	3.4605*	-1.0807	-6.2270***
	(1.742)	(1.413)	(1.839)	(1.825)	(1.403)	(1.813)
N	5535	5558	5559	5583	5577	5461
Adj. R2	0.040	0.020	0.015	0.021	0.020	0.093

Notes: All reported estimates are weighted nonstandardized regression coefficients. All models include controls for child age, gender, a dummy for having at least one sibling, number of individuals in the household, household net income, parental situation (married, single, living with new partners), parental age, parental education, parental employment, parental occupation, type of region (rural area, small/medium/large city), and East Germany. Robust standard errors clustered on the sampling point level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The school domain, however, shows a systematically lower level for migrant children, one that does not change with child care experience. Overall, however, with a few exceptions, we observe stronger effects for elementary than for secondary school children, which suggests that long term effects are smaller than short term effects.

Finally, we investigate the heterogeneous impact of child care on other child outcomes. Columns 1 and 2, Table 3.5, for example, show the estimates for the children's self-assessed KINDL sum score, which is only available for ages 10 to 17. For the total sample (columns 2 and 4, - Table 3.2), migrant children score significantly worse overall, although the effect of child care is positive albeit not statistically significant. The next four columns report outcomes based on the SDQ, whose first subscale assesses pro-social behavior using a sum score of multiple subscales measuring child behavioral problems (on all SDQ subscales, a higher value equals a greater amount of the behavior measured). As column 4 shows, children with early child care experience exhibit slightly less pro-social behavior. The estimates for the overall SDQ sum score (column 5) mimic the well-being regressions. Children with preschool child care experience and migrants show generally higher scores, which deviates from previous results by (Datta Gupta & Simonsen, 2010) who only find negative effects among boys from low SES households in family care, but not preschool child care. However, the interaction term in our regression tends to mediate both effects, suggesting that migrants tend to benefit from child care or are at least not negatively affected by it.

The estimates in Table 3.5, also reveal a significant relation between preschool child care and math and German skills, measured on the local 6-point valuation scheme, whose highest score of 1 means that positive coefficients signal negative outcomes. Based on the estimates in columns 7 to 10, all else being equal, migrants perform better overall in math; however, child care participation seems to slightly reduce performance later in school. Kita experience has a significantly negative coefficient (column 8), indicating that math performance is higher overall among children who experience child care at a very young age. Migrant children that attend kindergarten only, however, tend to perform worse in school. The estimates for German skills reveal a similar significantly negative relation with Kita participation, which is lower for migrant children, albeit not significantly so.

Table 3.5: RELATION BETWEEN CHILD CARE AND SELF-ASSESSED KINDL SUM SCORE, SDQ SUM SCORE, AND MATH/ LANGUAGE TEST SCORES

	OLS estimates for different child outcomes									
	Children aged 10 to 17		Children aged 2 to 17				Children aged (6) 8 to 17			
	KINDL sum score (self-assessed)		SDQ subscale Pro-social Behavior		SDQ sum score		Math score		German score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Child stayed at home as reference									
Child care	0.0788		-0.1031		0.6000***		0.0097		-0.0199	
	(0.556)		(0.062)		(0.159)		(0.044)		(0.035)	
Kindergarten		0.0102		-0.0857		0.6049***		0.0427		0.0078
		(0.563)		(0.064)		(0.163)		(0.044)		(0.036)
Kita		0.6322		-0.1516**		0.5033**		-0.1081**		-0.1321***
		(0.636)		(0.072)		(0.195)		(0.051)		(0.044)
Child care*Migrant	1.4433		0.1191		-0.6326*		0.1871**		0.0535	
	(1.474)		(0.160)		(0.368)		(0.086)		(0.090)	
Kindergarten*Migrant		1.5878		0.1268		-0.5381		0.2209**		0.0315
		(1.392)		(0.164)		(0.376)		(0.090)		(0.094)
Kita*Migrant		-0.3862		-0.0005		-0.2695		0.0785		0.0878
		(2.162)		(0.244)		(0.588)		(0.152)		(0.141)
Migrant	-2.5367*	-2.2870*	-0.0722	-0.0457	0.8411**	0.7540**	-0.2519***	-0.2697***	0.0949	0.0835
	(1.307)	(1.316)	(0.153)	(0.151)	(0.363)	(0.360)	(0.085)	(0.088)	(0.086)	(0.087)
N	4771	4628	10897	10596	10885	10586	6477	6297	6468	6286
Adj. R^2	0.051	0.054	0.032	0.033	0.092	0.092	0.120	0.125	0.170	0.176

Notes: All models include controls for child age, gender, dummy for having at least one sibling, number of individuals in the household, household net income, parental situation (married, single, living with new partners),parental age, parental education, parental employment, parental occupation, type of region (rural area, small/medium/large city), and East Germany. Robust standard errors clustered on the sampling point level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.6 Conclusions

In analyzing the relation between preschool child care and child well-being, as well as other child outcomes, we focus particularly on a possible heterogeneous experience for migrants. Overall, our results suggest that child care experience is associated with a slightly lower level in child well-being, with some evidence on the intensive margin that more years of preschool child care attendance reduce overall well-being. Although in general our findings support the existing literature (Belsky, 2006; Loeb et al., 2007; Magnuson et al., 2007), they differ in their focus on well-being as the primary measure for potential child care effects rather than the well-established child outcome variables such as cognitive skills (e.g., math or language scores) or amount of externalizing behavior. We are therefore able to make a valuable contribution to the knowledge gap on the relationship between well-being and early preschool child care. Of particular interest is our finding of higher well-being levels among migrant children, particularly those of elementary school age. Those findings are in line with recent results on the effect of preschool child care on school readiness in Germany (Dustmann et al., 2016; Felfe & Lalive, 2014). Even though we cannot fully rule out selection into child care based on unobservable characteristics that positively affect well-being measures, the mostly positive experience for migrant children emphasizes the importance of preschool child care for more than mere school success.

Seemingly, for children with a migratory background, not participating in preschool child care is associated with substantially lower levels of well-being. One possible explanation for this positive relationship with preschool child care (reflected by the large coefficients in columns 4 and 5, Table 3.2) may be that migrants with no preschool care experience enter compulsory elementary education without any familiarity with public education facilities, leading them to have more problems adapting to the new environment. Nonetheless, although this supposition is supported by the lack of evidence for a long term child care effect on the well-being of migrant children, it is contradicted by the persistently lower scores on the KINDL subscale for school readiness (column 6,

Table 3.4). The math and language skill results also raise questions about preschool child care's ability to increase the "school readiness" of migrant children in Germany.

Rather, the analyses of the single domain KINDL sum scores suggest that well-being is more driven by a social component. For example, the measures for psychological well-being and friends are significantly higher for both the full sample and both subsamples (elementary and secondary school), suggesting a long term benefit. Likewise, self-esteem tends to be systematically lower among migrant than nonmigrant children. An additional analysis of the SDQ sum score, which measures child behavioral problems, supports this pattern. Whereas overall preschool child care seems to be associated with more behavioral problems, migrant children seem unaffected. We also identify a negative association between child care and migrant children's math performance even though early child care seems to improve math and German skills in the total sample.

The differences we observe between migrants and nonmigrants make it difficult to extrapolate general policy implications. Nevertheless, the negative outcomes for German children, although rather small, raise questions about the implementation of mandatory kindergarten laws. Admittedly, our results might be driven by the recent shift in preschool child care toward more educative goals. Yet even though Kita attendees seem to perform slightly better in math and German, they show the strongest decline in well-being. It may be, therefore, that a focus on school outcomes comes at the expense of child well-being.

The results of our analysis also emphasize that migrant children who experience no preschool child care are much worse off than their German native counterparts in terms of the KINDL and most other scores. Hence, promoting preschool child care for migrant children might increase their overall well-being. Such promotion might take the form of information campaigns especially targeted at migrant families that explain the huge benefits of preschool child care and highlight the opportunities migrant children would miss by not attending.

4 The Effectiveness of a Population-Based Skin Cancer Screening Program: Evidence from Germany[7]

Abstract

In this paper, we analyze how a nationwide population-based skin cancer screening program (SCS) implemented in Germany in 2008 has impacted the number of hospital discharges following malignant skin neoplasm diagnosis and the malignant melanoma mortality rate per 100,000 inhabitants. Our panel data, drawn from the Eurostat database, cover subregions in 22 European countries, measured at the lowest nomenclature of territorial units for statistics (NUTS) level for 2000–2013. Applying fixed effects methods, we find a significantly positive and robust effect of the German SCS on the number of patients diagnosed with malignant skin neoplasm. However, the program does not significantly influence the melanoma mortality rate. This finding conflicts with the decreased melanoma mortality rate found for the pilot SCS program in northern Germany. Our results indicate that Germany's nationwide SCS program is effective in terms of a higher diagnosis rate for malignant skin neoplasms and thus may contribute to an improvement in the early detection of skin cancer.

[7] This chapter is based on joint work with Christopher Schreckenberger and Jörg Schiller both from the University of Hohenheim. The candidate´s individual contribution focused mainly on the set-up of the empirical estimation strategy and the writing. The work was published by Springer as 'Kaiser, M., Schiller, J. Schreckenberger, C. (2017): The Effectiveness of a Population-Based Skin Cancer Screening Program: Evidence from Germany, *The European Journal of Health Economics',* and is used for this thesis with permission of Springer. The work is available online: https://link.springer.com/article/10.1007%2Fs10198-017-0888-4. The author wants to thank the anonymous reviewers for their valuable comments.

4.1 Introduction

The incidence of skin cancer, the most frequently diagnosed cancer, is increasing in many industrialized countries, including the U.S. and Germany (American Cancer Society, 2017; MacKie, Hauschild, & Eggermont, 2009; Robert Koch Institut und die Gesellschaft für der epidemiologischen Krebsregister in Deutschland e. V., 2015; World Health Organization, 2016). In Germany, over 200,000 new cases of the commonest skin cancers – malignant melanoma (ICD–10 code C43), basal cell carcinoma, and squamous cell carcinoma (ICD–10 code C44), – were diagnosed in 2012. Moreover, although the mortality rate for malignant melanoma in Germany has remained relatively constant over the last 30 years, its age-standardized morbidity rate has more than tripled in the same time period. In 2012, over 20,000 individuals in Germany were diagnosed with malignant melanoma, and nearly 3,000 died from this disease (Robert Koch Institut und die Gesellschaft für der epidemiologischen Krebsregister in Deutschland e. V., 2015).

Skin cancer screening (SCS) may help to reduce morbidity and mortality from skin cancer by improved detection at an early stage (Breitbart et al., 2012; Choudhury, Volkmer, Greinert, Christophers, & Breitbart, 2012). In 2008, the German Statutory Health Insurance (SHI), which covers about 90% of the German population, introduced a nationwide population-based skin cancer screening (SCS) program. This program is the first of its kind worldwide (Choudhury et al., 2012) and primarily aims at reducing melanoma mortality (Eisemann, Waldmann, Garbe, & Katalinic, 2015) and thus mitigating the related health care expenditures by an early identification of skin cancer (Stang et al., 2016). Under this program, SHI enrollees who are at least 35 years old are entitled to a whole body examination every two years (Geller et al., 2010). General practitioners and dermatologists are eligible to screen patients after completion of a standardized training program (Geller et al., 2010; Veit, Lüken, & Melsheimer, 2015). Additionally, since these screenings are free of charge, the insured are encouraged to check suspicious lesions as early as possible.

With respect to the effectiveness of the nationwide SCS program in Germany, a microsimulation of melanoma mortality in Germany predicted about a 45% reduction twenty years after the implementation of a biennial population-based SCS in 2008. More interestingly, this simulation predicts a relative decline of the mortality rate by about 14% to 17% five years after the implementation of a SCS program with a 2-year screening interval (Eisemann et al., 2015). An evaluation of the SCS program also documented a greater improvement in the malignant skin tumor detection rate in the first years after implementation than figures from previous years would have predicted (Veit et al., 2015).

As a basis for implementing the program nationwide, a research-based pilot project, Skin Cancer Research to Provide Evidence for Effectiveness of Screening in Northern Germany (SCREEN), was carried out in the federal state of Schleswig-Holstein in Northern Germany between 2003 and 2004. Several studies have shown a substantial impact of this project on the incidence of melanoma and nonmelanoma and on melanoma mortality (Breitbart et al., 2012; Eisemann et al., 2014; Katalinic et al., 2012; Waldmann et al., 2012). Waldmann et al., for instance, report that the incidence of melanoma increased during the SCREEN period compared to the pre-SCREEN period, while it decreased after this pilot project. Additionally, the results were compared with incident rates in the state of Saarland, where the pilot study was not conducted. The authors found that in Saarland the incidence rate only slightly increased between the pre- and post-SCREEN period (Waldmann et al., 2012). Katalinic et al. identify an almost 50% decrease in melanoma mortality in Schleswig-Holstein between the prescreening period (1998-1999) and 2008-2009, while the melanoma mortality rate in other German regions remained relatively constant over the same period (Katalinic et al., 2012). The findings of a study by Stang and Jöckel, however, raised doubts about the SCS program’s effectiveness in reducing the melanoma mortality rate based on the fact that incorrect assignment of some skin melanoma deaths could upwardly bias the rate recorded for SCREEN (Stang & Jöckel, 2016). This doubt is supported by other research evidence of a slight increase in melanoma mortality five years after SCS implementation in 2008 (Katalinic, Eisemann, & Waldmann, 2015; Stang & Jöckel, 2016). This research inconclusiveness echoes a similar ambiguity in earlier evidence of the benefits and efficacy of population-based SCS programs in reducing mortality or increasing the proportion of skin cancers detected at earlier stages (Choudhury et al., 2012; Federman, Kirsner, & Viola, 2013).

In this paper, we analyze the SCS's impact on the number of hospital discharges following a diagnosis of malignant skin neoplasm (ICD–10 code C43_C44) and the mortality rate from malignant melanoma (ICD–10 code C43) per 100,000 inhabitants. Our main research question is whether the national SCS program introduced in Germany in 2008 has been effective for these outcome variables. To answer this query, we empirically compare the regional skin cancer diagnosis and mortality rates for Germany with those for other European countries between 2000 and 2013. By applying a fixed effects model to panel data to assess the effects of the SCS program implementation on our outcome variables, we make a valuable contribution to the literature.[8] As far as we know, ours is the first study to use advanced panel data methods to analyze the effectiveness of a population-based SCS program. Our results contribute especially to the literature assessing the effectiveness of SCS programs.

4.2 Materials and Methods

Using Eurostat data for 22 European countries, we derive a valid proxy for SCS effectiveness by extracting hospital discharges by diagnosis (*diagnosis)* and causes of death (*mortality*) per 100,000 inhabitants. Because we are interested in skin cancer diagnosis and mortality, we focus on the diagnosis rate for malignant skin neoplasms (ICD–10 code C43_C44) and the mortality rates from malignant melanoma (ICD–10 code C43). We also obtain the following covariates from Eurostat: the proportion of individuals aged 65 and older (*age 65+*), the sexratio (*sexratio*), the proportion with tertiary education (*educ*), the proportion of medical doctors per 100,000 inhabitants (*docdens*), the GDP per capita in logarithmic scale (*logGDP*), and employment rates (*employ*). All variables cover the same 2000–2013 time span except *mortality*, which covers only 2000 to 2012. The choice of our covariates is mainly based on the seminal work of Grossman which implies that the optimal choice of health investment is essentially influenced by the age, wealth and education of an

[8] Since there was SCS in Schleswig-Holstein before the implementation of the nationwide SCS program in 2008, we take the SCREEN project in Schleswig-Holstein in two ways into account. While we exclude Schleswig-Holstein in the descriptive statistics, we consider this project in our empirical model.

individual (Grossman, 1972). Furthermore, past studies that evaluate socioeconomic risk factors of different cancer types tend to rely on a similar set of variables (Aarts, Lemmens, Louwman, Kunst, & Coebergh, 2010; Conway et al., 2015; Little & Eide, 2012). To obtain information about regional differences in the variables in our sample, we decompose every country into the lowest possible regional level indicated by the nomenclature of territorial units for statistics (NUTS) (Eurostat, 2016). As Table 4.1 shows, our final sample includes 1,512 observations from 22 countries that are divided into 108 subregions.

