SOEP The German Socio-Economic Panel study

946 2017

SOEPpapers
on Multidisciplinary Panel Data Research

SOEP – The German Socio-Economic Panel study at DIW Berlin 946-2017

Commuting and Sickness Absence

Laszlo Goerke and Olga Lorenz

DIW SOEP

ISSN: 1864-6689 (online)

German Socio-Economic Panel (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: soeppapers@diw.de

Commuting and Sickness Absence*

Laszlo Goerke
IAAEU - Trier University +, IZA, Bonn and CESifo, München

+ Institute for Labour Law and Industrial Relations in the European Union, Campus II
D – 54286 Trier, Germany
E-mail: goerke(at)iaaeu.de

Olga Lorenz
IAAEU - Trier University ++

++ Institute for Labour Law and Industrial Relations in the European Union, Campus II
D – 54286 Trier, Germany
E-mail: lorenz(at)iaaeu.de

November 28, 2017

Abstract: We investigate the causal effect of commuting on sickness absence from work using German panel data. To address reverse causation, we use changes in commuting distance for employees who stay with the same employer and who have the same residence during the period of observation. In contrast to previous papers, we do not observe that commuting distances are associated with higher sickness absence, in general. Only employees who commute long distances are absent about 20% more than employees with no commutes. We explore various explanations for the effect of long distance commutes to work and can find no evidence that it is due to working hours mismatch, lower work effort, reduced leisure time or differences in health status.

Keywords: sickness absence, absenteeism, commuting, health, labour supply

JEL: I10, J22, R2, R40

---

* We thank Adrian Chadi, Marco de Pinto, Sabrina Jeworrek, Mario Mechtel and the participants of the IZA Workshop on Health and Labor Markets in Bonn, the 2016 Annual Conference of the European Society for Population Economics (ESPE) in Berlin, the 2016 CESifo Area Conference on Employment and Social Protection in Munich, the 7th Economic Workshop in Trier, the 2015 Colloquium on Personnel Economics (COPE) in Vienna, the 2015 Spring Meeting of Young Economists (SMYE) in Ghent, the 2015 Annual Conference of the Verein für Socialpolitik (VfS) in Münster, the Lüneburg Workshop in Economics and a seminar at U Trier for helpful comments and suggestions.

1. Introduction

Each day millions of employees commute between home and work. The frequency of commuting and the average duration of commutes have risen in the last decades (Kirby and LeSage, 2009). According to the German Federal Statistical Office, the percentage of commutes which took less than 10 minutes to work is declining, while the share of those people who commute 30 to 60 minutes to work has risen from 17% in 1996 to 23% in 2012. A similar picture emerges regarding commuting distance (Federal Statistical Office, 2013). This trend is not unique to Germany. In the UK, for example, commuting times have increased from 48 to 54 minutes per day, the average commuting trip length has increased by 1.3 kilometres between the mid-nineties and 2012 to reach 14.5 kilometres (National Travel Survey 2012). In Spain and Italy, commuting times have increased from 31 to 34 minutes and 22 to 35 minutes, respectively, over the period 1997 – 2006, according to the European Survey on Working Conditions (EWCS). These facts show that commuting is an important and growing component of daily life.

On the one hand, commuting may be viewed positively as it increases the density of labour markets and, hence, allows for better matches between jobs and individuals. Moreover, commuting enables employees to live in places where there are no adequate jobs, without forsaking their income. On the other hand, commuting is usually argued to be problematic from an environmental point of view and to be detrimental to the health of employees.[1]

The aim of this paper is to empirically examine the impact of commuting distance on the number of sickness absence days. If commuting is negatively related to health, employees who commute are more likely to be absent from work (Zenou, 2002). In addition, the gain from absence in terms of hours which can be used for other purposes than work, such as recuperation, is likely to be higher for individuals who commute. However, from a theoretical vantage point, the effect of commuting on sickness absence may also be negative. Individuals would not choose to have a longer commute unless they were compensated for it, for example, in the form of improved job characteristics (including pay) or better housing prospects (Stutzer and Frey, 2008). Hence, individuals who commute may have better, more motivating jobs and be able to achieve a better work-life balance. Furthermore, willingness to travel to work may be associated

[1] Several predominantly U.S. studies have found that work commutes induce stress due to their unpredictability and the perceived loss of control (Gottholmseder et al., 2009). Furthermore, commuting has been shown to be associated with increased heart rate and blood pressure (Novaco et al., 1979; Schaeffer et al., 1988). Further, commuting translates into shorter sleeping times and sleep disorders (Costa et al., 1988; Walsleben et al., 1999; Hansson et al., 2011), a lower social capital and participation (Mattisson et al., 2015), which has in turn been associated with health outcomes (Putnam, 2000; Lindström, 2004; Besser et al., 2008), negative mood (Gulian et al., 1989), emotional arousal (Hennessy and Wiesenthal, 1997), lower well-being and life satisfaction (Stutzer and Frey, 2008; Roberts et al., 2011; Olsson et al., 2013) as well as higher levels of workplace aggression (Hennessy, 2008), poor concentration levels (Matthews et al., 1991) and a higher risk of mortality (Sandow et al., 2014).

positively with work effort. Accordingly, the net effect of the commute to work on sickness absence is theoretically ambiguous.

An understanding of this relationship is important for a number of reasons: First, if absence affects productivity and profitability, a firm's employment and location decisions may be influenced by its (prospective) employees' commuting behaviour. Second, since absence can cause externalities, for example, if absent employees are entitled to sick pay, health policy requires knowledge of the relationship between commuting and absence from work. Third, policies which alter the mobility of the workforce and the integration of economic regions need to take into account the effects of commuting on absence. Finally, an analysis of the relationship between the work commute and sickness absence enhances our knowledge of the economic costs of commuting.

Despite this imminent importance, few studies have analysed the effect of commuting on sickness absence thus far. Early contributions, surveyed by Kluger (1998), are often based on cross-sectional, firm-specific data and tend to find positive correlations. More recently, panel data have been utilised. Magee et al. (2011), for example, employ data from the Australian household, income, and labour dynamics data set (HILDA) for the years 2005 and 2008. They find a positive correlation between commuting time and absence. Künn-Nelen (2016) uses the British Household Panel Survey (BHPS) data for 1991 to 2008 and detects no robust correlation between commuting time and being absent.[2] Hassink and Fernandez (2018) focus on the relocation of a food processing plant in the United States as a source of exogenous variation in commuting time. While there is no effect on the incidence of monthly absence for the entire sample of about 180 workers, the authors observe a positive impact of commuting for employees characterised by low work morale. The study most relevant to our analysis is that by van Ommeren and Gutiérrez-i-Puigarnau (2011). Using data from the German Socio-Economic Panel (SOEP), they examine the impact of changes in commuting distance on workers' productivity as manifested through the number of sickness absence days. van Ommeren and Gutiérrez-i-Puigarnau (2011) find that commuting distance induces illness-related absence, which they interpret as shirking behaviour by employees, with an elasticity of about 0.07 to 0.09.

Against this background, our paper makes a number of contributions: First, by analysing the impact of employer-induced changes in commuting distance on absence and by using a fixed-effects (FE) framework that includes important predictors of sickness absence and – novel to the literature – measures of compensation for commuting in the labour or housing market, e.g. indicators of satisfaction with the job, leisure and housing situation, we present a more integrated

[2] Moreover, there are some empirical analyses of absence behaviour which include an indicator of commuting as covariate, without looking at the relationship in detail. Allen (1981) and De Paola (2010), for example, report no correlation.

approach for explaining the relationship between sickness absence and commuting. We hence, provide a more precise analysis of the effects of the work commute. Second, we are able to ascertain whether absence behaviour of employees who do not commute differs from that of employees who make short, middle or long distance commutes. Third, we allow for discontinuities in the effect of commuting on absence. Finally, we investigate potential channels determining the relationship between commuting and sickness absence. This enables us to obtain a fuller picture of how commuting affects behavioural (lifestyle) factors that, in turn, influence absence behaviour.

Our empirical analyses are conducted using data from the German Socio-Economic Panel (SOEP) for the period 2002 – 2011. First, an ordinary least square and a negative binomial model are estimated. Second, fixed-effects models are used to remove time invariant unobserved heterogeneity. One major issue in the empirical study of the effect of commuting on sickness absence is reverse causation. In order to address this issue, we employ an identification strategy which is based on employer-induced changes in commuting distance, because these changes are exogenous from an employee's perspective. In particular, we look at employees who stay with the same employer and have the same place of residence during the period of observation.[3] We show that employees who commute middle (between 25 and 49 kilometres) or long distances (more than 50 kilometres), are absent more often than comparable employees who do not commute or who travel shorter distances. In particular, the average number of absence days amounts to 10.36 days for the entire sample, while long (middle) distance commuters exhibit 11.86 (10.43) absence days. These descriptive findings are confirmed when accounting for observable characteristics in a pooled sample as well as in the panel structure of our data. In contrast to previous papers, we do not observe that shorter commuting distances are associated with higher sickness absence. Moreover, we find no evidence that the effect of commutes on absence from work is due to working hours mismatch (respectively, a lower work effort or shirking), reduced leisure time or differences in health status.

The remainder of this paper is organized as follows: Section 2 describes the data and variables. Section 3 focuses on our identification strategy and outlines the econometric method. Section 4 reports the results, including several robustness checks and the analysis of mechanisms through which commuting might affect absence behaviour. Section 5 concludes the study.