Table 4.1: NUMBER OF SUBREGIONS AND OBSERVATIONS BY COUNTRY

Country	Subregions	Observations
Austria	9	126
Belgium	1	14
Bulgaria	1	14
Switzerland	1	14
Czech Republic	8	112
Germany	16	224
Denmark	1	14
Spain	18	252
Finland	1	14
France	22	308
Croatia	1	14
Hungary	1	14
Ireland	1	14
Italy	19	266
Netherlands	1	14
Norway	1	14
Poland	1	14
Portugal	1	14
Romania	1	14
Sweden	1	14
Slovenia	1	14
Slovakia	1	14
$\sum \textbf{countries} = 22$	$\sum \text{regions} = 108$	$\sum \text{observations} = 1{,}512$

A comparison of the descriptive statistics of Europe and Germany (excluding Schleswig-Holstein) given in Table 4.2 shows that on average, the ratio of males to females, i.e. the sexratio, is almost the same in Europe and Germany. Furthermore, the density of medical doctors is nearly the same in Europe and in Germany, whereas the GDP per capita, employment rate, tertiary education rate, and proportion aged 65+ tend to be lower for non-German regions. The same holds true for both the average number of skin cancer diagnoses (*diagnosis*) and the average melanoma mortality rate (*mortality*). Expressed in percentages, the differences in *diagnosis* and *mortality* per 100,000 inhabitants between Germany and other European countries are about 125% (99.74 vs. 44.42) and about 5% (3.00 vs. 2.86), respectively.

Table 4.2:SUMMARY STATISTICS FOR EUROPE AND GERMANY

	Europe (excluding Germany)				Germany (excluding Schleswig-Holstein)			
Variable	Obs.	Mean (SD)	Min	Max	Obs.	Mean (SD)	Min	Max
age 65+ (in %)	1,287	17.365 (3.049)	10.8	27.7	210	19.436 (2.288)	14.2	24.7
sexratio	1,288	0.954 (0.025)	0.886	1.051	210	0.9572 (0.011)	0.930	0.983
docdens (no. per 100,000)	1,245	366.159 (109.080)	192.653	976.253	210	366.406 (63.352)	259.073	588.665
GDP (in EURO)	1,288	23,585.02 (10,090.5)	1,800	79,000	210	28,269.05 (8,834.7)	15,800	54,600
educ (in %)	1,286	20.831 (8.345)	6.5	46.8	210	26.175 (4.573)	15.3	37.2
employ (in %)	1,286	68.628 (6.262)	46.3	83.3	210	74.839 (3.654)	65.5	81.5
diagnosis (no. per 100,000)	1,212	44.420 (37.044)	1.2	305	210	99.741 (27.293)	47.2	184.7
mortality (no. per 100,000)	1,148	2.869 (1.010)	0.6	7.2	195	3.000 (0.686)	0.2	4.82

Notes: Sample size varies slightly within each variable because of missing values.

A more sophisticated representation of the diagnosis and mortality rate differences can be obtained from the time trends for both variables, represented in Figure 1 as 2000–2013 skin cancer-related hospital discharges for Europe versus Germany. Whereas the trends for Europe appear stable across the period, Germany shows not only a monotonically increasing pattern of hospital discharges (that begins to intensify around 2006), but also a higher number of diagnoses. Interestingly, this sharp increase in diagnosis rate starts a few years before SCS implementation, which is consistent with program evaluations documenting the highest increase in reported skin cancer cases between 2007 and 2008 (Veit et al., 2015). These different trends can also be seen in Table 4.3, which splits the sample into four subsamples. Here, the 84.88 mean value of diagnoses per 100,000 persons in German regions (excluding the federal state of Schleswig-Holstein) between 2000 and 2007 rises by about 41% to 119.56 between 2008 and 2013.[9] Compared with this substantial increase, the 7% rise in other European regions (from 50.53 to 53.97) seems paltry.

The 2000–2012 pattern of average melanoma mortality, in contrast, is similar in both Europe and Germany, with both lines in Figure 1 characterized by an overall increase. At the same time, the skin cancer mortality rate is slightly higher in Germany than in the other European regions. Considering the pre- (2000–2007) and post- (2008–2012) implementation periods separately, Table 4.3 shows about a 25% increase in *mortality* for Germany (from 2.73 to 3.43) but only a 10% increase for other European regions (from 2.84 to 3.12).[10] Admittedly, however, these descriptive differences in the dependent variables *diagnosis* and *mortality* (y_{it}) before ($T_{it} = 0$) and after SCS implementation ($T_{it} = 1$) do not address the hypothetical counterfactual of diagnosis and mortality trends in Germany had the program not been implemented. Hence, comparing the average outcomes of different German regions before and after SCS implementation leads to a selection bias and therefore to biased estimates (cf. Angrist & Pischke, 2009). To circumvent this selection bias while measuring the SCS's effect on diagnoses and mortality

[9] We excluded the federal state of Schleswig-Holstein from Germany due to the SCREEN project, which was carried out between 2003 and 2004. However, the mean values of *diagnosis* and *mortality* for Germany only slightly change when Schleswig-Holstein is included.

[10] A policy intervention, such as the implementation of the SCS program in Germany, may lead to special effects around the year of implementation. Hence, we additionally compared the means for the diagnosis and mortality rate between the time period between 2000-2007 and 2009-2013. The differences between these periods only slightly differ from our main specification of the time periods. Thus, we suggest that there are not any special effects around the year of implementation.

rates, we set up a fixed effects model[11] that uses a hypothetical control group of all the European countries in our sample except Germany.

Table 4.3: OUTCOME VARIABLES FOR GERMANY AND EUROPE BEFORE AND AFTER 2008

Outcome	2000-2007		2008-2013	
	Germany (excluding Schleswig-Holstein)			
	Observations	Mean (no. per 100,000)	Observations	Mean (no. per 100,000)
diagnosis	120	84.878	90	119.559
mortality	120	2.733	75	3.429
	Europe (excluding Germany)			
	Observations	Mean (no. per 100,000)	Observations	Mean (no. per 100,000)
diagnosis	805	50.532	673	53.969
mortality	791	2.841	614	3.155

Notes: The sample does not include data on mortality in 2013.

For example, this method has been employed by Sabates and Feinstein (2006) to test education's effect on cervical cancer screening.[12] We estimate the impact of the SCS program on our outcome variables using the following model:

$$y_{it} = \alpha T_{it} + \beta X'_{it} + \gamma_i + \delta_t + u_{it} \quad (1)$$

where y_{it} is either hospital discharges by *diagnosis* for skin neoplasms or *mortality* from melanoma per 100,000 inhabitants in subregion i in year t. As Figure 2 shows, *mortality* seems to be normally distributed, while *diagnosis* is skewed to the right.

[11] Since we reject the null hypothesis of the Hausman specification test, we use a fixed effect instead of random effect technique in our model. In addition, we use a negative-binomial fixed effect for *diagnosis* as dependent variable, since we can reject the null hypothesis of the Kolmogorow-Smirnow test for an underlying normal distribution. Moreover, we can reject the null hypothesis of a zero value of alpha by a likelihood-ratio X^2 test, which suggest overdispersion of the data. Hence, a negative-binomial model is superior to a Poisson model.

[12] Cf. Angrist & Pischke (2009) or Wooldridge (2013) for the mechanisms underlying fixed effects regression techniques.

Figure 4.1: *Trends in skin cancer-related hospital discharges per 100,000 inhabitants from 2000 to 2013 (a) and in malignant melanoma mortality rates from 2000 to 2012 (b) for Europe (excluding Germany) and Germany (excluding Schleswig-Holstein)*

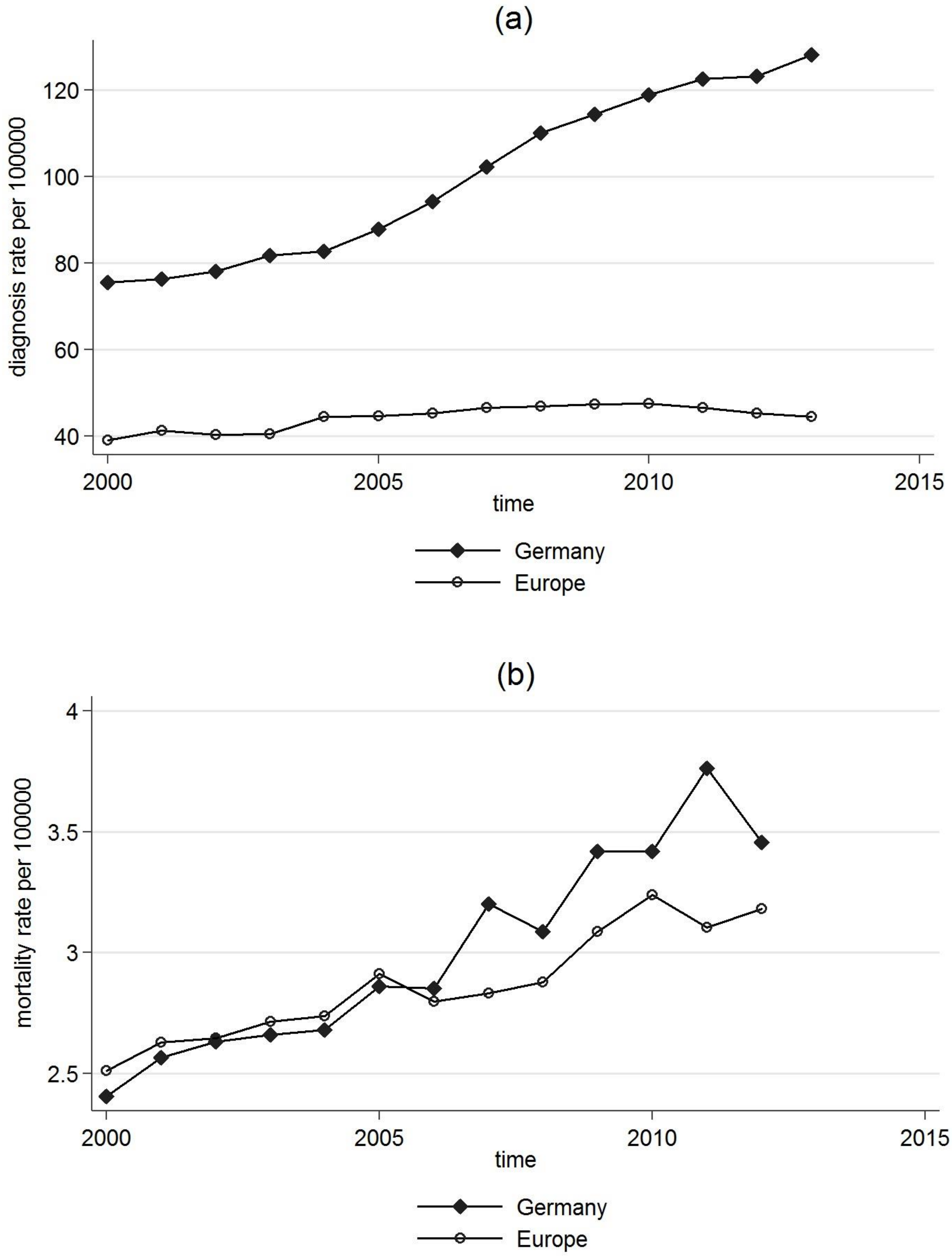
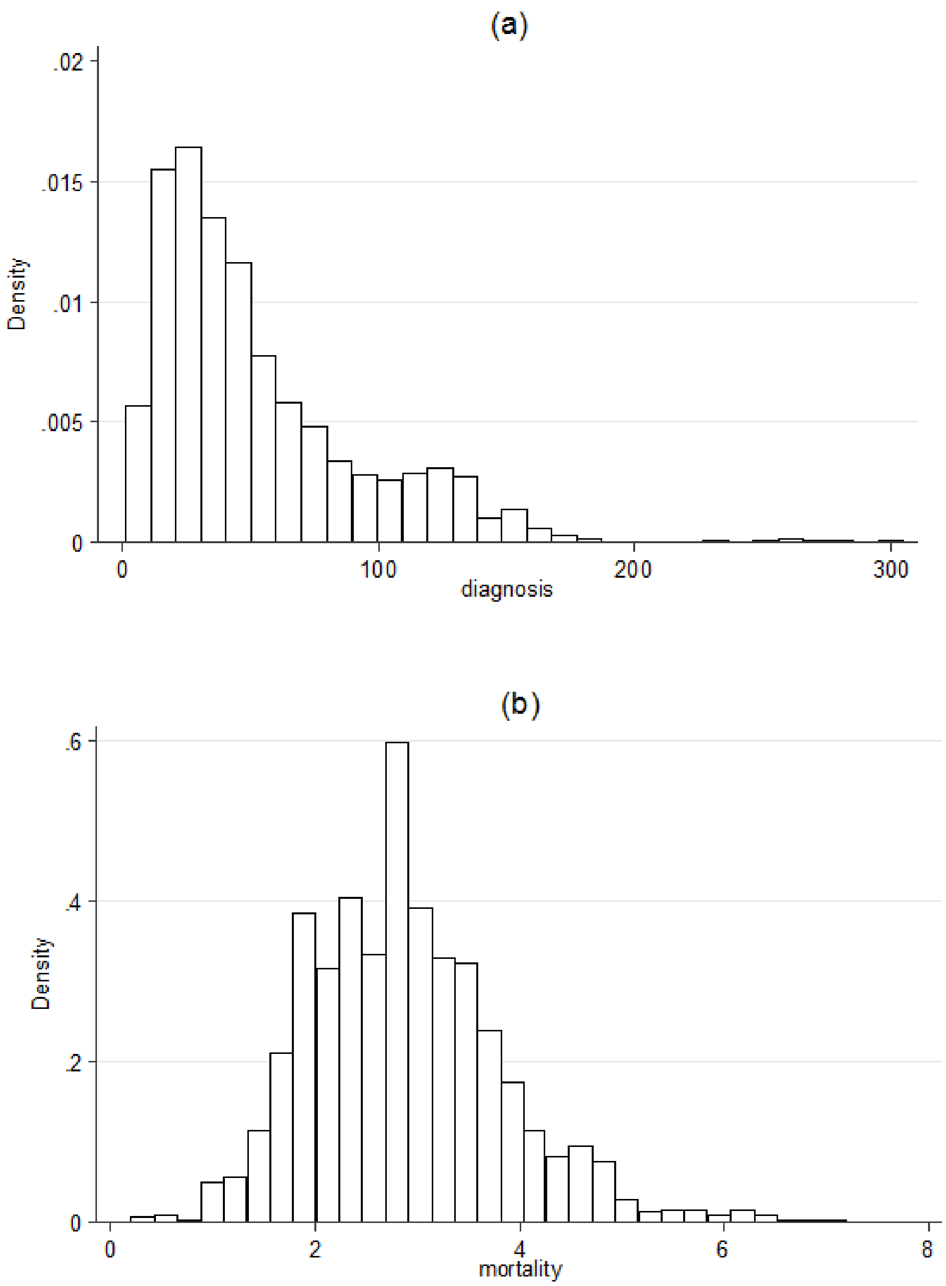


Figure 4.2: *Distribution of skin cancer-related hospital discharges per 100,000 inhabitants (a) and of malignant melanoma mortality rates (b)*



We therefore use a negative-binomial fixed effect for *diagnosis* as the dependent variable. The treatment[13] variable T_{it} is a dummy equal to 1 if subregion *i* had a SCS program in year *t* (i.e., it is 1 for German subregions only) and 0 otherwise. We also consider the SCREEN pilot by assigning the value 1 to this variable for Schleswig-Holstein in 2003 and 2004. X'_{it} is a $1 \times N$ vector of covariates with additional information about certain characteristics of region i in year *t.* To control for unobserved but constant regional heterogeneity and for heterogeneity among the different time periods, we include dummies for regional (γ_i) and annual (δ_t) fixed effects.