[3] Similar identification strategies have been used by other authors, as well, looking at different issues. Zax (1991) and Zax and Kain (1996) analyse job and residential moving behaviour. Gutiérrez-i-Puigarnau and van Ommeren (2010) investigate labour supply patterns. van Ommeren and Gutiérrez-i-Puigarnau (2011) and Hassink and Fernandez (2018) examine the impact of commuting on workers' productivity and effort, proxied by their absence behaviour, and Roberts et al. (2011) consider the effect of an exogenous change in commuting time on psychological health in a robustness check. Finally, Carta and De Philippis (2015) investigate the impact of commuting on the labour supply of couples.

2. Data and Variables

The current study is based on information from the German Socio-Economic Panel (SOEP) for the years 2002 – 2011. The SOEP is a longitudinal, nationally representative survey of private households in Germany. Currently around 30,000 people in approximately 15,000 households participate in the survey. The SOEP includes rich information on labour market status, wealth, income and standard of living, health and life satisfaction as well as on family life and socio-economic variables.[4] To the best of our knowledge, the SOEP is the only person-level dataset for Germany providing detailed information on both absence from work and commuting distance.

The SOEP provides a self-reported measure of the annual number of days absent from work due to sickness in the previous year. The exact question reads as follows: "*How many days were you unable to work in 20XX due to illness? Please state the total number of days, not just the number of days for which you had an official note from your doctor: (a) None (b) A total of X days.*" The advantage of this question is that it provides information on the total number of absence days, and not only with respect to those, for example, for which a medical certificate is required.[5] However, there is no data in the SOEP on the annual number and the duration of specific sickness spells. Therefore, in the following multivariate regressions we use 'days absent', i.e. the total number of days the employee has been absent during the previous year, as our dependent variable. Moreover, the SOEP provides information on whether an employee experienced one or more absence spells exceeding 30 days. We use this evidence and a dummy variable measuring the incidence of absence in robustness checks (cf. Section 4.4).

The SOEP, furthermore, requires respondents to report on commuting distance. The question reads: "*How far (in kilometres) is it from where you live to where you work? (a) X km (b) Difficult to say, location of workplace varies (c) Workplace and home are in the same building/same property.*" We define all respondents for whom either part (c) of the question applies or who state that the distance D between home and workplace is less than ten kilometres (part (a) of the question) as non-commuters (i.e., $0 \leq D < 10$ km). All respondents who travel ten or more kilometres to work are defined as commuters. Of these respondents, those who travel to work over ten kilometres and less than 25 kilometres are short distance (i.e., $10 \leq D < 25$ km), those who travel over 25 kilometres and less than 50 kilometres are middle distance (i.e., $25 \leq D$

[4] Further information about the SOEP is provided by Wagner et al. (2007) and can also be found at: http://www.diw.de/english/soep/29012.html. We use the SOEP long v29 dataset.

[5] In Germany, dependent employees with a minimum tenure of four weeks can basically take sick leave without a durational restriction. From the fourth day of the sickness spell onwards, employees are legally required to present an official note by a doctor. During the first six weeks of an absence period, the employer has to continue to pay wages. Once an employee's absence period exceeds six weeks, the mandatory health insurance will cover the cost of sick pay which drops to at most 90% of the net wage.

< 50 km) and those who cover 50 or more kilometres are long distance commuters.[6] This approach allows for qualitatively different effects of, for example, shorter and longer commuting distances on absence. Moreover, it is not sensitive to minor reporting errors. As robustness check we have experimented with several functional forms and categorisations for commuting distance (cf. Section 4.4). Finally, those who report working in different places (part (b) of the question) are excluded from the analysis as it is difficult to determine their actual commuting distance.

It is worth mentioning that the SOEP provides direct information about commuting time and commuting mode only in 2003. In other years, it is possible to imprecisely ascertain commuting time by calculating the difference between daily working hours, including travel time to and from work, and the usual daily working hours. We use this information in a robustness check (see footnote 16).

The choice of the other explanatory variables is informed by the literature on the determinants of sickness absence as well as on commuting. Correlates or determinants of absence can be categorized as follows (e.g. Block et al. 2014; Livanos and Zangelides, 2013; Ziebarth and Karlsson, 2010; Frick and Malo, 2008; Dionne and Dostie, 2007). The first group contains variables on personal characteristics such as gender, marital status, children, age, current health status as well as educational attainment. The second set incorporates variables on job-related aspects: Tenure, working time, type of employment contract (temporary), occupational position, size of company, sector information, industry dummies, and income. We also include region as well as year dummies. Furthermore, studies on commuting (e.g. Costa et al., 1988; Stutzer and Frey, 2008; Lyons and Chatterjee, 2008) suggest that compensation for commuting may be provided in the housing or labour market. Hence, we also include indicators of satisfaction with dwelling and the amount of leisure time, job satisfaction and a household crowding index as explanatory variables.

Our estimation sample consists of 18 to 65-year-old individuals in paid employment and does not include self-employed (cf. Roberts et al.; 2011). As information on sickness absence is provided for the year prior to the interview and commuting distance is measured at the interview date, we ensure that commuting distance applies to the same year t for which sickness absence days are reported by using the information on absence reported in the interview in the subsequent year $t + 1$. Furthermore, to affirm that information on commuting distance and sickness absence refers to the same employer we additionally confine our sample to workers who have a minimum tenure of two years.[7] As part of our identification strategy explained in the next section we focus

[6] No standard definition of commuting is used internationally or in Germany. We build our categories in line with definition used by the Federal Statistical Office of Germany.

[7] Since we estimate worker fixed-effect specifications, there is unlikely to be a selection bias because the FE specification (at least in the FE OLS specification) controls for worker-specific time-invariant heterogeneity.

on workers who stay with the same employer and have the same residence. Our working sample then consists of 6,459 individuals with 31,567 observations.

Tables A.1 and A.2 in Appendix A show our variable definitions and a complete list of covariates and descriptive statistics.

3. Identification Strategy and Econometric Methods

3.1 Identification Strategy

A worker's commuting distance is often self-chosen and may, thus, be affected by the endogenously determined residence and employer. To account for the endogeneity in the absence-commute relationship, we focus our analyses on a subset of individuals who experience a (presumably exogenously induced) change in their commuting distance. Therefore, we stipulate that an employee changes neither employer nor household location during the period of observation. A variation in commuting distance will, thus, only occur if a firm alters its location.[8] Thus, the change in commuting distance will be employer-driven and can be viewed as exogenous to the employee.[9] Such changes in workplace location due to firm relocation have been shown to be quite common (Gutiérrez-i-Puigarnau and van Ommeren, 2010; Gutiérrez-i-Puigarnau et al., 2014). For example, about 16.5% of firms in Germany are involved in relocation decisions each year (Federal Statistical Office, 2008). Using this approach, in our sample the average transition probability for the change in the categorical commuting distance is about 10%. We will explicitly address the potential bias of this selection by comparing results of different samples (cf. Section 4.4).

3.2 Models for Cross-Sectional Data

Since absence days can only take on non-negative integer values we estimate a negative binomial model with heteroskedasticity robust standard errors, which is also a convenient way for dealing with overdispersed data, such as we are examining (overdispersion parameter: α = 2.73). Additionally, we make use of an OLS regression model with heteroskedasticity robust standard errors. This approach is feasible since we need not get the functional form perfectly right to obtain valid estimates of the average partial effects. The idea for the empirical test is captured in the following regression equation:

[8] It is important to note that in the data available there is no information on whether the worker's firm relocated or not. So it is not possible to distinguish between true changes (because of firm relocation) and misreporting. Since we are treating commuting distance as a categorical variable, our results are not sensitive to minor reporting errors. Hence, the downward bias in our estimate is likely to be small. We additionally address this problem by excluding observations referring to absolute changes in commuting distance smaller than two kilometres (cf. Section 4.4).

[9] A similar approach is used by van Ommeren and Gutiérrez-i-Puigarnau (2011) who, however, use combinations of worker, residence and employer fixed effects while keeping everyone in the sample.

$$A_i = \alpha + \beta D_i + \gamma X_i + \theta T_i + \delta R_i + \varepsilon_i. \tag{1}$$

A_i equals the total number of days absent from work for individual i. D_i is an indicator for commuting distance and X_i represents a vector of independent variables (e.g. relating to personal and job characteristics and compensation for commuting). In order to capture region and time specific effects we also consider region (R_i) and year dummies (T_i). β, γ, θ and δ are coefficients, and ε_i denotes the error term. Our main interest lies in β. The pooled estimators identify the effect of commuting on the reported number of days absent, based on the variation in these variables between people and for each individual over time. It is assumed that unobserved characteristics, as well as measurement errors, are captured in the error term of the estimation.

3.3 Models for Panel Data

Additionally, we assess the impact of a change in commuting distance, respectively, a change in commuting distance categories on a change in the outcome variable using fixed-effects models because causal inference is better supported using panel data, rather than cross-sectional data (Wooldridge, 2010). Accordingly, we eliminate the risk that time-invariant variables confound the relationship between commuting and sickness absence.