4.3 Results

Table 4.4 shows the impact of Germany's SCS program on our outcome variables using the fixed effects model in equation (1). The coefficient of our treatment variable representing the SCS's effect on diagnosed cases of malignant skin neoplasm is significantly positive (0.276) in the base model without covariates (column 1a). This positive effect remains significant with the addition of the different covariates (columns 1a-5a) even though the coefficient declines to 0.181 in the full model (column 5a). The coefficient for the treatment dummy indicates that sub-regions which participated in the SCS program are characterized by an incidence-rate-ratio of about 1.2. Specifically, this means that the implementation of the SCS in Germany has caused an increase of 20% in the rate of diagnoses of malignant skin neoplasms compared to the hypothetical counterfactual scenario. Looking at our covariates, we find that physician density, employment rate and the GDP per capita are the greatest contributors to this decreasing treatment coefficient, with higher physician density, employment rate, and GDP per capita associated with more diagnoses of malignant skin neoplasms. Moreover, the sexratio also

[13] In this paper, the term "treatment" generally refers to the implementation of the German SCS program and the SCREEN pilot program in Schleswig-Holstein. To avoid any confusion from the difference between the SCS program implementation as a treatment effect in our empirical model and a medical treatment of skin cancer, we will explicitly call the latter "medical treatment".

positively correlates with the number of diagnoses, while a higher proportion of 65+ and tertiary educated individuals negatively affect the number of diagnoses.

With respect to the melanoma mortality rate, Table 4.4 shows that it is only significantly and positively correlated by the SCS program (*treatment*) in the base model (column 1b); once the covariates are added in, the relation is not significantly different from zero (column 2b–5b). These results indicate that the SCS did not significantly influence the malignant melanoma mortality rate between 2008 and 2012, meaning that we cannot confirm the finding of an increased melanoma mortality rate five years after SCS implementation (Katalinic et al., 2015; Stang & Jöckel, 2016). In contrast to the findings for diagnosis, mortality rate is significantly and positively linked only to the proportion of those aged 65+ and physician density and negatively associated with the employment rate.

To test the robustness of our results, we first apply a pooled regression model with the malignant skin neoplasms diagnostic rate as dependent variable[14] and the same covariates as in equation (1). As shown in Table 4.5, the results using this model confirm our findings with respect to the effect of the SCS program. Moreover, we apply a difference-in-difference-in-difference technique in order to exclude any possible unmeasured confounding factors that are likely to affect the diagnosis for both malignant skin neoplasms (ICD–10 code C43_C44) and other neoplasms (ICD–10 code C00_D48), i.e. general risk factors for cancer, such as smoking behavior, diet or alcohol consumption. Besides considering the pre- and post-SCS implementation period for Germany and Europe, we also include the diagnosis of neoplasms other than malignant skin neoplasms. In Table 8.2 presented in the Appendix, we sketch our difference-in-difference-in-difference approach.

[14] The rationale why we base our pooled regression model on a negative-binomial distribution of *diagnosis* is equivalent to the rationale why we base our fixed effect model on this distribution.

Table 4.4: EFFECT OF THE GERMAN SCS ON HOSPITAL DISCHARGES BY DIAGNOSIS FOR MALIGNANT SKIN NEOPLASM AND MALIGNANT MELANOMA MORTALITY RATE

	Dependent Variable									
	Hospital discharges by diagnosis per 100,000 inhabitants for malignant skin neoplasm (ICD–10 code C43_C44)					Malignant melanoma mortality rate (ICD–10 code C43) per 100,000 inhabitants				
	(1a)	(2a)	(3a)	(4a)	(5a)	(1b)	(2b)	(3b)	(4b)	(5b)
Treatment	0.276***	0.291***	0.219***	0.214***	0.181***	0.242**	0.143	0.119	0.117	0.077
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.08)	(0.09)	(0.09)	(0.09)	(0.10)
Proportion of individuals aged 65 and older (in %)r		-0.004	0.004	-0.010	-0.023**		0.048*	0.0459*	0.046*	0.094***
		(0.01)	(0.01)	(0.01)	(0.01)		(0.02)	(0.02)	(0.02)	(0.03)
Sexratio		7.548***	5.723***	6.130***	3.432**		-1.047	-0.391	-0.340	4.034
		(1.11)	(1.08)	(1.07)	(1.05)		(2.30)	(2.26)	(2.28)	(3.35)
Physician density (no. per 100,000)			0.001***	0.001***	0.001***			0.000+	0.000+	0.001*
			(0.00)	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)
Proportion of individuals (aged 25–64) with a tertiary education (in %)				-0.121***	-0.124***				-0.000	0.012
				(0.00)	(0.00)				(0.01)	(0.01)
Employment rate (in %)					0.032***					-0.018+
					(0.00)					(0.01)
Natural log of GDP per capita (in EURO)					0.307***					0.030
					(0.07)					(0.19)
Constant	4.454***	-2.595*	-0.830	-0.668	-3.034**	2.506***	2.710	1.948	1.894	-2.390
	(0.12)	(1.07)	(1.05)	(1.05)	(1.12)	(0.09)	(2.29)	(2.25)	(2.26)	(3.47)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1,436	1,435	1,397	1,396	1,396	1,356	1,355	1,315	1,314	1,314

Notes: The treatment variable equals 1 for all German subregions as of 2008 and for the German federal state of Schleswig-Holstein in 2003 and 2004. For *diagnosis* as dependent variable, we counter the right-skewed distribution by using a negative binomial fixed effects model. Standard errors are in parentheses. +p < 0.1, * p < 0.5, ** p < 0.01, *** p < 0.001.

Table 4.5: EFFECT OF THE GERMAN SCS ON THE OUTCOME VARIABLES USING POOLED REGRESSION

	Dependent Variable									
	Hospital discharges by diagnosis per 100,000 inhabitants for malignant skin neoplasm (ICD–10 code C43_C44)					Malignant melanoma mortality rate (ICD–10 code C43) per 100,000 inhabitants				
	(1a)	(2a)	(3a)	(4a)	(5a)	(1b)	(2b)	(3b)	(4b)	(5b)
Treatment	0.919***	1.165***	1.155***	1.188***	0.852***	0.279*	0.280+	0.251+	0.263+	-0.231
	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)	(0.14)	(0.15)	(0.15)	(0.16)	(0.16)
Proportion of individuals aged 65 and older (in %)		-0.055*	-0.054*	-0.052*	-0.040+		0.004	0.012	0.012	0.014
		(0.03)	(0.03)	(0.03)	(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
Sexratio		-12.434***	-11.141***	-10.389**	-13.105***		-4.638	-4.145	-3.827	-7.222*
		(2.99)	(3.19)	(3.45)	(2.71)		(3.74)	(3.67)	(3.54)	(3.07)
Physician density (no. per 100,000)			0.001+	0.001+	0.001*			0.000	0.000	0.000
			(0.00)	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)
Proportion of individuals (aged 25–64) with a tertiary education (in %)				-0.008+	-0.031**				-0.003	-0.038***
				(0.01)	(0.01)				(0.01)	(0.01)
Employment rate (in %)					0.059***					0.079***
					(0.01)					(0.01)
Natural log of GDP per capita (in EURO)					0.027					0.314*
					(0.12)					(0.15)
Constant	3.828***	16.515***	14.945***	14.343***	12.880***	2.507***	6.853+	6.103+	5.846+	1.336
	(0.07)	(2.88)	(3.06)	(3.24)	(2.73)	(0.09)	(3.65)	(3.61)	(3.50)	(3.88)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1,436	1,435	1,397	1,396	1,396	1,356	1,355	1,315	1,314	1,314

Notes: The treatment variable equals 1 for all German subregions as of 2008 and for the German federal state of Schleswig-Holstein in 2003 and 2004. For *diagnosis* as dependent variable, we counter the right-skewed distribution by using a negative-binomial regression model. Standard errors are in parentheses. +p < 0.1, * p < 0.5, ** p < 0.01, *** p < 0.001.

Table 4.6: EFFECT OF THE GERMAN SCS ON THE HOSPITAL DISCHARGES BY DIAGNOSIS PER 100,000 INHABITANTS FOR MALIGNANT SKIN NEOPLASM USING A DIFFERENCE-IN-DIFFERENCE-IN-DIFFERENCE TECHNIQUE

	Dependent Variable Hospital discharges by diagnosis per 100,000 inhabitants for malignant skin neoplasm (ICD–10 code C43_C44)				
	(1)	(2)	(3)	(4)	(5)
Treatment	0.358***	0.324***	0.311***	0.312***	0.306***
	(0.02)	(0.02)	(0.11)	(0.02)	(0.02)
Proportion of individuals aged 65 and older (in %)		0.022***	0.022***	0.019***	0.015***
		(0.00)	(0.00)	(0.00)	(0.00)
Sexratio		0.932*	0.195	0.160	-0.890*
		(0.38)	(0.39)	(0.38)	(0.39)
Physician density (no. per 100,000)			0.000***	0.000***	0.000***
			(0.00)	(0.00)	(0.00)
Proportion of individuals (aged 25–64) with a tertiary education (in %)				-0.002+	-0.003*
				(0.00)	(0.00)
Employment rate (in %)					0.008***
					(0.00)
Natural log of GDP per capita (in EURO)					0.244***
					(0.03)
Constant	5.221***	4.048***	4.717***	4.844***	3.094***
	(0.05)	(0.37)	(0.38)	(0.38)	(0.43)
Year dummies	Yes	Yes	Yes	Yes	Yes
N	2,872	2,870	2,794	2,792	2,792

Notes: The treatment variable equals 1 for all German subregions as of 2008 and for the German federal state of Schleswig-Holstein in 2003 and 2004. we counter the right-skewed distribution by using a negative-binomial regression model. Standard errors are in parentheses. +p < 0.1, * p < 0.5, ** p < 0.01, *** p < 0.001.

As reported in Table 4.6, our finding that the German SCS program significantly and positively affects the number of hospital discharges following malignant skin neoplasm diagnosis using the difference-in-difference-in difference technique confirms our result of our fixed effect estimation in Table 4.4.[15]

However, because of the right skewed distribution of *diagnosis*, we cannot exclude the possibility of inefficiently estimated coefficients of our fixed effects or pooled regression model. We therefore estimate a finite mixture model (FMM) that accounts for distribution heterogeneity (Deb, Gallo, Ayyagari, Fletcher, & Sindelar, 2011). As Table 4.7 shows, we confirm a positive effect of the SCS in subregions with a lower malignant skin neoplasm diagnosis rate (*component 1*), whereas the impact of the SCS program is, although still significantly positive, considerably smaller in subregions with a comparatively high rate (*component 2*).

Additionally, we apply a pooled OLS regression using the malignant melanoma mortality rate as the dependent variable and again identify a significantly positive effect of the SCS program in the base model (see column 1b, Table 4.5). We find no significant correlation between the implementation of the SCS program and the malignant melanoma mortality rate once all covariates are added into the model (column 5b, Table 4.5). These findings are consistent with our results of the fixed effects estimation in Table 4.4. Finally, we test for a "placebo effect" by applying the treatment variable in our model to Austria, which shows similar trends for both our outcome variables as well as for all covariates. Estimating our model as though Austria had also implemented an SCS program in 2008, however, yields no significant impact of the SCS program implementation on the diagnosis rate once all covariates are controlled for (see column 5a, Table 4.8). This finding indicates that although SCS implementation in Germany has affected the malignant skin neoplasm diagnosis rate, its implementation in Austria has had no such effect. Moreover, as Table 4.8

[15]Furthermore, we assume that any unobserved factors that could influence the diagnosis rate for malignant skin neoplasms do not systematically differ between 2000 and 2013. To assess this assumption, we re-estimated our fixed effect model for shorter time periods as this should exclude much of any unobserved confounding factors (not presented in the paper). We still find a significantly positive effect of the SCS program on diagnoses for skin neoplasms if we only consider the time period between 2006 and 2010 as well as the time period between 2007 and 2009.

also shows, the association between the SCS program and malignant melanoma mortality rate in Austrian subregions is not significantly different from zero, even in the base model without covariates.

Table 4.7: Effect of the German SCS on Hospital Discharges by Diagnosis for Malignant Skin Neoplasms using FMM

	Dependent Variable									
	Component 1					*Component 2*				
	(1a)	(2a)	(3a)	(4a)	(5a)	(1b)	(2b)	(3b)	(4b)	(5b)
Treatment	1.244***	1.408***	1.468***	1.446***	1.611***	0.181	0.245+	0.157+	0.192*	0.379***
	(0.08)	(0.07)	(0.07)	(0.08)	(0.12)	(0.12)	(0.14)	(0.08)	(0.08)	(0.07)
Proportion of individuals aged 65 and older (in %)		-0.048***	-0.035***	-0.042***	-0.053***		-0.028+	-0.001	0.050***	-0.001
		(0.01)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)	(0.01)
Sexratio		-12.313***	-10.544***	-11.957***	-15.513***		-7.303**	8.740***	14.422***	-9.750***
		(0.83)	(1.08)	(1.25)	(1.31)		(2.29)	(2.08)	(1.73)	(1.42)
Physician density (no. per 100,000)			-0.000	-0.000	-0.000			0.004***	0.005***	0.002***
			(0.00)	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)
Proportion of individuals (aged 25–64) with a tertiary education (in %)				0.000	-0.020***				-0.041***	-0.043***
				(0.00)	(0.09)				(0.00)	(0.00)
Employment rate (in %)					0.039***					0.063***
					(0.01)					(0.00)
Natural log of GDP per capita (in EURO)					-0.107					0.063
					(0.10)					(0.05)
Constant	3.512***	16.179***	14.214***	15.770***	17.926***	4.582***	12.119***	-5.254*	-11.084***	8.774***
	(0.05)	(0.85)	(1.13)	(1.26)	(1.69)	(0.08)	(2.16)	(2.10)	(1.70)	(1.54)
Year dummies	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
N	1,436	1,435	1,397	1,396	1,396	1,436	1,435	1,397	1,396	1,396

Notes: The treatment variable equals 1 for all German subregions as of 2008 and for the German federal state of Schleswig-Holstein in 2003 and 2004. We counter the right-skewed distribution of *diagnosis* by using a negative-binomial distribution in both components. Standard errors are in parentheses. +p < 0.1, * p < 0.5, ** p < 0.01, *** p < 0.001.

Table 4.8: PLACEBO EFFECTS OF APPLYING THE TREATMENT VARIABLE TO AUSTRIAN SUBREGIONS (FIXED EFFECTS MODEL)

	Dependent Variable									
	Hospital discharges by diagnosis per 100,000 inhabitants for malignant skin neoplasm (ICD–10 code C43_C44)					Malignant melanoma mortality rate (ICD–10 code C43) per 100,000 inhabitants				
	(1a)	(2a)	(3a)	(4a)	(5a)	(1b)	(2b)	(3b)	(4b)	(5b)
Treatment2	0.160***	0.123***	0.042+	0.034	0.017	0.105	0.043	-0.003	0.009	0.011
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.10)	(0.10)	(0.10)	(0.11)	(0.11)
Proportion of individuals aged 65 and older (in %)		0.035***	0.037***	0.026***	0.005		0.089***	0.088***	0.099***	0.104***
		(0.01)	(0.01)	(0.01)	(0.01)		(0.02)	(0.02)	(0.02)	(0.02)
Sexratio		6.362***	4.533***	4.913***	2.168*		1.261	2.169	2.256	3.960
		(1.15)	(1.09)	(1.10)	(1.07)		(3.01)	(3.18)	(3.18)	(3.35)
Physician density (no. per 100,000)			0.001***	0.001***	0.001***			0.001*	0.001*	0.001*
			(0.00)	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)
Proportion of individuals (aged 25–64) with a tertiary education (in %)				-0.009**	-0.010**				0.011	0.013
				(0.00)	(0.00)				(0.01)	(0.01)
Employment rate (in %)					0.037***					-0.017+
					(0.00)					(0.01)
Natural log of GDP per capita (in EURO)					0.284***					-0.019
					(0.07)					(0.19)
Constant	4.144***	-2.215*	-0.387	-0.372	-2.735*	2.512***	-0.142	-1.229	-1.691	-2.488
	(0.11)	(1.11)	(1.07)	(1.07)	(1.16)	(0.05)	(2.96)	(3.08)	(3.12)	(3.47)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1,436	1,435	1,397	1,396	1,396	1,356	1,355	1,315	1,314	1,314

Notes: The treatment2 variable equals 1 for all Austrian subregions as of 2008. For *diagnosis* as the dependent variable, we use a negative binomial fixed effects model. Standard errors are in parentheses. +p < 0.1, * p < 0.5, ** p < 0.01, *** p < 0.001.

4.4 Conclusions

This analysis of the effectiveness of Germany's nationwide population-based SCS program identifies a significant, robust effect of the SCS program on the number of malignant skin neoplasm diagnoses per 100,000 people. This finding is consistent with earlier reports of a positive effect of the SCREEN project and the national SCS program in Germany on the melanoma and nonmelanoma incidence (Breitbart et al., 2012; Eisemann et al., 2014; Waldmann et al., 2012). It should be noted, however, that patients diagnosed with ICD–10 code C44 skin cancers are often ambulatory (Robert Koch Institut und die Gesellschaft für der epidemiologischen Krebsregister in Deutschland e. V., 2015), meaning that the SCS's impact on the diagnosis rate could be higher than that recorded here for inpatient discharges.[16] Nevertheless, once we control for the different covariates, we identify no significant program effect on the melanoma mortality rate, so our findings fail to support prior evidence of a decreasing melanoma mortality produced by SCREEN (Breitbart et al., 2012; Katalinic et al., 2012). Furthermore, we cannot confirm the 14% to 17% decline of the malignant melanoma mortality rate in Germany five years after the implementation of the national biennial SCS, predicted in the microsimulation model of Eisemann et al. (2015).

The insignificant effect on the mortality rate may be partly explained by the complexity of determining the cause of death, which could bias disease-specific mortality rates (Black, Haggstrom, & Welch, 2002). Furthermore, the analysis is based on aggregate data, which do not allow us to control for individual skin cancer risk characteristics such as solarium attendance, sunburn prevention behavior, skin type, and/or existing moles. Furthermore, because our data cover comparatively few post-SCS periods, our results may not capture

[16] Both melanoma and non-melanoma skin cancer are more likely diagnosed by outside hospital services. Our data include information about the hospital discharges by diagnosis per 100,000 inhabitants for malignant skin neoplasm. It is fair to say that the number of these discharges include both the cases solely diagnosed in hospitals and cases already diagnosed by outside hospital services which are transferred to a hospital for treatment. The latter are also reported as hospital discharges by diagnosis per 100,000 inhabitants for malignant skin neoplasm. However, our data does not include cases that were diagnosed and solely treated by outside hospital services. Hence, our finding of a positive effect of Germany's SCS program on the malignant skin neoplasm diagnosis rate may underestimate the true effect.

possible long run effects on the mortality rate. On the other hand, while Welch and Black suggest that an increasing diagnosis rate combined with no significant change in mortality rate may indicate skin cancer overdiagnosis (Welch & Black, 2010)[17], such a finding could also be explained by a rising number of erroneously diagnosed benign skin lesions (Carli et al., 2003).[18]

Overall, our results indicate that Germany's nationwide population-based SCS program has been effective in terms of a higher diagnosis rate for malignant skin neoplasms and thus may have helped improve early detection of skin cancer. Future research on SCS effectiveness might consider extending our analysis by using longer time periods and individual data. Moreover, the German SCS program could additionally be evaluated in terms of process outcomes, such as the awareness of individuals for SCS and skin cancer prevention, improved quality of diagnosis among physicians or the acceptability among patients for SCS. Finally, a cost-effectiveness analysis should also be conducted to identify the extent to which prior findings on the cost-effectiveness of melanoma screenings can be confirmed for Germany's SCS program (Losina et al., 2007). In addition to assessing the justifiability of associated costs, such an analysis would provide reliable insights for policy decisions.

[17]Welsh and his colleagues, for instance, find evidence supporting melanoma overdiagnosis in the U.S. as a result of raised diagnostic scrutiny Welch, Woloshin, & Schwartz (2005)

[18]A rising number of erroneously diagnosed benign skin lesions would lead to systematic measurement error, which we do not consider a major problem in our analysis.

5 Benford's law as an indicator of income data reliability[19]

Abstract

This paper analyzes how closely different income measures conform to Benford's law, a mathematical predictor of probable first digit distribution across many sets of numbers. Because Benford's law can be used to test data set reliability, we use a Benford analysis to assess the quality of six widely used survey data sets. Our findings indicate that although income generally obeys Benford's law, almost all the data sets show substantial discrepancies from it, which we interpret as a strong indicator of reliability issues in the survey data. This result is confirmed by a simulation, which demonstrates that household level income data do not manifest the same poor performance as individual level data. This finding implies that researchers should focus on household level characteristics whenever possible to reduce observation errors.

[19]The following chapter is single authored work. The underlying analysis is based primarily on income data from the harmonized datasets and codebooks developed by Gateway to Global Aging, with funding help from the National Institute on Ageing (R01 AG030153, RC2 AG036619, R03 AG043052; see www.g2aging.org). Additional GDP data are taken from the World Bank (http://data.worldbank.org/indicator/NY.GDP.MKTP.CD). The manuscript was submitted to *Review of Income and Wealth* and is currently under review.

5.1 Introduction

The widespread use of survey data across both social and life sciences has led to the development over recent decades of a multitude of econometric and statistical methods designed to detect causal relations and make the most precise predictions possible. One crucial issue that remains, however, is measurement error, which may severely bias estimates.[20] Hence, although research streams in both econometrics and statistics are already concentrating on how to deal with measurement error-induced reliability problems, researchers could benefit greatly from advance knowledge about data quality.[21] Such knowledge is even more important for policy makers, who do not apply econometric techniques to the survey data on which they base their policy inferences, especially when those data have been collected for particular government purposes.

Yet although myriad methods already exist for checking the quality of a particular data set, most rely on expensive procedures like matching employee and employer data to enable comparison (for a detailed description, see Duncan & Hill, 1985; Mellow & Sider, 1983). An alternative, and less costly, option is to apply Benford's law of likely first digit distribution in a data set, and detect anomalies and possible quality problems by measuring the number of deviations from the theoretical pattern. This rule, however, although commonly used by tax authorities to detect fraud, is still not widely applied in the social sciences.[22] Authors who have used it to evaluate (survey) data quality include Judge and Schechter (2009), mainly for agricultural data; Nigrini and Miller (2007), for hydrological data; Sandron (2002), for population numbers; Mir (2014), for religious data; and Ausloos,

[20]Survey data tend to be subject to two different kinds of (nonsampling) measurement errors: random errors, such as those caused by interviewer inattentiveness, or nonrandom (systematic) errors, such as the inaccurate responses generated when the survey is badly constructed or includes problematic survey items (for a detailed discussion, see Groves, 2004; Saris & Gallhofer, 2014; Bowling, 2005)

[21]Saris and Revilla (2016) provide a useful overview of existing correction techniques for measurement errors in survey data. Moreover, a wide body of literature focuses on measurement error in time-series data. Such techniques usually exploit the data's distributional properties to overcome the uncertainty from measurement problems (e.g., the Kalman (1960) filter, a powerful tool for improving predictions in an evolving system).

[22]For instance, a Benford analysis facilitated the uncovering of the 2001 Enron accounting fraud (Nigrini, 2012, p. 207) .

Herteliu, and Ileanu (2015), for long-term birth numbers. The literature thus lacks large-scale analyses of how individual income data conforms to Benford's law, knowledge that could be used to improve the assessment of survey data quality. Given the widespread use of income data in both econometric and policy analyses, however, correct assessment of data quality is crucial for accurate inference.

This paper attempts to meet this need by making the following contributions to the existing literature: First, we show plausible reasons to consider income as generally in compliance with Benford's law. Second, we use the law to assess income data quality in six harmonized survey data sets that are widely used in economics and social science. Third, by using three different variables, we detect systematic differences in the quality and design of these income measures. Fourth, we introduce a simple but efficient simulation algorithm that improves the validity of a Benford analysis for any particular survey data set.

The paper is organized as follows: Section 5.2 describes Benford's law and the conditions for its applications (5.2.1), explains the analytical method used for our analysis (5.2.2), and describes the data sets analyzed (5.2.3). Section 5.3 reports the main results, and section 5.4 presents the specifications for and outcomes of the simulation test for robustness. Section 5.5 then discusses the results and concludes the paper.

5.2 Data and Methods

5.2.1 Benford's law

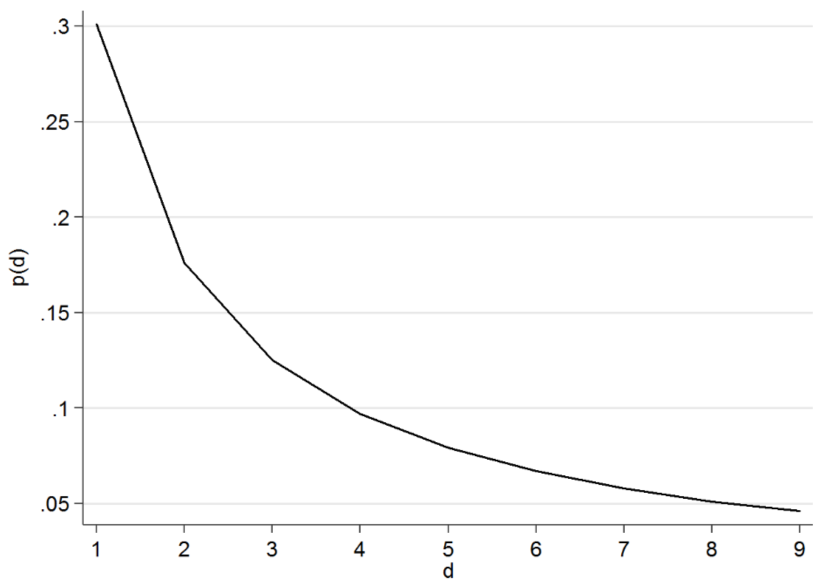
Although named for American physicist Frank Benford (1938), the phenomenon on which Benford's law is based was first reported by Francis Newcomb (1881), who noted a more frequent use of logarithmic tables that included numbers beginning with low digits. From this observation, he derived a mathematical rule for the probability p of first digits d occurring in the numbers of a given data set. This rule is characterized by the following logarithmic function (with B as the logarithmic base),

$$p(j) = log_B\left(1 + \frac{1}{j}\right) \tag{1}$$

which empirically predicts the occurrence of first digits in a broad variety of data sets. Benford (1938) independently made this same observation over a half century later and published his own first-digit law.[23]

In Figure 5.1, we illustrate Benford's law by mapping first digits $j \in \{1,2,3 \dots,9\}$ onto a probability space to produces a monotonically decreasing graph with higher probability values for lower digits and almost uniform values for higher digits. Whereas the probability of observing a number beginning with $j = 1$ in a given data set is approximately 6.5 times higher than that for a number beginning with $j = 9$, the probability for a number beginning with $j = 8$ (rather than $j = 9$) is only 1.1 times higher.

Figure 5.1: *Benford first digit distribution of numbers in a given data set*



Benford's law does not, however, characterize every data set: rather, its occurrence requires the presence of various criteria. Pinkham (1961), for instance, shows that changing the measurement scale should not change the first digit distribution in a set of numbers.

---

[23] A more generalized formulation of Benford's law describes the probability of the occurrence of a particular number j as the nth digit in the following form: $p(j) = \sum_{k=\lfloor B^{n-1} \rfloor}^{B^n-1} log_B(1 + \frac{1}{k \cdot B + j})$. Readers interested in the law's application beyond the first digits addressed here will find additional information in Hill (1995b) and Durtschi, Hillision, & Pacini (2004).

Thus, if the first digit occurrence probability in a data set expressed in kilometers changes when the same data are expressed in meters, the data are probably not following Benford's law. In subsequent work, Hill (1995a) further shows that Benford's law is characterized by base invariance, meaning that the first digit occurrence probability must not change with a change in the base B of the underlying logarithmic function (see equation 1). Moreover, as Pietronero, Tosatti, Tosatti, and Vespignani (2001) point out, the occurrence of Benford's law is a result of multiplicative processes, implying necessary Benford compliance by data generating processes that follow a Markov chain (Berger, Hill, Kaynar, & Ridder, 2011). This latter throws valuable light on why many observable (economic) data obey Benford's law: many economics processes (e.g., GDP growth, employment rates, or income development) are describable by Markov chains (Le Gallo, 2004).

The fact that the random numbers used in data generating processes need not be identically distributed may also explain the frequent adherence to Benford's law of "real world" data sets (Hill, 1995). For example, although the numbers used in scientific publications are invariably taken from a mix of different probability distributions, they tend to obey Benford's law to some extent (Tödter, 2009). The latter observation is also in line with Formann's (2010) simulation evidence that data from right-tailed distributions, particularly, are likely to obey Benford's law. Again, this adherence is likely if the resulting random variables in the data set stem from a mix such as the ratio of two half-normal distributions. The latter conversely implies, however, that data resulting from a symmetric distribution is unlikely to obey Benford's law, which in fact is seldom the case in economic data (e.g., income) because of its tendency to be log normally distributed (Clementi & Gallegati, 2005).

Wallace (2002) thus proposes a rule of thumb for data set adherence to Benford's law that expects fulfillment of two criteria: the mean of the data set should be higher than the median, and the data set should be characterized by a positive skewness value. This rule implies that to obey Benford's law, the data must have sufficient volatility, including a sufficiently broad range of numbers (Durtschi, Hillison, & Pacini, 2004). Durtschi et al. (2004) conversely propose the following exclusion restriction: a data set is not likely to obey Benford's Law if it is, for instance, "influenced by human thought," has a "built in minimum or maximum," or "is comprised of assigned numbers" (p. 24). As a result of the above, economic data are widely accepted to generally conform to Benford's law. Hence, any discrepancies in the data sets evaluated here may indicate a serious data reliability problem.

5.2.2 Hypothesis testing

To make the most accurate evaluations of data reliability, we incorporate several of the many methods for testing adherence to the first digit law but interpret the separate results as an aggregate. In particular, we combine a graphic analysis with three statistical methods: Pearson's (1900) chi-squared (χ^2) test as well as the Kolmogorov-Smirnov (1948) test for goodness of fit, and a distance measure developed by Leemis, Schmeiser, and Evans (2000).

The Pearson's goodness-of-fit uses the following test statistic:

$$\chi^2 = \sum_{j=1}^{m} \frac{(N_j - n_j)^2}{n_j} \tag{2}$$

which computes the sum of the squared (and standardized) deviations of the empirical number of observations N_j (with $n = \sum_{j=1}^{9} N_j$) for each of the first digits $j \in \{1,2,3 \ldots,9\}$ and the expected frequency $n_j = p(j) \times n$ as proposed by Benford's law. Given a particular significance level α, we will not reject the null hypothesis that the data obey Benford's law if the test statistic does not exceed the corresponding critical value (20.09 for $\alpha = 0.01$ and 15.51 (13.36) for $\alpha = 0.05$ ($\alpha = 0.10$), respectively).