Since our dependent variable is a count variable, we employ a conditional fixed-effects negative binomial regression model as a benchmark (see Hilbe; 2007). Conditional estimation of the fixed-effects model is obtained using maximisation of the log likelihood conditional on the sum of the number of counts during the period during which the individual is observed. Although this method is used frequently, it has been criticised as not being a true fixed-effects approach (Allison, 2009; Allison and Waterman, 2002). A complication is that the conditional fixed-effects negative binomial model is based on a regression decomposition of the overdispersion parameter, rather than the usual regression decomposition of the mean. As a result, the model only removes individual fixed-effects equal to the logarithm of the overdispersion parameter, implying that the conditional fixed-effects specification does not control for all stable covariates. For this reason we additionally revert to the fixed-effects OLS regression which controls for all of its stable predictors.[10] Furthermore, a fixed-effects OLS regression is less contingent on distributional assumptions and easier to interpret than the alternatives. The basic model specification can be denoted by:

$$A_{it} = \beta D_{it} + \gamma X_{it} + \iota_i + \mu_t + \varepsilon_{it}, \tag{2}$$

[10] Given our identification strategy, gender and regional dummies 'drop out' of the fixed-effects OLS specification, while all other controls vary over time.

where A_{it} is a measure of the number of days absent for a worker i in year t, D_{it} is an indicator for commuting distance, X_{it} are a set of conditioning variables, β and γ refer to parameters to be estimated. μ_t are defined as year fixed-effects and ι_i are individual fixed-effects.

4. Results

4.1 Sample Description

Table 1 reports the associations between commuting distance, the number of days absent and the incidence of absence. The average number of days lost through sickness absence amounts to 10.36 days. The standard deviation of 24.7 days indicates that there is a lot of cross-sectional variation. The distribution of absence days is heavily skewed with a mass point at zero. The full distribution of sickness absence days is depicted in Appendix A, Figure A.1.

About half of the individuals in our dataset (54%) are short, middle or long distance commuters. The annual number of days absent increases by about two days, as one-way commute distance increases from under 10 kilometres (non-commuter) to over 50 kilometres (long distance commuter). Those workers who commute long distances have on average 11.9 absence days. The incidence rate is also higher. Approximately 70% of long distance commuters have stayed home sick at least once in the last 12 months, whereas only 63% of non-commuters did so. Hence, the descriptive evidence suggests that being a commuter is associated with a higher incidence of absence and more absence days per year.

Table 1
Descriptive statistics for full sample and for commuter categories. Source: SOEP 2002 – 2011.

	Full sample				Non-Commuter		Short distance commuter		Middle distance commuter		Long distance commuter	
	Mean	SD	Min	Max	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Absence:												
# of days	10.36	24.70	0	365	9.96	24.31	10.59	25.12	10.43	23.72	11.86	27.62
Incidence	0.65	0.47	0	1	0.63	0.48	0.65	0.47	0.67	0.46	0.70	0.45
N	31,567				14,113		10,435		5,129		1,890	
%	100%				46%		33%		16%		5%	

Notes: Summary statistics only for key variables. SD = Standard deviation. Appendix A shows the detailed descriptive statistics in Table A.2.

The descriptive statistics furthermore indicate that commuters are more often male, are better educated, work longer hours, and are less likely to work part time. In addition, they have a higher labour income, shorter tenure and tend to work more often in large firms. Finally, commuters appear to be less satisfied with their leisure time and work than non-commuters.

In our data, the average commuting trip length has increased by 2.3 kilometres between 2002 and 2011 to reach 20.13 kilometres. The average one-way commuting distance of workers is 19 kilometres. This is in line with a range of other studies employing German data (OECD, 2007; Schulze, 2009; Federal Institute for Research on Building, Urban Affairs and Spatial Development, 2012). Hence, our sample selection is likely unrelated to commuting behaviour. The full distribution of commuting distances can be found in Appendix A, Figure A.2.

4.2 Cross-Sectional Evidence

Table 2 reports results for cross-sectional, multivariate regression models. Model I estimates a pooled negative binomial regression (NEGBIN). While the estimated coefficient of being a short distance commuter is insignificant, being a middle distance commuter instead of a non-commuter is associated with a 7.05% change in the expected number of days absent, or equivalently, the conditional mean is 1.07 times larger.[11] Being a long distance commuter leads to a 0.201 proportionate change or 20% change in the number of sickness absence days. The effect is, for example, comparable to the impact due to being a female (Model I: $\beta_{\text{female}} = 0.191,\ p < 0.001$; Model II: $\beta_{\text{female}} = 2.073,\ p < 0.001$, see Appendix A, Table A.3).[12]

Table 2
Estimation results using cross-sectional data. Dependent variable: Days absent. Source: SOEP 2002 – 2011.

	Model I Pooled NEGBIN	Model II Pooled OLS
Short distance commuter	0.0385	0.572
	(1.45)	(1.86)
Middle distance commuter	0.0705*	0.846*
	(2.17)	(2.17)
Long distance commuter	0.201***	2.173***
	(3.72)	(3.36)
No. Obs.	31,567	31,567

Notes: Only the coefficients for the commuting variables are reported. Non-commuters are treated as the reference category. The following control variables are included: female, age, age squared, married, children, college degree, education, health status, working hours, regular part-time, temporary job, blue-collar worker, firm size, public sector, tenure, tenure squared, log(monthly wage), satisfaction with work, leisure and dwelling, household crowding index, business sector dummies, region dummies, year dummies. Appendix A shows the results for control variables in Table A.3. All models are estimated using robust standard errors. *t* statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Model II estimates a pooled least squares regression (OLS). The regression results are almost identical to the ones reported above, indicating that greater commuting distances are associated with more sickness absence. For example, long distance commuters are on average

[11] Recall that the dependent variable is a count variable, and that a negative binomial regression approach models the log of the expected count as a function of the predictor variables. We can interpret the negative binomial regression coefficients as follows: for a one unit change in the predictor variable, the difference in the logs of expected counts is expected to change by the respective regression coefficient, given the other predictor variables in the model are held constant. Hence, the coefficients displayed are equal to the proportionate change in the conditional mean if the regressors change by one unit. For indicator variables the coefficient reflects a proportionate change from the base level. For a detailed explanation, see Cameron and Trivedi (2008).

[12] There is a huge body on literature showing that sex is a strong predictor of sickness absence rates with higher incidence and duration of sickness absence for women predominantly due to the high total workload and the double-exposure situation, e.g. responsibility for household chores and child care, see Leigh (1983), Vistnes (1997), Krantz et al. (2006).

about 2.17 days more absent than those who commute less than 10 kilometres. Since the raw difference in the duration of absence observed between long distance commuters and non-commuters is 1.90 days (Table 1), this difference tends to underestimate the impact of commuting.

4.3 Fixed-Effects Analyses

We next present the findings from fixed-effects estimations to cater for the potential impact of time-invariant, unobservable characteristics on absence behaviour. In Table 3, Model III reports the results for a fixed-effects negative binomial estimation (FE NEGBIN).

Table 3

Estimation results using panel data. Dependent variable: Days absent. Source: SOEP 2002 – 2011.

	Model III FE NEGBIN	Model IV FE OLS
Short distance commuter	0.044	1.262
	(1.91)	(1.89)
Middle distance commuter	0.109***	2.543**
	(3.65)	(2.73)
Long distance commuter	0.191***	3.245*
	(4.35)	(2.50)
Number of observations	31,567	31,567
Number of groups	6,459	6,459

Notes: Only the coefficients for the commuting variables are reported. Non-commuters are treated as the reference category. The following control variables are included: age, age squared, married, children, college degree, education, health status, working hours, regular part-time, temporary job, blue-collar worker, firm size, public sector, tenure, tenure squared, log(monthly wage), satisfaction with work, satisfaction with leisure, satisfaction with dwelling, household crowding index, business sector dummies, year dummies, region dummies and being female (in Model III). Appendix A shows the results for control variables in Table A.3. All models are estimated using robust standard errors. *t* statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The overall effect of commuting distance on the number of days absent is positive and statistically significant.[13] Further, the expected number of days absent is about 11% higher for middle distance and 20% higher for long distance commuters compared to non-commuters.

Since the conditional negative binomial method has been criticised for not being a true fixed-effects model, we also estimate a fixed-effects least squares model (FE OLS; Model IV). Model IV in Table 3 shows that the overall effect of commuting distance is positive and statistically significant.[14] Comparing the magnitudes of the estimated coefficients in Models II and IV clarifies that controlling for time-invariant characteristics tends to increase the effect of being a middle or long distance commuter on the number of days absent. Hence, cross-section estimation of the effect of commuting on sickness absence negatively biases the results. One plausible explanation for this bias is that individuals with unobserved positive attitudes to work

[13] The three degree-of-freedom chi-square test indicates that commuting distance is a statistically significant predictor of absence ($\chi^2(3) = 25.27; \mathrm{p} = 0.0000$).

[14] The F-test indicates that commuting distances are jointly significant at the 5% level ($F(3, 6458) = 3.23$, p-value = 0.0214).

are more likely to accept jobs which require commuting longer distances and are also less likely to be absent. So, the conditional method estimates reported above are conservative. In consequence, being a long distance commuter instead of a non-commuter is associated with 3 absence days more on average ($p < 0.05$), while being a middle distance commuter goes along with 2 more absence days on average ($p < 0.01$). The relationship between commuting short distances and the number of sickness absence days is either not significant or at the borderline of being weakly significant in the main models (see Table 2 and Table 3). Additionally, the sensitivity checks provide no evidence of a link for short distance commuters (see Section 4.4).

Thus, the descriptive evidence and the results of the pooled estimations are confirmed: Compared to commutes of less than ten kilometres, short distance commutes have no impact on sickness absence, middle and long distance commutes increase the duration of absence.

4.4 Robustness Checks

In Table A.4 (Appendix A) we report a number of robustness checks on our results of the fixed-effects negative binomial model. The first two models ((i) and (ii)) are estimated for men and women separately as the determinants of absence may be gender-specific (Leigh, 1983; Vistnes, 1997). The estimated coefficients of the commuting variables are not statistically different from each other. Hence, we obtain no evidence that the effect of commuting distance on absence is gender-specific.[15]

The third model (iii) is estimated for those individuals who do not work in the public sector as sickness absence in the public sector is higher than in the private sector (Winkelmann, 1999; Frick and Malo, 2008). For this sample, the results are almost identical to the ones in the main model, indicating that the observed impact of commuting distance on absence is certainly not a public-sector phenomenon.

The next two models ((iv) and (v)) are estimated for rural and urban communities. There are a number of reasons why the relationship between commuting and absence may vary with such spatial characteristics (Eibich and Ziebarth, 2014). First, employees living in rural regions may be healthier, have a higher quality of life compared to individuals living in urban regions (Ziersch et al., 2009; Zeng et al., 2015) and may commute longer distances. Second, in small, rural communities the observability of the behaviour of others is likely to be more pronounced than in urban communities characterised by greater anonymity. Hence, the impact of commuting on absence may be weaker in rural communities. We indeed find that individuals who are living in rural regions have less sickness absence days (results not documented). Table A.4 indicates

[15] The coefficients across models are not statistically different from each other (Long distance: $\chi_1^2 = 2.45$, p-value = 0.12; middle distance: $\chi_1^2 = 0.49$, p-value = 0.48; short distance: $\chi_1^2 = 0.02$, p-value = 0.87).

that the effects of long distance commutes are similar to those reported in the main model for both regions. Further, the coefficient of short distance commutes becomes significant for the urban community sample, whereas it loses its significance for individuals who commute middle distances. Therefore, our main finding – long distance commutes increase absence – holds for individuals who live in rural and urban communities. The evidence for other distances reveals no clear spatial pattern.

In a further robustness check, we exclude all observations of employees who stated that they had experienced at least one absence spell lasting 30 days or more (model (vi)). Excluding these observations (outliers) makes the sample more homogeneous because such workers no longer receive a wage replacement but a lower level of sick pay instead. The results for this restricted sample are virtually identical to those presented for the baseline. This suggests that unobserved wage reductions due to long sickness absence periods do not affect our results.

Model (vii) tests the sensitivity of our results to reporting error by excluding observations that report small distance changes, namely absolute variations in the reported (continuous) commuting distance of 2 km or less. Hence, we also exclude observations that do not transcend the distance categories defined above. Such small changes will more likely refer to measurement error in the reported commuting distance as respondents in one year will, for example, report 13 km and in the next year 15 km without changing actual commuting distance. We see that the effect of commutes is very similar to that reported in the main model, indicating that the effect of commuting to work on sickness absence is not due to measurement error.[16]

In models (viii) and (ix) we have experimented with several functional forms and categorisations for commuting distance. In model (viii) we classify commuting distance as a dummy variable (equals 1 if individual commutes more than 50 km) and in model (ix) we estimate a log-linear specification of commuting distance. In model (viii) the coefficient of the dummy variable indicates that the expected number of days absent is about 14% higher for those who travel more than 50 kilometres compared to those who travel fewer kilometres. In model (ix) the point estimate of the continuous commuting distance variable (in log), and therefore the elasticity, is 0.0461 (s.e. 0.009). Thus, if the average logarithm of commuting distance, 2.22 in our data, falls to about 0, sickness absence days will decline by about 10% (0.0461×2.22). van

[16] Another attempt to deal with measurement error is to calculate a proxy for commuting time. To obtain the commuting time we built the difference between the daily working hours including travel time to and from work (taken from the question: "How many hours per normal workday do you spend on job, apprenticeship, second job (including travel time to and from work)?") and the usual daily working hours (taken from the question: "And how many hours do you generally work, including any overtime?") divided by 5 workdays. Again, we only find a positive and statistically significant effect of long commutes, particularly of commutes which take more than 45 minutes ($\beta_{46\,\min and\, more} = 0.0559, p = 0.013$). We do not use this measure of commuting time as our focal explanatory variable as it is calculated in an imprecise manner. Further, commuting time may be influenced by many factors, for instance by changes in congestion or infrastructure or even commuting modes.

Ommeren and Gutiérrez-i-Puigarnau (2011) use a similar measure of commuting distance and find that commuting distance induces absence or shirking behaviour with an elasticity twice as large as the one we find (0.07 to 0.09). Since they also use data from the SOEP and employ a similar identification strategy, we can test whether it is the choice of the explanatory variables that is driving the difference in the results. To do this we re-run our analysis including only explanatory variables similar to those used by van Ommeren and Gutiérrez-i-Puigarnau (2011) (not reported). We find a point estimate of 0.062, indicating that failure to include additional confounders into the estimations is likely to result in overestimates of the strength of the association between commuting and sickness absence. We also estimate the latter model using the categorical commuting distance variable instead of the continuous measure. We find that only middle and long distance commutes are associated with higher sickness absence days, while short distance commutes are not. It is thus apparent that the effect documented by van Ommeren and Gutiérrez-i-Puigarnau (2011) does not hold in general, as we find no evidence of an impact of short distance commutes on absence in our application.[17]

The next models ((x) to (xiv)) are an attempt to address a potential selection bias since our estimates are, thus far, based on a sample of workers who change neither employer nor residence. In our setting, endogeneity might, first, result from the self-selection of employees in a group of workers who do not change residence or employer. Strictly speaking we cannot exclude the possibility that individuals with unobserved positive attitudes to work are more likely to accept jobs at longer distances and are also less likely to be absent. Second, employees may move residence or job as a reaction to employer-induced workplace relocation. Third, if an employer needs some employees to move to a different part of the firm at a different location, employees are usually asked whether they are willing to move or not. To tackle the potential bias resulting from these issues, we additionally employ two strategies. The first is to estimate the fixed-effects negative binomial model on other, less-selective samples. In particular, we include data on employees who change the employer only (model (x)), employees who change residence only (model (xi)), and employees who change both employer and residence (model (xii)). Second, we replace commuting distance with lagged values of commuting distance, in order to avoid the influence of sickness absence on contemporaneous commuting distance (model (xiii) and (xiv)). This strategy is based on the assumption that lagged commuting distance is uncorrelated with the current sickness absence residual, which assumes no serial correlation in the sickness absence residuals for the two periods. This approach reveals that especially commuting long distances translates into higher sickness absence days for the next two years.

[17] Our main results remain robust when we use more fine grained commuting classes or models with linear splines: Individuals who commute less than 20 to 25 km do not have a higher number of absence days than non-commuters. These additional robustness checks are available upon request.

Overall, none of these analyses yields any other qualitative finding than those reported above, indicating that the effect of commuting long distances to work on sickness absence is not due to self-selection based endogeneity bias.[18]

Finally, in model (xv) we alter the methodology to see whether the choice of the dependent variable affects the results. We therefore estimate a random-effects probit model where we only distinguish between 'never having been absent' and 'having been absent at least once'. The results indicate that being a middle or long distance commuter increases the probability of being absent at least once in a given year. This finding also supports the hypothesis that only longer commuting distances positively affect sickness absence from work.

4.5 Transmission Channels

The previous analysis has uncovered a robust impact of commuting longer distances on the number of days of absence from work. In this subsection, we investigate various hypotheses concerning the underlying mechanism of this relationship.

As outlined in the introduction, there is substantial evidence that commuting is associated with increased levels of illness. Since absence is negatively related to health (Puhani and Sonderhoff, 2010 and Goerke and Pannenberg, 2015; for example, present according evidence for Germany), the impact of commuting on absence may be due to health effects. To accommodate this possibility, the estimations presented thus far include a subjective measure of health. We further analyse health as transmission mechanism by, first, omitting the health variables included in the estimations depicted in Table 3. Second, we include additional health indicators, such as satisfaction with health, concern about individual's own health, degree of disability, the number of overnight hospital stays and the number of annual doctor visits. An appreciable change in the coefficients of the commuter variables would suggest that the baseline coefficient is biased. The results obtained with additional controls should than be closer to the causal effect of commuting on sickness absence we seek to uncover.[19]

Table 4 depicts the results for the fixed-effect specifications and reveals that more healthy people are indeed less absent from work. Moreover, we see that the magnitudes of the estimated coefficients of the commuter variables decline to some extent if health indicators are included. However, the estimated coefficients of the commuter variables across models are not statistically

---

[18] Given that commuting long distances is positively associated with absence days, we would expect someone who stops commuting long distances to have fewer sickness absence days. Hence, we also investigated the effect of quitting commuting on absence. An individual who stops commuting long distances decreases the number of sickness absence days significantly by about 16%. This evidence supports our identification strategy.

[19] We have also analysed the effect of interactions of distance with health indicators, but have not found any significant effects. This indicates that an employee's marginal costs of commuting do not depend on the individual's state of health.

different from each other. Accordingly, this effect does not explain the observed impact of commuting on absence. Otherwise, the significant coefficients of the commuting covariates would become statistically insignificant when controlling for health.

Table 4

Transmission channels. Fixed-effects estimates. Dependent variable: Days absent. Source: SOEP 2002 – 2011.