The Kolmogorov-Smirnov (1948) test is expressed as

$$D_n = sup_x |F_n(x) - F(x))| \tag{3}$$

where $F(x) = p(x \leq j)$ represents the cumulative distribution function of the Benford distribution and $F_n(x) = \Pr_n(x \leq j)$ denotes the empirical cumulative distribution function (cumulative frequency) for all n observations.[24] Here, we reject the null hypothesis if the test statistic D_n exceeds the critical values, calculated as follows (Sachs, 2004, p. 427–431):

$$\frac{b_\alpha}{\sqrt{n}} \tag{4}$$

where $b_\alpha \in \{1.224, 1.358, 1.628\}$ depends on significance level. The distance measure (Leemis et al., 2000) used to test the degree of similarity between the first digit distribution in the data sets analyzed and those in the Benford distribution is then expressed by

$$m = max_{j=1,2,\ldots,9}\{|\Pr_n(j) - p(j)|\} \tag{5}$$

5.2.3 Data

The analysis evaluates six different longitudinal data sets from the Health and Retirement Study (HRS) family of studies, originated by the U.S. National Institute on Aging (NIA). All six focus on a broad range of health, wealth, and income issues, and include quality of life measures that provide insights into the life situations of older citizens. To accurately identify and compare the quality differences in the individual data sets, in five cases, instead of the originals, we employ the harmonized data sets provided by Gateway to Global Aging.[25] A major advantage of harmonization is that the resulting data sets tend to include similar variables, which facilitates both the analysis and interpretation of the separate analytical results:

HRS (America). The original Health and Retirement Study (HRS), funded by the NIA, whose first wave (1992–1993) served as a baseline for the remaining 11 waves (ending in 2014–2015), which all closely mirrored its structure. From an original sample size of 12,600 individuals over 51, the sample size increased to 18,700 by 2014.

---

[24]Here, $\Pr_n(j)$ equals the probability of observing a number beginning with digit j in a given data set with n observations (i.e., the relative frequency of numbers beginning with j in a given data set).

[25]For more information, see https://g2aging.org/?.

Harmonized ELSA (England). The English Longitudinal Study of Ageing (ELSA), administered to individuals over 50, was funded by the NIA and three different UK government departments. [26] Its sample declined from 12,000 in 2002–2003 to approximately 9,600 in 2014–2015. The harmonized dataset covers 4 of the original 6 waves.

Harmonized SHARE (Europe).[27] The Survey of Health, Ageing, and Retirement in Europe (SHARE), structured like the HRS and ELSA and funded by the European Commission, was administered to individuals over 50. Having been conducted in 19 European countries plus Israel, SHARE offers notably larger samples: 30,700 in the first wave and 68,200 in the latest. The harmonized data set covers four SHARE waves: 2004–2005, 2006–2007, 2010–2011, and 2012–2013.

Harmonized CRELES (Costa Rica). The Costa Rican Longevity and Healthy Aging Study (CRELES) was a joint project of the University of Costa Rica`s Centroamericano de Población, Instituto de Investigaciones, and the University of California, Berkeley. In contrast to its sister studies, the CRELES, administered in 5 waves from 2004-2005 to 2012-2013, includes two different cohorts: those over 55 and those over 60.

Harmonized TILDA (Ireland). The Irish Longitudinal Study on Ageing (TILDA), conducted with individuals over 50, was funded by the Department of Health, Atlantic Philanthropies, and Irish Life. The harmonized TILDA data set compiles two survey waves: 2010–2011 (sample = 8,500) and 2012–2013 (sample = 7,200).

Harmonized LASI (India). The Longitudinal Aging Study in India (LASI), funded by the NIA, the Government of India, and the United Nations Population Fund, differs from the other surveys in that it was only administered once, in 2010, to individuals over 45. Nonetheless, although the sample only includes 1,600 individuals, the survey is structured similarly to the HRS.

---

[26]The Department of Health, the Department of Work and Pensions, and the Department for Transport.

[27]SHARE data cover the following European countries: Austria, Belgium, Czech Republic, Denmark, Estonia, France, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, Netherlands, Poland, Portugal, Slovenia, Spain, Sweden ,and Switzerland.

To test whether the numbers in the data sets come close to following Benford's law, we focus on three different (income) variables measured over the previous 12 months: total household income (HITOT) from all sources; total respondents earnings (RIEARN) from both labor and trade; and spousal employment earnings (SIEARN) from both labor and trade.[28] One reason for choosing these particular variables is that not all the individual surveys covered by the harmonized data sets necessarily address the same topics, which makes it hard to compare data set quality. For instance, although it would be interesting to evaluate the reliability of individual health data (e.g., hospital stays per year), this information is only available in some surveys. Moreover, the health related data contained in almost every survey tend to refer to different time spans.[29]

At the same time, because the conditions for a set of numbers (or variables) to obey Benford's law are relatively strict, not every variable can be exploited for data reliability assessment using a Benford analysis. Information on cigarette intake per day, for example, although an interesting candidate for reliability testing, does not obey Benford's law because of its built-in maximum (Durtschi et al., 2004). Income, however, is a widely used analytic variable across scientific disciplines – especially in economics or social sciences – so the reliability of these income variables is central to assessing the validity of the corresponding analytic conclusions. In particular, because wealth related policy decisions tend to be based on major survey data, the data underlying income distribution information must be reliable.[30]

To avoid analytic distortion, our summaries of the mean, median, skewness, and number of observations for these three variables in each data set (Tables 5.1–5.3) exclude

[28]Minor differences in income variable composition among the different surveys include the inclusion (HRS) or exclusion (TILDA) of second job earnings in constructing the RIEARN and SIEARN variables (see the respective data set codebooks for more information).

[29]For instance, although almost all harmonized data sets include information about drinking behavior, some questionnaires ask respondents for their daily alcohol intake (e.g., ELSA or LASI), whereas others ask for the total number of drinks if the respondent is currently drinking (e.g., TILDA).

[30]For example, the German Federal Ministry of Labor and Social Affairs regularly publishes a Poverty and Wealth Report (*Armuts-und Reichtumsbericht*), which deals with income dynamics among Germans. The report's major analyses and conclusions are based on income data from the German Socio-Economic-Panel (GSOEP), a large-scale, national, longitudinal survey (see http://www.armuts-und-reichtumsbericht.de/DE/Startseite/start.html).

observations in which each variable has a zero value. For all data sets, the mean values for HITOT, RIEARN and SIEARN (Tables 5.1, 5.2, and 5.3, respectively) are higher than the medians, and all distributions appear positively skewed. This pattern is a strong indicator that the income variables used should generally obey Benford's law and hence be suitable for detecting reliability problems in the data (Wallace, 2002).

TABLE 5.1: DESCRIPTIVE STATISTICS FOR TOTAL HOUSEHOLD INCOME FOR THE DIFFERENT DATA SETS

HITOT	*Mean*	*Median*	*Skewness*	*Obs.*
HRS	57510.78	34800	189.72	224,287
ELSA	23854.1	18616	11.463	61,742
CRELES	16221.65	1440	102.56	10,703
LASI	124161.7	58900	5.62	1,504
SHARE	36651.25	22933.14	5.97	25,028
TILDA	59086.01	32028	19.534	12,579

Notes: The values are measured in the following units: HRS – nominal dollars; ELSA – nominal pounds; CRELES – 1,000 Costa Rican colons; LASI – Indian rupees; TILDA – euros; SHARE – euros, except for Denmark, Sweden, Switzerland, Poland, Czech Republic, Hungary, and Estonia. The first wave of SHARE includes before-tax income, whereas all subsequent waves consider after-tax income.

TABLE 5.2: DESCRIPTIVE STATISTICS FOR RESPONDENT EMPLOYMENT EARNINGS FOR THE DIFFERENT DATA SETS

RIEARN	*Mean*	*Median*	*Skewness*	*Obs.*
HRS	36446.42	25000	30.89	84,375
ELSA	14736.67	12009.24	12.70	19,796
CRELES	4015.614	1800	6.39	2,435
LASI	32656.59	13750	2.61	92
SHARE	22318.6	18000	1.44	9,836
TILDA	30899.19	25000	1.04	2,421

Notes: The values are measured in the following units: HRS – nominal dollars before taxes and other deductions; ELSA – nominal pounds after taxes and other deductions; CRELES – 1,000 Costa Rican colons; LASI – Indian rupees; TILDA – euros; SHARE – euros, except for Denmark, Sweden, Switzerland, Poland, Czech Republic, Hungary, and Estonia. The first wave of SHARE includes before-tax income, whereas all subsequent waves consider after-tax income.

TABLE 5.3: DESCRIPTIVE STATISTICS FOR SPOUSAL EMPLOYMENT EARNINGS FOR THE DIFFERENT DATA SETS

SIEARN	*Mean*	*Median*	*Skewness*	*Obs.*
HRS	38561.26	27000	31.91	63,085
ELSA	15047.19	12309.47	13.65	16,229
CRELES	3696.734	2400	3.86	1,390
LASI	36030.64	15000	2.40	70
SHARE	22827.07	18511.64	1.43	6,488
TILDA	32301.43	26000	0.977	1,389

Notes: The values are measured in the following units: HRS – nominal dollars before taxes and other deductions; ELSA – nominal pounds after taxes and other deductions; CRELES – 1,000 Costa Rican colons; LASI – Indian rupees; TILDA – euros; SHARE – euros, except for Denmark, Sweden, Switzerland, Poland, Czech Republic, Hungary, and Estonia. The first wave of SHARE includes before-tax income, whereas all u subsequent waves consider after-tax income.

5.3 Results

First, to verify the assumption that income measures should generally be Benford distributed, we also include World Bank GDP data from 1960 onward measured in current U.S. dollars for 264 countries. The total number of observations is 11,315, with a mean and median of US$966 billion and US$13.8 billion, respectively. This higher mean than median, combined with the positively skewed (8.44) GDP distribution, strongly suggests that the GDP data should obey Benford's law.

By comparing the relative frequencies for the first digits in the GDP data set with those proposed by Benford's law (Figure 5.2), we reveal a relatively good fit, with only minor deviations. This finding is confirmed by the fact that neither the chi-squared nor Kolmogorov-Smirnov tests exceed their respective critical values (see Table 5.4), leading to acceptance of the null hypothesis. Likewise, the maximum deviation in the distance test is 0.0033, which corresponds to a 7 percent maximum (occurring at the ninth digit), a negligible discrepancy. Both these results provide strong evidence for the assumption that income data obey Benford's law, with any deviation merely an indicator of a reliability issue in the underlying data.

We then compare the frequency graph for HITOT in the harmonized data sets with that for the Benford distribution (Figure 5.3), revealing clearly that the first digit distribution for the income variable in the SHARE, HRS, and LASI data sets is close to that of the Benford distribution. The income data from the harmonized ELSA and CRELES, in contrast, although still showing a general pattern, perform comparatively poorly in terms of Benford similarity, whereas the harmonized TILDA, although generally patterned close to Benford, shows a major discrepancy in the first number. The results revealed by this graphic analysis are largely confirmed by the statistical analysis (see Table 5.5). Although the HITOT variable seems to fully obey Benford's law only in the LASI data set (in which the null hypothesis cannot be rejected for either the chi-squared or Kolmogorov-Smirnov test), the deviations indicated by the chi-square values are lower for the HRS and SHARE data than for the ELSA and CRELES data. The chi-square for the TILDA analysis seems comparatively low.

Figure 5.2: *Relative frequencies for GDP and Benford*

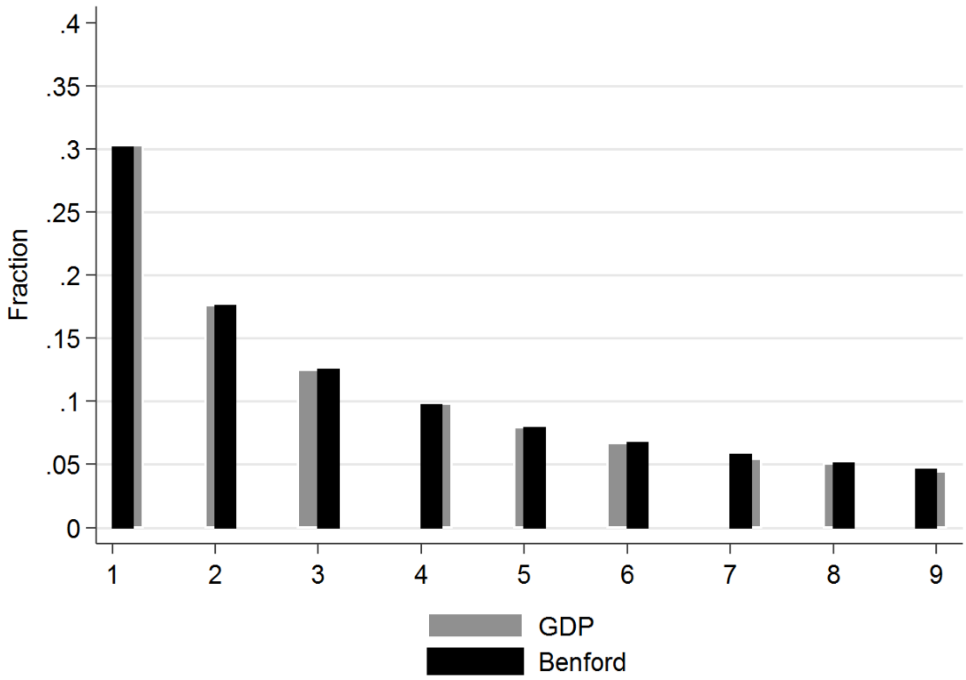
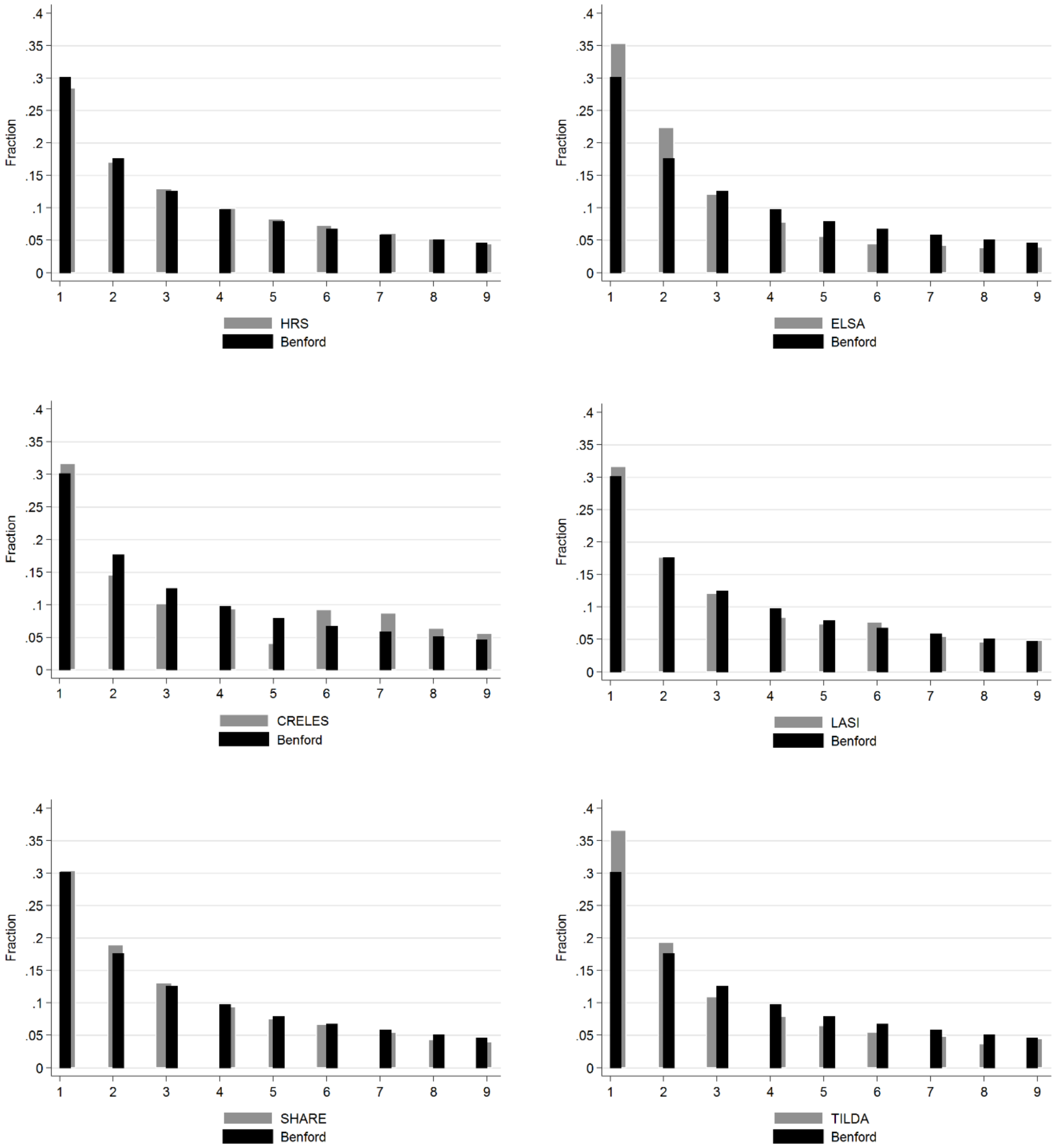


TABLE 5.4: TEST STATISTIC VALUES FOR GDP

	GDP
χ^2	2.9737
D_n	0.0046
m	0.0033
n	11315

Notes: This table lists the test statistics values for the chi-squared (χ^2), the Kolmogorov-Smirnov (D_n), and distance (m) tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5.3: *Relative frequencies for HITOT and Benford for the different data sets*



When the values for each maximum deviation are compared, however, the accuracy pattern implied by the graphics holds true: the values for the ELSA, TILDA, and CRELES are higher than those for the other data sets. Hence, taken together, the graphic and empirical analyses point to major differences in the reliability of the HITOT variable across the different data sets, with particularly poor performance in the ELSA, TILDA, and CRELES data sets.

As regards RIEARN and SIEAR (Figures 5.4 and 5.5, respectively), the graphic analysis suggests that the first digits of these earnings variables do not fit Benford's law as well as does household income. Although the generally decreasing probability pattern is still present, all data sets show more or fewer major deviations for both variables. This observation is supported by the statistical analysis in which both the chi-square and Kolmogorov-Smirnov values (see Table 5.5) show statistically significant differences from the Benford for the first digit distributions of RIEAN and SIEARN in all data sets except the LASI. In this latter, however, acceptance of the null is mainly attributable to the low number of observations and should thus be treated with caution, as should the comparatively low chi-square values for the CRELES and TILDA data. This caveat is supported by the overall occurrence in the distance measure (for all data sets except the TILDA) of much larger values than observed in the HITOT analysis. Given the results of both the statistical and graphic analyses, a general reliability issue with the RIEARN and SIEARN variables across all data sets analyzed is highly probable.

Figure 5.4: *Relative frequencies for RIEARN and Benford for the different data sets*

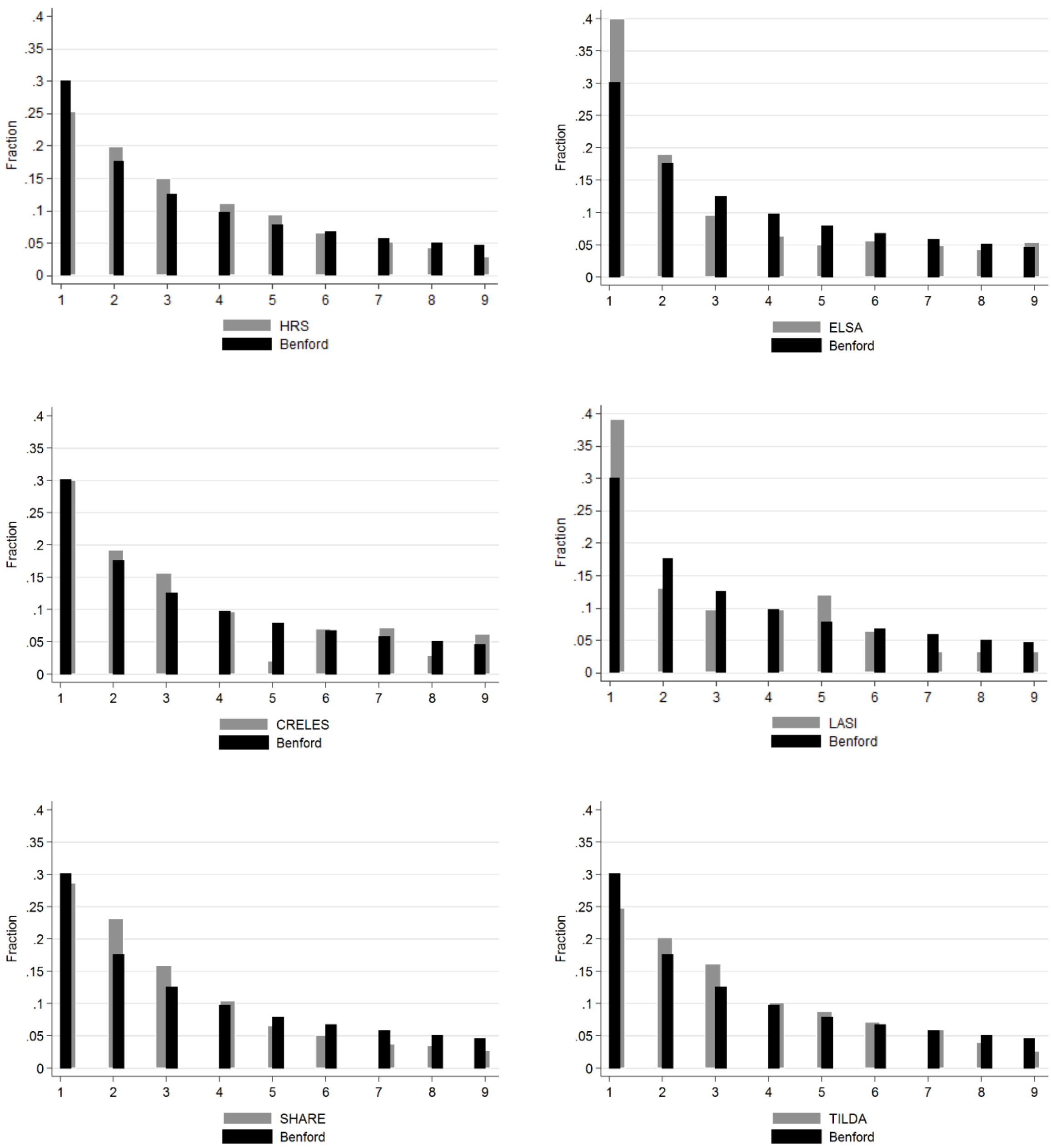


Figure 5.5: *Relative frequencies for SIEARN and Benford for the different data sets*

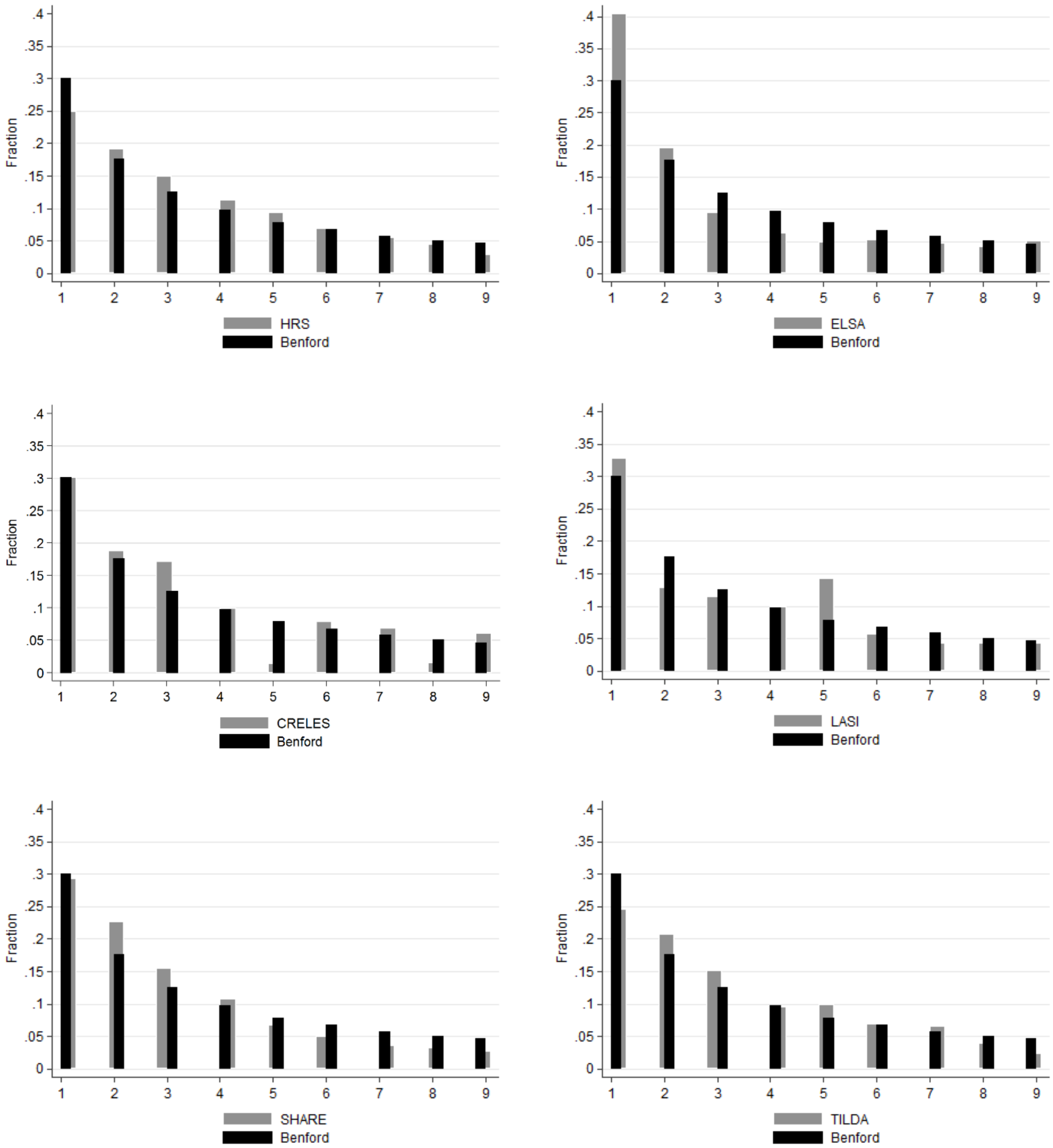


TABLE 5.5: TEST STATISTICS VALUES FOR THE DIFFERENT DATA SETS AND VARIABLES

	Variable											
	HITOT				*RIEARN*				*SIEARN*			
	χ^2	D_n	m	n	χ^2	D_n	m	n	χ^2	D_n	m	n
HRS	510.2509***	0.0211***	0.0159	224285	2314.7734***	0.0479***	0.0479	84375	1735.7136***	0.0511***	0.0511	63085
ELSA	2904.6676***	0.1001***	0.0523	61742	1375.7038***	0.1135***	0.099	19796	1212.4607***	0.1228***	0.1034	16229
CRELES	630.3626***	0.0783***	0.0380	10703	174.0999***	0.0468***	0.0582	2435	148.0976***	0.0621***	0.0647	1390
LASI	8.2094	0.0162	0.0154	1504	8.0090	0.0902	0.0902	92	5.2130	0.0361	0.0636	70
SHARE	90.1389***	0.0223***	0.0136	25028	518.7483***	0.0828***	0.0557	9836	322.6380***	0.0838***	0.0503	6488
TILDA	383.6861***	0.0821***	0.0649	12579	85.8518***	0.0527***	0.0527	2421	55.6337***	0.0548***	0.0548	1389
N												

Notes: This table lists the test statistics values for the chi-squared (χ^2), Kolmogorov-Smirnov (D_n), and distance (m) tests. The LASI results for RIEARN and SIEARN must be treated with caution because of the very small number of observations. * p < 0.1, ** p < 0.05, *** p < 0.01.

5.4 Simulation

Although the previous results raise doubts about the reliability of income data in longitudinal panel studies, these doubts are based on the assumption that income data should generally obey Benford's law. If the opposite were true, our analytic outcomes would inevitably lead to rejection of the null hypothesis because the first digit frequency patterns in the data sets analyzed would necessarily differ from those proposed by Benford. In that case, it would be impossible to use Benford's law for any valid assessment of data reliability. We thus avoid false inference by employing a simple Monte Carlo (MC) simulation in which we exploit information about the first and second moments of the HITOT variable in all our data sets to construct hypothetical samples, which are then tested for Benford adherence.[31]

The MC simulation is based on two assumptions:[32]

Assumption 1: $HITOT = X = (X_1, X_2, \dots X_n)$ is a vector of log normally distributed random variables $log_{10}(X) \sim N(\mu_n, \sigma_n^2)$, whose mean and variance are given by μ_n and σ_n^2, respectively.

Assumption 2: $Y = (Y_1, Y_2, \dots Y_n)$ is a vector of log normally distributed variables $log_{10}(Y) \sim N(\mu_n, \sigma_n^2)$, whose mean and variance are given by μ_n and σ_n^2, respectively.

[31] We apply the MC simulation only to the HITOT variable because the calculations are so computationally intensive.

[32] We base Assumption 1 on the demonstrable tendency of income data to be asymptotically log normally distributed (Clementi & Gallegati, 2005).

The simulation itself can be characterized by the following repetitive process:

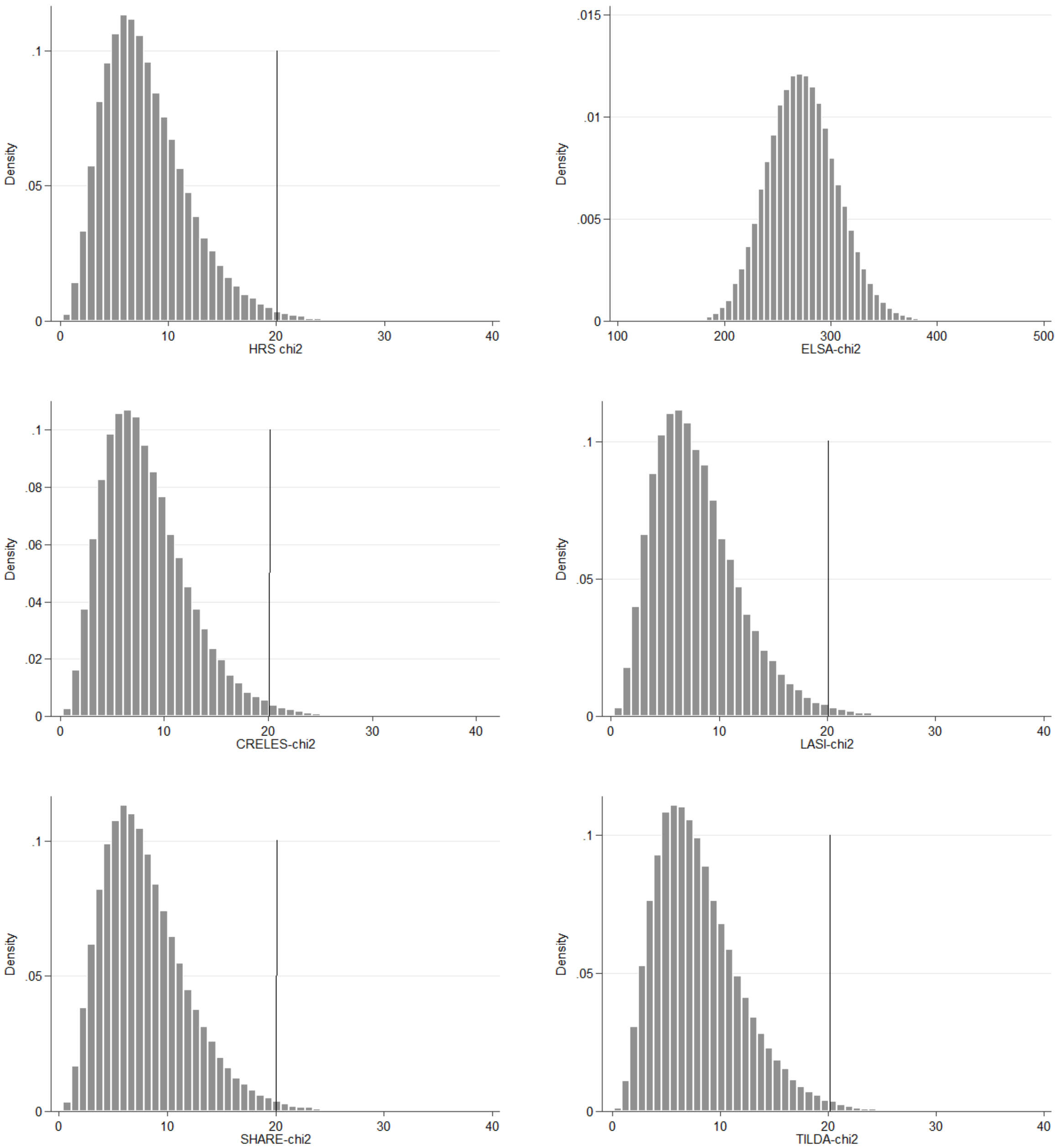
MC simulation pseudo-code:	
(1)	Create a vector T, with i elements
(2)	Calculate $E(X) = \mu_n$ and $Var(X) = \sigma_n^2$
(3)	Create new random vector Y^i containing n random variables
(4)	Create vector Z^i, with ${Z_n}^i = 10^{y_n}$
(5)	Conduct a Benford analysis and replace the *i*th element of *T* with the computed value for the χ^2 test statistic
(6)	Repeat steps 3 to 5 *i* times

As Forman (2010) shows, log normally distributed random variables obey Benford's law if the probability density function fulfils certain criteria that are dependent on the values of the first two moments. Hence, the distribution of the simulated chi-square values for each data set should contain information about whether the data set should generally obey Benford's law.