	Baseline (see Table 3)		Baseline without subjective health measures		Baseline with additional health measures	
	FE NEGBIN	FE OLS	FE NEGBIN	FE OLS	FE NEGBIN	FE OLS
Focal variable:						
Short distance commuter	0.044	1.262	0.0576*	1.320	0.0375	0.909
	(1.91)	(1.89)	(2.49)	(1.84)	(1.62)	(1.37)
Middle distance commuter	0.109***	2.543**	0.121***	2.740**	0.104***	2.271*
	(3.65)	(2.73)	(4.04)	(2.91)	(3.46)	(2.48)
Long distance commuter	0.191***	3.245*	0.205***	3.969**	0.195***	2.837*
	(4.35)	(2.50)	(4.67)	(2.84)	(4.42)	(2.22)
Health status: very good (ref.)						
good	0.157***	1.230**			0.104**	0.243
	(4.67)	(3.15)			(3.02)	(0.62)
acceptable	0.343***	2.567***			0.200***	-0.333
	(9.52)	(4.91)			(5.10)	(-0.61)
less good	0.581***	10.86***			0.308***	4.411***
	(14.20)	(10.47)			(6.37)	(4.08)
bad	0.947***	47.49***			0.503***	35.67***
	(14.08)	(9.90)			(6.53)	(7.76)
Health satisfaction					-0.0414***	-0.947***
					(-6.67)	(-5.44)
Life satisfaction					0.00360	-0.288
					(0.56)	(-1.82)
Concerned about health: very (ref.)						
somewhat					-0.0179	-1.729*
					(-0.69)	(-2.19)
not at all					-0.0393	-1.248
					(-1.25)	(-1.50)
Degree of disability					0.000357	-0.172**
					(0.50)	(-2.77)
# of hospital stays					-0.00633***	-0.082
					(-4.13)	(-1.16)
# of doctor visits					0.00802***	0.272***
					(18.21)	(9.63)
No. observations	31,567	31,567	31,567	31,567	31,354	31,354
No. groups	6,459	6,459	6,459	6,459	6,432	6,432

Notes: Only the coefficients for the commuting variables and those of potential health channels are reported. Non-commuters are treated as the reference category. The baseline models correspond to Model III and Model IV of Table 3. All models are estimated using robust standard errors. *t* statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

While health-related absence may be regarded as involuntary, the standard labour-supply perspective on absence views such behaviour as voluntary adjustment to predetermined and overly long or insufficiently flexible working hours (Allen, 1981). Since commuting increases the length of the total workday while simultaneously reducing time for private use, the need to adjust total working time to the preferred amount is likely to be greater for individuals who commute. In order to scrutinise this transmission channel, we estimate extended specifications of Models III and IV, as depicted in Table 3, and add two dummy variables which indicate whether individuals would like to work less or more hours than they actually do. The estimated coefficients of the commuter variables (not documented) are basically unaffected compared to those in the main model. This is also true if we include further working time indicators, such as the number of actual hours worked, overtime hours per week or having a second job. Therefore, commuting does not result in greater voluntary absence, which is often interpreted as shirking.[20]

In a substantial number of empirical studies, job (in-) security has been found to affect absence from work (see e.g. Staufenbiel et al., 2010; Bratberg et al., 2015). Moreover, reduced job security has a disciplining effect, suggesting that workers are more likely to accept jobs at longer distances and are also less likely to be absent. Hence, job insecurity may influence both the probability of becoming a commuter and of being absent from work. We investigate this transmission channel by including a variable in extended specifications of Models III and IV that indicates whether the respondent is concerned about its own job security. Individuals who are not concerned about their job indeed have higher absence (results not documented).[21] However, the estimated coefficients of the commuting variables are basically the same as shown for the baseline. Alternatively, we use the unemployment rate (at the level of federal states) as a proxy of job insecurity. Its inclusion does not substantially alter the estimated coefficients of interest. Consequently, the impact of commuting on absence does not arise because commuters are concerned about their jobs.

Research on the determinants of sickness absence has shown that higher wages are associated with lower sickness absence rates (see e.g. Drago and Wooden, 1992; Piha et al., 2009). Further, standard urban economic theory (see e.g. Lucas and Rossi-Hansberg, 2002) suggests that workers should be willing to accept longer commutes only if they are compensated by higher wages. Hence, wages may influence both the probability of commuting long distances and of being absent from work, which could mean that the impact of commuting on absence is due to wage effects (Ross and Zenou, 2008). To account for this possibility, the estimations presented thus far include information on monthly labour income. We further analyse income as transmission mechanism by, first, omitting the income variable included in the estimations depicted in Table 3. Second, we include an additional variable, indicating whether income

[20] In a further step, we have also included information on private time use, for instance, the average time per day spent on running errands, housework, child care, care for people with disabilities and other dependants living in the household, leisure time, time for repairs and garden work. The estimated coefficients of the commuter variables are basically unaffected by the inclusion of the private time use variables.

[21] These results complement findings by Ichino and Riphahn (2005) who find that average absence substantially rises, once the probability of being fired decreases.

increased, decreased or remained constant in comparison to the previous year (e.g. due to promotion). Excluding or adding further income controls does not change our main results.[22]

We conclude that long distance commuting raises absence. This effect is partially due to health consequences but cannot be explained by it. Moreover, it does not arise because of a change in job security or because commuters face a greater mismatch between actual and desired working time. One reason why absence from work is much lower for those traveling shorter distances might be that they more frequently show up at work, despite anticipating an upcoming illness. Such behaviour could arise because short distance commuters can more easily return home if their health condition deteriorates than employees who have to travel longer distances to reach their place of residence. Consequently, we might expect long-distance commuters to exhibit lower levels of presenteeism. Unfortunately, we are not able to investigate to which extent employees with different commuting distances go to work although being sick. This is the case since our data does not provide information on presenteeism days. This limitation may be worth addressing in future research.

Alternatively, one may hypothesise that commuting is associated with lower work effort and, hence, more absenteeism. Our data does not enable us to directly provide further evidence on this kind of transmission mechanism. However, when we consider weekly overtime or actual weekly work time as proxies for work effort and as our dependent variables, we find that commuting distance has a positive effect on working overtime or working more hours than the number which has been contractually agreed upon. Hence, one should be cautious with the interpretation of sickness absence as inverse measure of productivity or work effort.[23]

5. Conclusion

In this paper we enrich the literature on the relationship between commuting distance and sickness absence using panel data for Germany. Empirically, we know very little about this linkage. We address a possible reverse causality bias by exploiting variation of commuting distance within individuals when there are no changes in residence and employer. For Germany, we find a causal effect of commuting distance on sickness absence. We show that individuals with long commutes have around 20% more sickness absence days than similar individuals with

[22] A full set of the results of the specification described in this sub-section are available upon request.

[23] One additional hypothesis we considered is that income, working hours or the desired working hours (work more or less hours) might be proxy indicators of work effort. Since the coefficient of long distance commutes is basically unaffected by the inclusion of these variables one may argue that the impact of commuting on absence does not arise because commuters provide lower work effort. However, as with other proxy indicators, there is a difficulty in ensuring that the claimed relationship is not confounded by other variables. Nevertheless, there is a growing body of literature showing that commuting increases the number of working hours and, hence, labour supply (Gutiérrez-i-Puigarnau and van Ommeren, 2010; 2014).

no commutes. The effect of middle distance commutes is much lower, i.e. about 11%. The effect becomes zero at commuting distances of less than 25 kilometres. The results are robust across specifications and when accounting for selection effects. Furthermore, we explore potential explanations for the effect of commutes to work on absence. We find that the impact is not due to working hours mismatch or poor health. A deeper investigation of the determinants shows that differences in personnel characteristics, job related aspects and factors compensating for commuting are not able to explain the gap in sickness absence from work either.

Our findings have a number of implications. First, we demonstrate that sickness absence due to commuting is an important characteristic of the (German) labour market, which is in line with a range of theoretical models (Zenou, 2002). Second, the present study suggests that commuting may have far-reaching consequences for both employees and the financial performance of employers. Hence, evidence of an absence-commute relationship puts a price on the work commute and should be considered in cost-benefits assessments, since absence from work causes sizeable costs not only for the employer but also for the employee. Consequently, our findings point to the economic benefit from transport infrastructure improvements as well as to potential costs savings for the health care system. Third, it is important to consider the positive effect of commuting on sickness absence when discussing the expansion of economic regions or increasing the mobility of the workforce. Hence, there is a need for an integrating different policy areas concerning commuting, such as planning policy, transport policy, policies at the workplace, social policies and innovative policies.

References

Allen, S.G., (1981). An empirical model of work attendance. The Review of Economics and Statistics 63 (1), 77 – 87.

Allison, P.D., (2009). Fixed effects regression models. Series: Quantitative applications in the social sciences 07 (160).

Allison, P.D., Waterman, R.P., (2002). Fixed-effects negative binomial regression models. Social Methodology 32 (1), 247 – 265.

Besser, L.M., Marcus, M., Frumkin, H., (2008). Commute time and social capital in the US. American Journal of Preventive Medicine 34 (3), 207 – 211.

Block, J., Goerke, L., Millán, J.M., Román, C. (2014). Family employees and absenteeism. Economics Letters 123 (1), 94 – 99.

Bratberg, E., Monstadt, K. (2015). Worried sick? Worker responses to a financial shock. Labour Economics 33, 111 – 120.

Cameron, C., Trivedi, P., (2008). Regression analysis of count data. Cambridge University Press.

Carta, F., De Philippis, M., (2015). You've come a long way, baby. Effects of commuting times on couples' labour supply. Bank of Italy Temi di Discussione (Working Paper) No. 1003.

Costa, G., Pickup, L., Di Martino, V., (1988). Commuting – a further stress factor for working people: Evidence from the European community. An empirical study. International Achieves of Occupational and Environmental Health 60 (5), 377 – 385.

De Paola, M., (2010). Absenteeism and peer interaction effects: evidence from an Italian public institute. The Journal of Socio-Economics 39 (3), 420 – 428.

Dionne, G., Dostie, B., (2007). New evidence on the determinants of absenteeism using linked employer-employee data. Industrial and Labor Relations Review 61 (1), 108 – 120.

Drago, R., Wooden, M., (1992). The determinants of labor absence: Economic factors and workgroup norms across countries. Industrial and Labor Relations Review 45 (4), 764 – 778.

Eibich, P., Ziebarth, N.R., (2014). Examining the structure of spatial health effects in Germany using Hierarchical Bayes Models. Regional Science and Urban Economics 49 (C), 305 – 320.

Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR), (2012). Spatial Development Report 2011. Bonn. Germany.

Federal Statistical Office, (2008). Relocation of economic activities. Wiesbaden. Germany. (https://www.destatis.de/DE/Publikationen/Thematisch/UnternehmenHandwerk/VerlagerungAktivitaeten5529301069004.pdf?__blob=publicationFile).

Federal Statistical Office, (2013). Census. Special microcensus evaluations: Commuting 2012. Wiesbaden. Germany.

Frick, B., Malo, M.A., (2008). Labor market institutions an individual absenteeism in the European Union: the relative importance of sickness benefit systems and employment protection legislation. Industrial Relations 47 (4), 505 – 529.

Goerke, L., Pannenberg, M., (2015). Trade union membership and sickness absence: Evidence from a sick pay reform. Labour Economics 33, 13 – 25.

Gottholmseder, G., Nowotny, K., Pruckner, G.J., Theurl, E., (2009). Stress perception and commuting. Health Economics 18 (5), 559 – 576.

Gulian, E., Matthews, G., Glendon, A.I., Davies, D.R., Debney, L.M., (1989). Dimensions of driver stress. Ergonomics 32 (6), 585 – 602.

Gutiérrez-i-Puigarnau, E., van Ommeren, J.N., (2010). Labour supply and commuting. Journal of Urban Economics 68 (1), 82 – 89.

Gutiérrez-i-Puigarnau, E., van Ommeren, J.N., (2014). Labour supply and commuting revisited. Urban Studies 52 (14), 2551 – 2563.

Gutiérrez-i-Puigarnau, E., Mulalic, I., van Ommeren, J.N., (2014). Do rich households live farther away from their workplaces?. Journal of Economic Geography 16 (1), 177 – 201.

Hansson, E., Mattisson, K., Björk, J., Östergren, P., Jakobsson, K., (2011). Relationship between commuting and health outcomes in a cross-sectional population survey in southern Sweden. BMC Public Health, 11:843.
Hassink, Wolter H. J. and Roberto M. Fernandez (2018), Worker morale and effort: Is the relationship causal?. The Manchester School (forthcoming).
Hilbe, J.M., (2007). Negative binomial regression. Cambridge University Press.
Hennessy, D.A., (2008). The impact of commuter stress on workplace aggression. Journal of Applied Social Psychology 38 (9), 2315 – 2355.
Hennessy, D.A., Wiesenthal, D.L., (1997). The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours. Ergonomics 40 (3), 348 – 361.
Ichino, A., Riphahn, R.T., (2005). The effect of employment protection on worker effort: Absenteeism during and after probation. Journal of the European Economic Association 3 (1), 120 –143.
Kirby, D.K., LeSage, J.P., (2009). Changes in commuting to work times over the 1990 to 2000 period. Regional Science and Urban Economics 39 (4), 460 – 471.
Kluger, A.N., (1998). Commute variability and strain. Journal of Organizational Behavior 19 (2), 147 – 165.
Krantz, G., Lundberg, U., (2006): Workload, work stress, and sickness absence in Swedish male and female white-collar employees. Scandinavian Journal of Public Health 34 (3), 238 – 246.
Künn-Nelen, A., (2016). Does commuting affect health?. Health Economics 25 (8), 984 – 1004.
Leigh, J.P., (1983). Sex differences in absenteeism. Industrial Relations 22 (3), 349 – 361.
Lindström, M., (2004). Social capital, the miniaturisation of community and self-reported global and psychological health. Social Science & Medicine 59 (3), 595 – 607.
Livanos, I., Zangelides, A., (2013). Unemployment, labor market flexibility and absenteeism: a pan-european study. Industrial Relations 52 (2), 492 – 515.
Lucas, R., Rossi-Hansberg, E., (2002). On the internal structure of cities. Econometrica 70 (4), 1445 – 1476.
Lyons, G., Chatterjee, K., (2008). A human perspective on the daily commute: Costs, benefits and trade-offs. Transportation Reviews 28 (2), 181 – 198.
Magee, C., Stefanic, N., Caputi, P., Iverson, D., (2011). Occupational factors and sick leave in Australian employees. Journal of Occupational & Environmental Medicine 53 (6), 627 – 632.
Matthews, G., Dorn, L., Glendon, A.I., (1991). Personality correlates of driver stress. Personality and Individual Differences 12 (6), 535 – 549.
Mattisson, K., Hakansson, C., Jakobsson, K., (2015). Relationships between commuting and social capital among men and women in southern Sweden. Environment and Behaviour 47 (7), 734 – 753.
National Travel Survey: 2012, (2013). National travel survey statistics. Department for Transport. https://www.gov.uk/government/collections/national-travel-survey-statistics
Novaco, R.W., Stokols, D., Campbell, J., Stokols, J., (1979). Transportation, stress, and community psychology. American Journal of Community Psychology 7 (4), 361 – 380.
OECD, (2007). OECD Rural Policy Review: Germany. Paris. France.
Olsson, L.E., Gärling, T., Ettema, D., Friman, M., Fujii, S., (2013). Happiness and satisfaction with work commute. Social Indicators Research 111 (1), 255 – 263.
Piha, K., Laaksonen, M., Martikainen, P., Rahkonen, O., Lahelma, E., (2009). Interrelationships between education, occupational class, income and sickness absence. The European Journal of Public Health 20 (3), 276 – 280.
Puhani, P.A., Sonderhoff, K., (2010). The effects of a sick pay reform on absence and on health-related outcomes. Journal of Health Economics 29 (2), 285 – 302.
Putnam, R.D., (2000). Bowling alone: The collapse and revival of American community. Touchstone. New York.

Roberts, J., Hodgson, R., Dolan, P., (2011). “It’s driving her mad”: Gender differences in the effects of commuting on psychological health. Journal of Health Economics 30 (5), 1064 – 1076.

Ross, S.L., Zenou, Y., (2008). Are shirking and leisure substitutable? An empirical test of efficiency wages based on urban economic theory. Regional Science and Urban Economics 38 (5), 498 – 517.

Sandow, E., Westerlund, O., Lindgren, U., (2014). Is your commute killing you? On the mortality risk of long-distance commuting. Environment and Planning A 46 (6), 1496 – 1516.

Schaeffer, M.H., Street, S.W., Singer, J.E., Baum, A., (1988). Effects of control on the stress reactions of commuters. Journal of Applied Social Psychology 18 (11), 944 – 957.

Schulze, S., (2009). Einige Beobachtungen zum Pendlerverhalten in Deutschland. Hamburg Institute of International Economics (HWWI) Policy Paper, No. 1-19.

Staufenbiel, T., König, C. (2010). A model for the effects of job insecurity on performance, turnover intention and absenteeism. Journal of Occupational and Organizational Psychology 83 (1), 101 – 117.

Stutzer, A., Frey, B.S., (2008). Stress that doesn’t pay: The commuting paradox. The Scandinavian Journal of Economics 110 (2), 339 – 366.

van Ommeren, J.N., Gutiérrez-i-Puigarnau, E., (2011). Are workers with a long commute less productive? An empirical analysis of absenteeism. Regional Science and Urban Economics 41 (1), 1 – 8.

Vistnes, J.P., (1997). Gender differences in days lost from work due to illness. Industrial and Labor Relations Review 50 (2), 304 – 323.

Wagner, G.G., Frick, J.R., Schupp, J., (2007). The German Socio-Economic Panel study (SOEP) – evolution, scope and enhancements. Journal of Applied Social Science (Schmollers Jahrbuch) 127 (1), 139 – 169.

Walsleben, J.A., Norman, R.G., Novak, R.D., O’Malley, E.B., Rapoport, D.M., Strohl, K.D., (1999). Sleep habits of Long Island rail road commuters. Sleep 22 (6), 728 – 734.

Winkelmann, R., (1999). Wages, firm size and absenteeism. Applied Economics Letters 6 (6), 337 – 341.

Wooldridge, J.M., (2010). Econometric analysis of cross section and panel data. MIT Press. Cambridge et al.

Zax, J.S., (1991). The substitution between moves and quits. The Economic Journal 101 (409), 1510 – 1521.

Zax, J.S., Kain, J.F., (1996). Moving to the suburbs: Do relocating companies leave their black employees behind?. Journal of Labor Economics 14 (3), 472 – 504.

Zeng, D., You, W., Mills, B., Alwang, J., Royster, M. Anson-Dwamena, R., (2015). A closer look at the rural health disparities: Insights from four major diseases in the Commonwealth of Virginia. Social Science and Medicine 140, 62 – 68.

Zenou, Y., (2002). How do firms redline workers? Journal of Urban Economics 52 (3), 391 – 408.

Ziebarth, N.R., Karlsson, M., (2010). A natural experiment on sick pay cuts, sickness absence, and labor costs. Journal of Public Economics 94 (11 – 12), 1108 – 1122.

Ziersch, A., Baum, F., Darmawan, I.G., Kavanagh, A.M., Bentley, R.J., (2009). Social capital and health in rural and urban communities in South Australia. Australian and New Zealand Journal of Public Health 33 (1), 7 – 16.

Appendix A

Table A.1
Variable definitions. Source: SOEP 2002 – 2011.