In Figure 5.6, we show the distribution of *T* calculated independently for each data set when $i = 100.000$, with a vertical line indicating the critical value for the 99% confidence interval of the chi-squared distribution. Except for the ELSA data set, most simulated chi-square values are below the critical value,[33] which reinforces the assumption that the Benford deviations reported in the Results section are, in fact, the consequence of a reliability issue in the income data. It must nevertheless be noted that this finding is unconformable for ELSA, which impedes the interpretation of reliability for that particular data set.

[33]The 99th percentile values are 20.0644 (HRS), 354.0885 (ELSA), 20.0874 (CRELES), 20.0644 (LASI), 20.1608 (SHARE), and 20.2041 (TILDA).

Figure 5.6: *MC simulation results for each data set*



Notes: The figure graphs the simulated χ^2 values for each data set, with the vertical lines indicating the critical value of the chi-squared distribution for the 99% confidence interval.

5.5 Conclusions

According to this analysis, Benford's law seems to hold for income information from both aggregate (GDP) and survey data. Nevertheless, although there is no significant difference between the Benford first digit distribution and the first digit pattern in the World Bank GDP data, HRS data adherence to Benford's law is less strict when measured statistically. Given income data's general tendency to conform with Benford's law, this finding strongly suggests that the income data from our surveys are subject to observational errors and should be treated with caution in terms of policy and econometric analysis. Such caution need not apply, however, to all data sets and variables analyzed. Whereas respondent (or spousal) personal earnings show poor reliability in all data sets, the household level income measure performs comparably well in the HRS (America), SHARE (Europe), and LASI (India) data. [34]

To some extent, this finding contradicts Judge and Schechter's (2009) conclusion that survey data from developing countries is generally less reliable. Rather, our analysis of harmonized data sets suggests that it is less a matter of origin and preparation than the framing of the question that determines the degree of reliability. In particular, respondents tend to make less reliable statements about their individual income than about household income.[35] The overall robustness of these findings in all but the ELSA data set is confirmed by a simulation implemented to determine whether Benford deviations do indeed indicate a reliability problem or are merely the result of a non-Benford distributed variable.

Our study results have two important implications for both econometric and policy analyses of survey data: First, our evidence of crucial reliability problems in data measured on the individual level strongly suggests the use of household level data whenever possible,

[34]However, since there exists no clear cutoff for each statistical measure which indicates a change from a good to a rather poor reliability, it is rather the relative discrepancy in the test statistics for different data sets which matters in the assessment of the quality. For instance, while the SHARE data shows a chi-square value of around 90 for the analysis of HITOT, the tests statistic is 7 times higher for the CRELES data. Given this discrepancy – together with the results of the Kolmogorow-Smirnov as well as the distance measure– the SHARE shows a rather high quality compared to the CRELES data.

[35]Because all the data sets employ detailed income measures, the differences in these measures' reliability do not stem from rounding errors.

Second, contrary to the accepted wisdom that survey data from developed countries perform better than those from developing countries, our results suggest that any data analyzed should be evaluated for quality using a Benford analysis confirmed via simulation. As illustrated by our simulation results, this technique provides valuable insights on the variable of interest's general behavior, thereby improving the quality assessment of the underlying data. Nonetheless, because we only analyze income's Benford adherence in a particular family of surveys, our results – whether for different variables or for different survey data sets – are not generalizable to the broader body of survey research. Future research might thus make use of the Benford analysis to evaluate additional variables and data from other surveys to enhance knowledge about the interrelation of data framing, origin, and reliability.

6 Conclusions

The underlying thesis analyzed the relationship between policy interventions as well as different socio-economic factors like unemployment and health. The main purpose was to work out how different policy measures and individual economic shifts affect the health outcomes of individuals. Knowledge about the health effects of such interventions is of particular importance for individuals as well as for policy makers in order to assess the changes in health outcomes due to probable changes in the economic environment as well as the structural framework of the society.

Therefore, in chapter two, it has been shown how unemployment impacts risky behaviors. Since unemployment can be seen as a strong economic change, the results of this part of the thesis reveal evidence about the general health impact of negative economic alterations. In contrast to previous literature, the results yield no evidence for a negative impact of unemployment on risky behaviors. In particular, by using data from the GSOEP four different measures for risky behaviors has been analyzed (diet, alcohol consumption, physical activity, and smoking). While the analysis shows no evidence for an adjustment of the consumption of addictive goods after becoming unemployed, individuals alter their food consumption as well as their level of physical activity in a positive way. This finding can probably be traced back on a shift in the opportunity costs of time. Since the preparation of healthy food and a high level of sports consumption is typically time consuming, but generally not expensive, additional time can be used to foster such activities.

The findings of chapter two mainly contradict the results of previous studies. This can be primarily explained by the incorporation of plant closure as an exogenous reason for being unemployed. This finding is even strengthened when using other reasons for unemployment in the analysis. In these cases the coefficients are biased. This implies that it is – to some extent - rather the choice of the amount of healthy lifestyles which determines unemployment than the other way round.

In chapter three, it has been analyzed to what extent preschool care impacts the well-being of children and adolescents. Therefore, data from the KIGGS survey has been exploited. Interestingly, the effects differ among the socio-economic as well as the cultural background of the children. While children from non-migrant families show a lower level of well-being if they have experienced preschool child care, the opposite is true for children with a migratory background. This relationship may be partially explained by the fact that migrant children may enter school with a lower level of school readiness. However, this explanation is doubtful since the math and language skill results of migrant children yield no evidence for this hypothesis. Moreover, the findings suggest that psychological components like self-esteem can be seen as the main driver for the difference in well-being among migrants and non-migrants. Overall, the results imply that policy interventions should especially target on the promotion of preschool child care for migrants since they are clearly better off in terms of child well-being.

In chapter four, panel data, drawn from the Eurostat database, has been exploited to assess the impact of a nationwide population-based skin cancer screening program in Germany. The results indicate that the program has been effective in terms of a higher diagnosis rate for malignant skin neoplasms. However, there is no prove of a reduction in the mortality rate due to skin cancer. The latter can be mainly explained by the comparatively short time period of the study, which does not allow to assess the long-term effect of the skin cancer screening program. The fact that the skin cancer screening program substantially and significantly increases the diagnosis rate yields strong evidence that well elaborated health care policy programs contribute to an overall increase in the health state of individuals as well as the society, in general.

Since good data quality is essential to draw a correct inference about the information in the underlying data set, appropriate measures to assess the reliability of survey data sets are important. Moreover, a good data reliability is particularly important for health economists as well as policy makers because a poor data quality may lead to a wrong inference and hence to false policy implications. Obviously, such a scenario would prevent the economy from producing a pareto-optimal outcome. For example, if the outcome of the analysis in chapter three would be wrong because of a problem in the data reliability, the policy implication would be wrong as well. This would lead to a wrong incentive for migrant parents. Instead of avoiding the hypothetical negative health outcome they would send their children to preschool care. However, in the long-run, this would lower the health state of the children and therefore produce a non pareto-optimal outcome of the economy.

Therefore, in chapter five, the data reliability from six widely used health and retirement survey has been analyzed. In particular, Benford’s law was used to assess the quality of different income measures of the individual data sets. The main findings strongly suggest that the income data shows observational errors which implies – to some extent - a poor data reliability. However, since income data measured on the household level shows a better performance in terms of reliability than income data measured on the individual level, the findings imply that researchers as well as policy makers should use data measured on the household level for analysis purpose, in general.

In summary, the results of this thesis show that economic circumstances and health policy decisions influence the health outcomes of individuals. Policy interventions should therefore always be analyzed in terms of the direct as well as the indirect effects on the health outcomes of the individuals and the economy as a whole. However, researchers as well as policy makers should be aware of reliability issues that may arise when they make use of survey data to draw an inference in the best possible way.

7 Literature

Aarts, M. J., Lemmens, V. E., Louwman, M. W., Kunst, A. E., & Coebergh, J. W. (2010). Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. *European journal of cancer, 46*(15), 2681–2695.

American Cancer Society (2017). *Cancer Facts & Figures 2017.* Atlanta: American Cancer Society.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton: Princeton University Press.

Arcaya, M., Glymour, M. M., Christakis, N. A., Kawachi, I., Subramanian, S. (2014). Individual and spousal unemployment as predictors of smoking and drinking behavior. *Social Science & Medicine 110*, 89–95.

Ausloos, M., Herteliu, C., & Ileanu, B. (2015). Breakdown of Benford's law for birth data. *Physica A: Statistical Mechanics and Its Applications*, *419*, 736–745.

Baetschmann, G., Staub, K. E., Winkelmann, R. (2015). Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 178*(3), 685–703.

Belsky, J. (2006). Early child care and early child development: Major findings of the NICHD study of early child care. *European Journal of Developmental Psychology*, *3*(1), 95–110.

Berger, A., Hill, T. P., Kaynar, B., & Ridder, A. (2011). Finite-state Markov Chains Obey Benford's Law. *SIAM Journal on Matrix Analysis and Applications*, *32*(3), 665–684.

Black, W. C., Haggstrom, D. A., & Welch, H. G. (2002). All-Cause Mortality in Randomized Trials of Cancer Screening. *Journal of the National Cancer Institute, 94*(3), 167–173.

Blau, D. M. (1999). The effect of pregnancy intention on child development. *The Journal of Human Resources, 34*(4), 786–822.

Bouchery, E. E., Harwood, H. J., Sacks, J. J., Simon, C. J., Brewer, R. D. (2011). Economic costs of excessive alcohol consumption in the U.S., 2006. *American journal of preventive medicine*, *41*(5), 516–524.

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, *27*(3), 281–291.

Breitbart, E. W., Waldmann, A., Nolte, S., Capellaro, M., Greinert, R., Volkmer, B., & Katalinic, A. (2012). Systematic skin cancer screening in Northern Germany. *Journal of the American Academy of Dermatology, 66*(2), 201–211.

Bullinger, M., Brütt, A. L., Erhart, M., & Ravens-Sieberer, U. (2008). Psychometric properties of the KINDL-R questionnaire: Results of the BELLA study. *European Child and Adolescent Psychiatry*, *17*, 125–132.

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly*, *25*(2), 140–165.

Care, E. C., & Development, C. (2002). Early Child Care and Children's Development Prior to School Entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *39*(1), 133–164.

Carli, P., Mannone, F., Giorgi, V. de, Nardini, P., Chiarugi, A., & Giannotti, B. (2003). The problem of false-positive diagnosis in melanoma screening: the impact of dermoscopy. *Melanoma research, 13*(2), 179–182.

Choudhury, K., Volkmer, B., Greinert, R., Christophers, E., & Breitbart, E. W. (2012). Effectiveness of skin cancer screening programmes. *The British journal of dermatology, 167*(s2), 94–98.

Clementi, F., & Gallegati, M. (2005). Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States. In A. Chatterjee, S. Yarlagadda, & B. K. Chakrabarti (Eds.), *Econophysics of Wealth Distributions: Econophys-Kolkata I* (pp. 3–14). Milano: Springer Milan.

Conway, D. I., Brenner, D. R., McMahon, A. D., Macpherson, L. M., Agudo, A., Ahrens, W., et al. (2015). Estimating and explaining the effect of education and income on head and neck cancer risk: INHANCE consortium pooled analysis of 31 case-control studies from 27 countries. *International journal of cancer, 136*(5), 1125–1139.

Crosnoe, R. (2007). Early child care and the school readiness of children from Mexican immigrant families. *International Migration Review*, *41*(1), 152–181.

Cryan, J. R., Sheehan, R., Wiechel, J., & Bandy-Hedden, I. G. (1992). Success outcomes of full-day kindergarten: More positive behavior and increased achievement in the years after. *Early Childhood Research Quarterly*, *7*(2), 187–203.

Datta Gupta, N., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics*, *94*(1–2), 30–43.

Dhuey, E. (2011). Who Benefits From Kindergarten? Evidence From the Introduction of State Subsidization. *Educational Evaluation and Policy Analysis*, *33*(1), 3–22.

Driessen, G. W. J. M. (2004). A large-scale longitudinal study of the utilization and effects of early childhood education and care in The Netherlands. *Early Child Development and Care*, *174*(7–8), 667–689.

Duncan, G. J., & Hill, D. H. (1985). An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data. *Journal of Labor Economics*, *3*(14), 508–532.

Durtschi, C., Hillison, W., & Pacini, C. (2004). The Effective Use of Benford ' s Law to Assist in Detecting Fraud in Accounting Data. *Journal of Forensic Accounting*, *5*(1), 17–34.

Dustmann, C., Raute, A., & Schönberg, U. (2016). *Who benefits from universal child care? Estimating marginal returns to early child care attendance* (Working Paper). Retrieved Februrary, 2016, from https://www.utaschoenberg.com/working-papers.

Dave, D. M., Kelly, I. R. (2012). How does the business cycle affect eating habits? *Social science & medicine*. *74*(2), 254–262.

Deb, P., Gallo, W. T., Ayyagari, P., Fletcher, J. M., & Sindelar, J. L. (2011). The effect of job loss on overweight and drinking. *Journal of Health Economics, 30*(2), 317–327.

Eisemann, N., Waldmann, A., Garbe, C., & Katalinic, A. (2015). Development of a microsimulation of melanoma mortality for evaluating the effectiveness of population-based skin cancer screening. *Medical Decision Making, 35*(2), 243–254.

Eisemann, N., Waldmann, A., Geller, A. C., Weinstock, M. A., Volkmer, B., Greinert, R., et al. (2014). Non-melanoma skin cancer incidence and impact of skin cancer screening on incidence. *The Journal of investigative dermatology, 134*(1), 43–50.

Erhart, M., Ellert, U., Kurth, B.-M., & Ravens-Sieberer, U. (2009). Measuring adolescents' HRQoL via self reports and parent proxy reports: an evaluation of the psychometric properties of both versions of the KINDL-R instrument. *Health and Quality of Life Outcomes, 7*(1), 77.