Variable	*Definition*
Dependent variable	
Days absent	Number of sickness absence days.
Focal variable	
Commuter	Commuting distance D measured one-way in kilometres. Categorical variable: 0 = "non-commuter ($D < 10$ km)", 1 = "short distance commuter ($10 \leq D < 25$ km)", 2 = "middle distance commuter ($25 \leq D < 50$ km)", 3 = "long distance commuter ($D \geq 50$ km)".
Personal characteristics	
Female	Dummy equals 1 for female.
Age	Age in years.
Age^2	Age squared.
Married	Dummy equals 1 if individual is living together with partner (either as a married or unmarried couple).
Children	Dummy equals 1 if children live in the household.
Education	A five point scale measuring highest level of education attainment: 0 = "no or other school certificate", 1 = "secondary general school certificate", 2 = "intermediate school degree", 3 = "leaving certificate from vocational high school", 4 = "college entrance exam".
College Degree	Dummy equals 1 if individual has completed college education.
Health status	A five point indicator of self-reported health status: 1 = "very good", 2 = "good", 3 = "acceptable", 4 = "less good", 5 = "bad".
Job related aspects	
Tenure	Number of years in present job.
$Tenure^2$	Job tenure squared.
Working hours	Contractually agreed hours of work per week.
Regular part-time	Dummy equals 1 if individual works part-time.
Temporary job	Dummy equals 1 if individual has a fixed-term employment contract.
Blue-collar worker	Dummy equals 1 if individual is a blue-collar worker.
Firm size	Size of company: 0 = "< 5 employees", 1 = "5 – 19 employees", 2 = "20 – 99 employees", 3 = "100 – 199 employees", 4 = "200 – 1999 employees", 5 = "2000 employees and over".
Public sector	Dummy equals 1 if individual works in the public sector.
Log (monthly wage)	Current gross labor income, being corrected by purchasing power parity and harmonized by consumer price index. Variable expressed in natural logarithms.
Variables compensating for commuting	
Satisfaction with work	Satisfaction with main job measured on an eleven point scale from 0 = "completely dissatisfied" to 10 = "completely satisfied".
Satisfaction with leisure	Satisfaction with leisure time measured on an eleven point scale from 0 = "completely dissatisfied" to 10 = "completely satisfied".
Satisfaction with dwelling	Satisfaction with dwelling measured on an eleven point scale from 0 = "completely dissatisfied" to 10 = "completely satisfied".
Household crowding index	Household crowding index defined as the number of usual residents in a dwelling divided by the number of rooms in the dwelling.
Industry	9 dummies equalling 1 for individuals working in the named industry: agriculture, energy, mining, manufacturing, construction, trade, transport, bank or insurance, services.
Region	Dummy variables for the 16 federal states of Germany.
Year	Dummy variables for each year covered by the sample.

Table A.2

Descriptive statistics for full sample and for commuter categories. Source: SOEP 2002 – 2011.

	Full sample				Non-Commuter		Short distance commuter		Middle distance commuter		Long distance commuter	
	Mean	SD	Min.	Max.	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Number of days absent	10.36	24.70	0	365	9.96	24.31	10.59	25.12	10.43	23.72	11.86	27.62
Incidence of absence	0.6529	0.47	0	1	0.6328	0.48	0.6589	0.47	0.6786	0.46	0.7005	0.45
Female	0.47	0.49	0	1	0.54	0.49	0.45	0.49	0.37	0.48	0.30	0.46
Age	45.25	9.00	19	64	45.64	9.00	45.08	9.01	44.71	8.82	44.87	9.34
Married	0.74	0.43	0	1	0.73	0.44	0.75	0.43	0.74	0.43	0.76	0.42
Children	0.40	0.48	0	1	0.38	0.48	0.41	0.49	0.39	0.48	0.40	0.49
Education:												
No school certificate (ref.)	0.09	0.25	0	1	0.07	0.27	0.09	0.25	0.07	0.21	0.05	0.15
Sec. general school certificate	0.25	0.43	0	1	0.27	.044	0.26	0.44	0.22	0.41	0.18	0.38
Intermediate school degree	0.38	0.48	0	1	0.39	0.48	0.37	0.47	0.38	0.48	0.40	0.49
Vocational high school	0.06	0.24	0	1	0.05	0.22	0.06	0.24	0.07	0.26	0.07	0.27
College entrance exam	0.22	0.41	0	1	0.19	0.39	0.22	0.41	0.26	0.43	0.30	0.45
College degree	0.25	0.43	0	1	0.23	0.42	0.22	0.41	0.30	0.45	0.34	0.47
Health status:												
Very good (ref.)	0.09	0.25	0	1	0.07	0.26	0.09	0.25	0.08	0.24	0.11	0.27
Good	0.46	0.49	0	1	0.47	0.49	0.45	0.49	0.45	0.49	0.44	0.49
Acceptable	0.34	0.47	0	1	0.33	0.47	0.35	0.47	0.36	0.48	0.33	0.47
Less good	0.10	0.30	0	1	0.09	0.29	0.10	0.31	0.10	0.30	0.11	0.31
Bad	0.01	0.10	0	1	0.01	0.11	0.01	0.10	0.01	0.10	0.01	0.12
Working hours	35.18	7.74	1.5	72.5	33.94	8.65	35.55	7.25	36.75	6.10	38.14	4.73
Regular part-time	0.21	0.40	0	1	0.27	0.44	0.19	0.39	0.12	0.33	0.06	0.25
Temporary job	0.02	0.15	0	1	0.02	0.14	0.02	0.15	0.02	0.14	0.03	0.17
Blue-collar worker	0.29	0.45	0	1	0.30	0.46	0.31	0.46	0.26	0.44	0.20	0.40
Firm size:												
< 5 employees (ref.)	0.06	0.19	0	1	0.05	0.22	0.06	0.16	0.04	0.13	0.05	0.14
5 – 19 employees	0.12	0.32	0	1	0.14	0.35	0.11	0.32	0.08	0.27	0.06	0.25
20 – 99 employees	0.19	0.39	0	1	0.22	0.41	0.18	0.39	0.15	0.36	0.15	0.36
100 – 199 employees	0.10	0.31	0	1	0.11	0.32	0.10	0.30	0.09	0.29	0.08	0.27
200 – 1999 employees	0.26	0.43	0	1	0.24	0.43	0.27	0.44	0.26	0.44	0.27	0.44
2000 employees and over	0.27	0.44	0	1	0.20	0.40	0.28	0.45	0.38	0.48	0.39	0.48
Public sector	0.34	0.47	0	1	0.37	0.48	0.33	0.47	0.32	0.46	0.31	0.46
Tenure	14.95	9.52	2	49.80	15.06	9.60	15.23	9.49	14.45	9.32	13.51	9.48
Log (monthly wage)	7.76	0.57	4.23	10.23	7.62	0.60	7.80	0.53	7.94	0.51	8.06	0.47
Satisfaction with work	6.95	1.86	0	10	7.04	1.86	6.90	1.85	6.88	1.84	6.79	1.99
Satisfaction with leisure	6.61	2.01	0	10	6.77	1.98	6.56	1.99	6.46	1.97	5.99	2.19
Satisfaction with dwelling	7.82	1.72	0	10	7.80	1.73	7.84	1.68	7.83	1.70	7.80	1.81
Household crowding index	0.70	0.28	0.09	10	0.71	0.29	0.70	0.28	0.67	0.28	0.69	0.27
Number of observations	31,567				14,113		10,435		5,129		1,890	
%	100%				46%		33%		16%		5%	

Table A.3

Estimation results. Dependent variable: Days absent. Source: SOEP 2002 – 2011.

	Model I Pooled NEGBIN	Model II Pooled OLS	Model III FE NEGBIN	Model IV FE OLS
Focal variable				
Non-commuter (ref.)				
Short distance commuter	0.0385	0.572	0.0442	1.262
	(1.45)	(1.86)	(1.91)	(1.89)
Middle distance commuter	0.0705^{*}	0.846^{*}	0.109^{***}	2.543^{**}
	(2.17)	(2.17)	(3.65)	(2.73)
Long distance commuter	0.201^{***}	2.173^{***}	0.191^{***}	3.245^{*}
	(3.72)	(3.36)	(4.35)	(2.50)
Personal characteristics				
Female	0.191^{***}	2.073^{***}	0.256^{***}	-
	(6.52)	(6.11)	(8.77)	
Age	-0.0362^{***}	-0.374^{**}	-0.0976^{***}	-1.892^{***}
	(-3.33)	(-2.96)	(-9.11)	(-5.04)
Age^2	0.000525^{***}	0.00565^{***}	0.000829^{***}	0.0231^{***}
	(4.20)	(3.69)	(6.92)	(5.34)
Married	0.0133	0.0644	0.00987	0.616
	(0.44)	(0.17)	(0.37)	(0.58)
Children	-0.0561	0.0875	0.0654^{**}	0.308
	(-1.91)	(0.27)	(2.75)	(0.46)
Education: No school certificate (ref.)				
Secondary general school certificate	0.0660	0.454	-0.00179	7.287^{***}
	(1.36)	(0.67)	(-0.04)	(5.88)
Intermediate school degree	-0.0864	-0.535	0.129^{*}	3.137^{**}
	(-1.78)	(-0.81)	(2.53)	(2.84)
Vocational high school	-0.0722	-1.000	0.239^{***}	0.875
	(-1.08)	(-1.30)	(3.56)	(0.57)
College entrance exam	-0.151^{**}	-1.231	0.322^{***}	5.674^{***}
	(-2.73)	(-1.83)	(5.55)	(4.58)
College degree	-0.181^{***}	-1.245^{**}	-0.0259	-1.505
	(-5.01)	(-3.18)	(-0.69)	(-0.90)
Health status: Very good (ref.)				
Good	0.270^{***}	1.386^{***}	0.157^{***}	1.230^{**}
	(5.80)	(5.33)	(4.67)	(3.15)
Acceptable	0.661^{***}	4.557^{***}	0.343^{***}	2.567^{***}
	(13.47)	(13.36)	(9.52)	(4.91)
Less good	1.335^{***}	15.65^{***}	0.581^{***}	10.86^{***}
	(23.28)	(19.74)	(14.20)	(10.47)
Bad	2.381^{***}	58.93^{***}	0.947^{***}	47.49^{***}
	(26.99)	(13.43)	(14.08)	(9.90)
Job related aspects				
Working hours	0.0173^{***}	0.165^{***}	0.00776^{***}	0.0736
	(6.29)	(5.30)	(3.67)	(1.18)
Regular part-time	0.149^{**}	1.244^{*}	0.0462	-0.544
	(3.23)	(2.12)	(1.30)	(-0.60)
Temporary job	-0.101	-1.471^{*}	-0.0922	-2.130^{*}
	(-1.39)	(-2.43)	(-1.59)	(-2.37)
Blue-collar worker	0.286^{***}	2.787^{***}	-0.0338	1.354
	(8.36)	(6.67)	(-1.18)	(1.40)
Firm size: < 5 employees (ref.)				
5 – 19 employees	0.127	0.987	0.249^{***}	-0.827
	(1.64)	(1.55)	(3.99)	(-0.66)
20 – 99 employees	0.293^{***}	2.608^{***}	0.371^{***}	2.456
	(3.87)	(4.00)	(5.95)	(1.61)
100 – 199 employees	0.366^{***}	3.304^{***}	0.383^{***}	2.337
	(4.52)	(4.60)	(5.87)	(1.47)
200 – 1999 employees	0.444^{***}	4.428^{***}	0.454^{***}	3.640^{*}
	(5.87)	(6.76)	(7.25)	(2.46)
2000 employees and over	0.468^{***}	4.578^{***}	0.446^{***}	2.682

	(6.15)	(6.90)	(7.10)	(1.79)
Public sector	0.158^{***}	1.487^{***}	0.177^{***}	0.373
	(4.76)	(3.98)	(6.31)	(0.48)
Tenure	-0.00542	0.0257	-0.000529	0.461^{**}
	(-1.15)	(0.46)	(-0.13)	(3.20)
Tenure^2	0.0000772	-0.00109	0.0000895	-0.006
	(0.64)	(-0.69)	(0.84)	(-1.60)
Log (monthly wage)	-0.0724	-1.107^{*}	0.184^{***}	-1.333
	(-1.88)	(-2.58)	(6.43)	(-1.43)
Variables compensating for commuting				
Satisfaction with work	-0.0496^{***}	-0.566^{***}	-0.0341^{***}	-0.401^{**}
	(-7.11)	(-5.46)	(-7.29)	(-2.68)
Satisfaction with leisure	0.0138^{*}	0.263^{***}	0.0139^{**}	0.208
	(2.25)	(3.35)	(3.01)	(1.76)
Satisfaction with dwelling	0.00636	0.176	-0.00291	-0.055
	(0.84)	(1.80)	(-0.53)	(-0.42)
Household crowding index	-0.0331	-0.418	-0.0826^{*}	-0.906
	(-0.70)	(-0.76)	(-2.08)	(-0.95)
Business sector dummies	Included	Included	Included	Included
Region dummies	Included	Included	Included	-
Year dummies	Included	Included	Included	Included
constant	1.962^{***}	10.48^{*}	-0.893^{**}	47.38^{***}
	(5.23)	(2.38)	(-2.72)	(4.14)
Number of observations	31,567	31,567	31,567	31,567
Number of groups			6,459	6,459

Notes: *t* statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.4
Robustness checks. Source: SOEP 2002 – 2011.

	(Baseline) Sample used for Table 3	(i) Female	(ii) Male	(iii) Excluding public sector	(iv) Rural region	(v) Urban Region	(vi) Excluding 'sickness absence days outliers'	(vii) Excluding small distance changes	(viii) Commuting distance as dummy variable	(ix) Commuting distance as log-linear specification	(x) Including employer change
Short distance	0.0442	0.0366	0.0555	0.0277	-0.0049	0.0623*	0.0499	0.0264			0.0398
commuter	(1.91)	(1.12)	(1.68)	(0.93)	(-0.10)	(2.37)	(1.93)	(1.00)			(1.74)
Middle distance	0.109***	0.1208**	0.102*	0.1537***	0.2352***	0.0639	0.1137***	0.0785*			0.110***
commuter	(3.65)	(2.65)	(2.54)	(4.04)	(3.75)	(1.86)	(3.39)	(2.38)			(3.72)
Long distance	0.191***	0.162*	0.212***	0.2586***	0.2425**	0.1671**	0.2238***	0.1546***			0.181***
commuter	(4.35)	(2.19)	(3.84)	(4.66)	(3.05)	(3.14)	(4.56)	(3.35)			(4.24)
Commutes: 50 km and more (ref.: 0 – 49 km)									0.1423*** (3.44)		
Commuting distance (in log)										0.0461*** (5.01)	
No. obs.	31,567	14,942	16,625	20,279	7,673	23,848	28,968	25,370	31,567	31,567	32,178
No. of groups	6,459	3,068	3,391	4,334	1,538	4,919	6,115	5,996	6,459	6,459	6,540

Table A.4 cont.
Robustness checks. Source: SOEP 2002 – 2011.

	(xi) Including residence change	(xii) Including employer and residence change	(xiii) 1-year-lag	(xiv) 2-year-lag	(xv) Dependent variable: incidence of absence
Short distance	0.0381	0.0360	0.0315	0.0046	0.0450
commuter	(1.90)	(1.84)	(1.01)	(0.13)	(1.75)
Middle distance	0.0884***	0.0927***	0.1320***	0.0602	0.0780*
commuter	(3.41)	(3.66)	(3.23)	(1.29)	(2.32)
Long distance	0.177***	0.175***	0.2334***	0.1990**	0.1410**
commuter	(4.74)	(4.84)	(3.69)	(2.83)	(2.82)
No. obs.	39,181	40,289	18,369	14,321	31,567
No. of groups	7,453	7,601	4,279	3,419	6,459

Notes: Only the coefficients for the commuting variables are reported. Models (i) – (xiv) are fixed-effects negative binomial models with the number of 'days absent' as dependent variable. Model (xv) is a random effects probit model with the 'incidence of absence' as dependent variable. In models (i) – (vii) and models (x) – (xv) non-commuters are treated as the reference category. Like in the main table, all control variables are included in all specifications. All models are estimated using robust standard errors. *t* statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure A.1
Distribution of absence days. Source: SOEP 2002 – 2011.

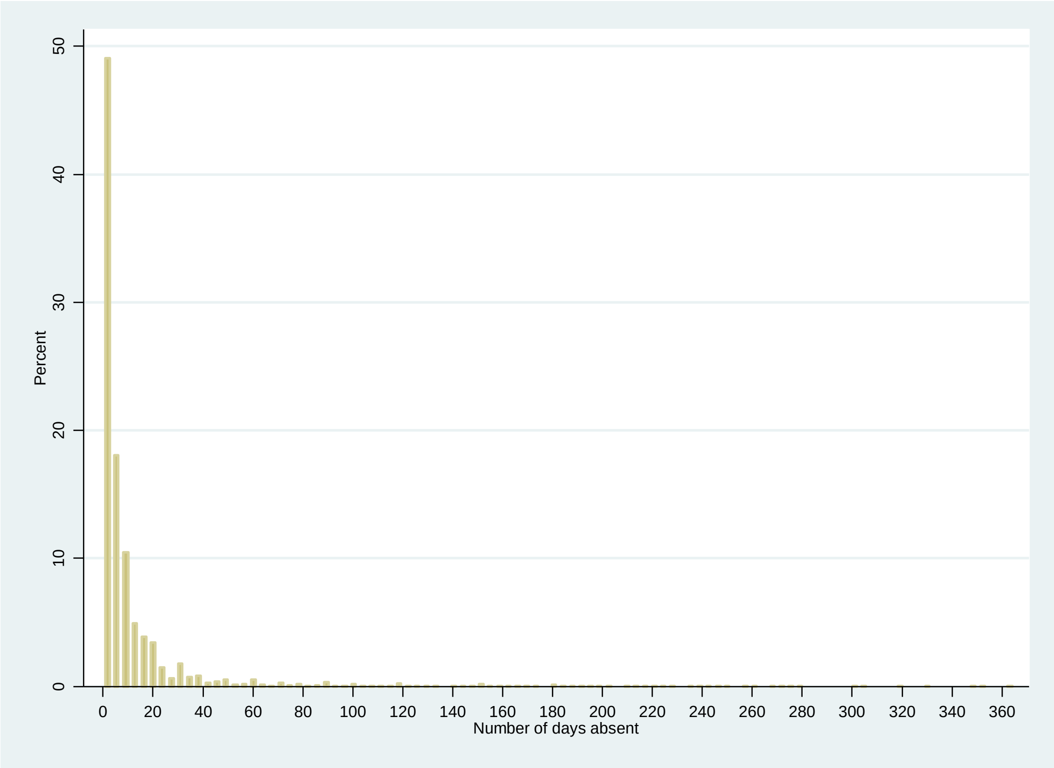

Figure A.2
Distribution of commuting distances. Source: SOEP 2002 – 2011.