Ettner, S. L., 1997. Measuring the human cost of a weak economy: does unemployment lead to alcohol abuse. *Social science & medicine, 44*(2), 251–260.

Eurostat (2016). *NUTS - Nomenclature of territorial units for statistics: Overview.* Retrieved May 04, 2016, from http://ec.europa.eu/eurostat/web/nuts/overview.

Evers, A., Lewis, J., & Riedel, B. (2005). Developing child-care provision in England and Germany: problems of governance. *Journal of European Social Policy, 15*(3), 195–209.

Federman, D. G., Kirsner, R. S., & Viola, K. V. (2013). Skin cancer screening and primary prevention: facts and controversies. *Clinics in dermatology, 31*(6), 666–670.

Felfe, C., & Lalive, R. (2014). Does Early Child Care Help or Hurt Children's Development? *IZA Discussion Paper Series*, (8484).

Formann, A. K. (2010). The Newcomb-Benford Law in its relation to some common distributions. *PLoS ONE*, *5*(5).

Freeman, D. G. (1999). A note onEconomic conditions and alcohol problems. *Journal of Health Economics*. *18*(5), 661-670.

Geller, A. C., Greinert, R., Sinclair, C., Weinstock, M. A., Aitken, J., Boniol, M., et al. (2010). A nationwide population-based skin cancer screening in Germany: proceedings of the first meeting of the International Task Force on Skin Cancer Screening and Prevention (September 24 and 25, 2009). *Cancer epidemiology, 34*(3), 355–358.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy, 80*(2), 223–255.

Groves, R. M. (2004). *Survey errors and survey costs* (Vol. 536). John Wiley & Sons.

Havnes, B. T., & Mogstad, M. (2011). No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes. *American Economic Journal: Economic Policy, 3*(2), 97–129.

Hemmerling, A. (2008). *Der Kindergarten als Bildungsinstitution: Hintergründe und Perspektiven*. Springer Science & Business Media.

Hill, T. P. (1995). A Statistical Derivation of the Significant-Digit Law. *Statistical Science*, *10*(4), 354–363.

Hill, T. P. (1995a). Base-invariance implies benford's law. *Proceedings of the American Mathematical Society*, *123*(3), 887–895.

Hill, T. P. (1995b). The Signifilcant Digit Phenomenon. *The American Mathematical Monthly*, *102*(4), 322–327.

Judge, G., & Schechter, L. (2009). Detecting Problems in Survey Data using Benford's Law. *Journal of Human Resources*, *44*(1), 1–24.

Kalman, R. E., & others. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35–45.

Katalinic, A., Eisemann, N., & Waldmann, A. (2015). Skin Cancer Screening in Germany. Documenting Melanoma Incidence and Mortality From 2008 to 2013. *Deutsches Ärzteblatt international, 112*(38), 629–634.

Katalinic, A., Waldmann, A., Weinstock, M. A., Geller, A. C., Eisemann, N., Greinert, R., et al. (2012). Does skin cancer screening save lives? an observational study comparing trends in melanoma mortality in regions with and without screening: An Observational Study Comparing Trends in Melanoma Mortality in Regions With and Without Screening. *Cancer, 118*(21), 5395–5402.

Khan, S., Murray, R. P., Barnes, G. E. (2002). A structural equation model of the effect of poverty and unemployment on alcohol abuse. *Addictive behaviors, 27*(3), 405–423.

Kurth, B. M., Kamtsiuris, P., Hölling, H., Schlaud, M., Dölle, R., Ellert, U., ... & Neuhauser, H. (2008). The challenge of comprehensively mapping children's health in a nation-wide health survey: design of the German KiGGS-Study. *BMC Public health*, *8*(1), 196.

Le Gallo, J. (2004). Space-time analysis of GDP disparities among European regions: A Markov chains approach. *International Regional Science Review*, *27*(2), 138–163.

Leemis, L. M. L., Schmeiser, B. W. S., & Evans, D. L. E. (2000). Survival Distributions Satisfying Benford' s Law. *The American Statistician*, *54*(3), 236–241.

Little, E. G., & Eide, M. J. (2012). Update on the current state of melanoma incidence. *Dermatologic clinics, 30*(3), 355–361.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*(1), 52–66.

Losina, E., Walensky, R. P., Geller, A., Beddingfield, F. C., Wolf, L. L., Gilchrest, B. A., & Freedberg, K. A. (2007). Visual Screening for Malignant Melanoma: A Cost-effectiveness Analysis. *Archives of Dermatological Research, 143*(1), 21–28.

MacKie, R. M., Hauschild, A., & Eggermont, A. M. M. (2009). Epidemiology of invasive cutaneous melanoma. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO, 20*(s6), 7.

Magnuson, K. a., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*(1), 33–51.

Magnuson, K., Lahaie, C., & Waldfogel, J. (2006). Preschool and School Readiness of Children of Immigrants. *Social Science Quarterly*, *87*(5), 1241–1262.

Mark Twain. (n.d.). BrainyQuote.com. Retrieved July 20, 2017, from BrainyQuote.com: https://www.brainyquote.com/quotes/quotes/m/marktwain105716.html

Mellow, W., & Sider, H. (1983). Accuracy of Response in Labor Market Surveys: Evidence and Implications. *Journal of Labor Economics*, *1*(4), 331–344.

Meurs, M. (2006). Decline in pre-school use in post-socialist societies: the case of Bulgaria. *Journal of European Social Policy*, *16*(2), 155–166.

Mir, T. A. (2014). The Benford law behavior of the religious activity data. *Physica A: Statistical Mechanics and Its Applications*, *408*, 1–9.

Montgomery, S. M., Cook, D. G., Bartley, M. L., Wadsworth, M. E. (1998). Unemployment, cigarette smoking, alcohol consumption and body weight in young British men. *The European Journal of Public Health*, *8*(1), 21–27.

Mossakowski, K. N., 2008. Is the duration of poverty and unemployment a risk factor for heavy drinking. *Social science & medicine*, *67*(6), S947–S955.

Network, N. I. of C. H. and H. D. E. C. C. R. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development*, *74*(4), 976–1005.

NICHD Early Child Care Research Network. (2001). Nonmaternal care and family factors in early development: An overview of the NICHD Study of Early Child Care. *Journal of Applied Developmental Psychology*, *22*(5), 457–492.

NICHD Early Child Care ResearchNetwork. (2004). Type of child care and children's development at 54 months. *Early Childhood Research Quarterly*, *19*(2), 203–230.

Nigrini, M. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection* (Vol. 586). John Wiley & Sons.

Nigrini, M. J., & Miller, S. J. (2007). Benford's Law applied to hydrology data - Results and relevance to other geophysical data. *Mathematical Geology*, *39*(5), 469–490.

Obeng, C. S. (2007). Immigrants Families and Childcare Preferences: Do Immigrants' Cultures Influence Their Childcare Decisions? *Early Childhood Education Journal*, *34*(4), 259–264.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, *72*(5), 1534–1553.

Pietronero, L., Tosatti, E., Tosatti, V., & Vespignani, A. (2001). Explaining the uneven distribution of numbers in nature: The laws of Benford and Zipf. *Physica A: Statistical Mechanics and Its Applications*, *293*(1–2), 297–304.

Pinkham, R. S. (1961). On the Distribution of First Significant Digits. *The Annals of Mathematical Statistics, 32*(4), 1223-1230.

Ravens-Sieberer, U., & Bullinger, M. (2000). KINDL R English Manual. *Questionnaire for Measuring Health-Related Quality of Life in Children and Adolescents Revised Version. KINDL, Germany.*

Robert Koch Institut und die Gesellschaft für der epidemiologischen Krebsregister in Deutschland e. V. (2015). *Krebs in Deutschland 2011/2012.* Retrieved May 04, 2016, from http://www.gekid.de/Doc/krebs_in_deutschland_2015.pdf.

Sabates, R., & Feinstein, L. (2006). The role of education in the uptake of preventative health care: the case of cervical screening in Britain. *Social science & medicine (1982), 62*(12), 2998–3010.

Sachs, L. (2004). *Angewandte Statistik*. Berlin Heidelberg: Springer.

Sandron, F. (2002). Les populations suivent-elles la loi des nombres anomaux? *Population*, *57*(4), 761–767.

Saris, W. E., & Gallhofer, I. N. (2014). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. John Wiley & Sons.

Saris, W. E., & Revilla, M. (2016). Correction for Measurement Errors in Survey Research: Necessary and Possible. *Social Indicators Research*, *127*(3), 1005–1020.

Scarborough, P., Bhatnagar, P., Wickramasinghe, K. K., Allender, S., Foster, C., Rayner, M., 2011. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs. *Journal of public health*, *33*(4), 527–535.

Schlack, R., Hölling, H., & Kurth, B. M. (2007). Inanspruchnahme außerfamiliärer vorschulischer Kindertagesbetreuung und Einfluss auf Merkmale psychischer Gesundheit bei Kindern. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, 50*(10), 1249-1258.

Schmitz, H. (2011). Why are the unemployed in worse health? The causal effect of unemployment on health. *Labour Economics, 18*(1), 71–78.

Schober, P. S., & Schmitt, C. (2017). Day-care availability, maternal employment, and satisfaction of parents: Evidence from cultural and policy variations in Germany. *Journal of European Social Policy*.

Smirnov, N. (1948). Table for Estimating the Goodness of Fit of Empirical Distributions. *The annals of mathematical statistics, 19(2),* 279–281.

Spieß, C. K. (2012). Betreuungsgeld widerspricht den Zielen nachhaltiger Familienpolitik. *DIW-Wochenbericht*, *79*(24), 24-24.

Spiess, C. K., Büchel, F., & Wagner, G. G. (2003). Children's school placement in germany: Does Kindergarten attendance matter? *Early Childhood Research Quarterly*, *18*(2), 255–270.

Stang, A., Garbe, C., Autier, P., Jöckel, K.-H., Autier, P., & Jockel, K.-H. (2016). The many unanswered questions related to the German skin cancer screening programme. *European journal of cancer, 64*, 83–88. Retrieved October 21, 2016.

Stang, A., & Jöckel, K.-H. (2016). Does skin cancer screening save lives? A detailed analysis of mortality time trends in Schleswig-Holstein and Germany. *Cancer, 122*(3), 432–437.

Statistisches Bundesamt (2012). *Kindertagesbetreuung in Deutschland 2012*. Retrieved February, 2016, from https://www.destatis.de/DE/PresseService/Presse/Pressekonferenzen/2012/kindertagesbetreuung/begleitmaterial_PDF.pdf?__blob=publicationFile.

Tödter, K. H. (2009). Benford's law as an indicator of fraud in economics. *German Economic Review*, *10*(3), 339–351.

Turney, K., & Kao, G. (2009). Pre-kindergarten child care and behavioral outcomes among children of immigrants. *Early Childhood Research Quarterly, 24*(4), 432–444.

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., & Vandergrift, N. (2010). Do Effects of early child care extend to age 15 years? Results from the nichd study of early child care and youth development. *Child Development, 81*(3), 737–756.

Veit, C., Lüken, F., & Melsheimer, O. (2015). *Evaluation der Screeninguntersuchungen auf Hautkrebs gemäß Krebsfrüherkennungs-Richtlinie des Gemeinsamen Bundesausschusses: Abschlussbericht 2009 - 2010.* Retrieved May 04, 2016, from BQS Institut für Qualität und Patientensicherheit: https://www.g-ba.de/downloads/17-98-3907/2015-03-11_BQS_HKS-Abschlussbericht-2009-2010.pdf.

Votruba-Drzal, E., Li-Grining, C. P., & Maldonado-Carreo, C. (2008). A Developmental Perspective on Full- Versus Part-Day Kindergarten and Childrens Academic Trajectories Through Fifth Grade. *Child Development, 79(4),* 957-978.

Waldmann, A., Nolte, S., Weinstock, M. A., Breitbart, E. W., Eisemann, N., Geller, A. C., et al. (2012). Skin cancer screening participation and impact on melanoma incidence in Germany--an observational study on incidence trends in regions with and without population-based screening. *British journal of cancer, 106*(5), 970–974.

Wallace, W. A. (2002). Assessing the quality of data used for benchmarking and decision-making. *The Journal of Government Financial Management*, *51*(3), 16–22.

Welch, H. G., & Black, W. C. (2010). Overdiagnosis in cancer. *Journal of the National Cancer Institute, 102*(9), 605–613.

Welch, H. G., Woloshin, S., & Schwartz, L. M. (2005). Skin biopsy rates and incidence of melanoma: population based ecological study. *British Medical Journal, 331*(7515), 481–484.

Winkler, J., & Stolzenberg, H. (1999). Der Sozialschichtindex im Bundes-Gesundheitssurvey. *Gesundheitswesen, 61(2),* 178-183.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). Mason, OH: South-Western Cengage Learning.

World Health Organization (2016). *Skin cancers: How common is skin cancer?* Retrieved May 04, 2016, from http://www.who.int/uv/faq/skincancer/en/index1.html.

8 Appendix

Table 8.1: INTERACTION EFFECTS

	Diet	Alcohol	Physical activity	Smoking
	(1)	(2)	(3)	(4)
Age				
exogenous layoff	-0.3142	-0.5197	-0.1403	**5.4557*** **
exogenous layoff*age	0.0009	0.0055	-0.0008	**-0.1134*** **
endogenous layoff	0.2379	-0.3241	**0.3197*** **	**-1.9161*** **
endogenous layoff*age	-0.0060	-0.0001	**-0.0058***	**0.0647*** **
Age	**-0.0269*** **	**-0.0256** **	**-0.0392*** **	**-0.0231*** **
Marital Status				
exogenous layoff	-0.1132	-0.4317	**-0.2274***	0.5854
exogenous layoff*married	-0.0760	-0.0502	**0.2422*** **	**-2.0525*** **
endogenous layoff	-0.0051	**-0.3824*** **	0.0325	**0.6022** **
endogenous layoff*married	-0.2241	0.2293	0.0959	0.1011
Married	0.0092	0.1004	**0.1149***	0.0364
Job type				
exogenous layoff (blue collar sample)	-0.1152	**-1.6013** **	**-0.3732*** **	-0.0118
endogenous layoff (blue collar sample)	**-0.3473*** **	-0.2095	0.1035*	**0.7084** **
exogenous layoff (white collar sample)	-0.2818	0.1906	0.0040	**0.9828** **
endogenous layoff (white collar sample)	0.0032	**-0.3819*** **	0.1188***	**0.5472*** **
Household Income				
exogenous layoff	-.4091782	**-1.201523** **	**-.6386494*** **	-.1919428
exogenous layoff* HH_Income(t)	.0000383	**.0003451** **	**.0001935*** **	.000335
endogenous layoff	-.0110409	-.1829726	-.0287517	**1.472201***
Endogenous layoff* HH_Income(t)	0.000	-.000057	**.0000525** **	**-.0003378***
HH_Income(t)	**-.0000255***	.0000211	**-.0000328** **	-.0000509

Notes: All regressions contain similar control variables as the main results in table 2 and 3. Columns 1-3 are estimated using BUC, column 4 uses OLS. Robust standard errors clustered on the individual level in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01

Table 8.2: DIFFERENCE-IN-DIFFERENCE-IN-DIFFERENCE APPROACH

Germany		
	Malignant skin neoplasms ICD–10 code C43_C44	Neoplasms ICD–10 code C00_D48 (excluding ICD–10 code C43_C44)
t<2008	x_G	y_G
t>2008	$x_G´$	$y_G´$
Difference	$x_G´ - x_G$	$y_G´ - y_G$
Difference-difference	$\Delta\varphi_G = (x_G´ - x_G) - (y_G´ - y_G)$	
Europe		
t<2008	x_E	y_E
t>2008	$x_E´$	$y_E´$
Difference	$x_E´ - x_E$	$y_E´ - y_E$
Difference-difference	$\Delta\varphi_E = (x_E´ - x_E) - (y_E´ - y_E)$	
Difference-difference-difference	$\Delta\varphi_G - \Delta\varphi_E$	