Research Data Centre (FDZ)
of the German Federal
Employment Agency (BA)
at the Institute for
Employment Research (IAB)

FDZ-Methodenreport 01/2018 EN

Methodological aspects of labour market data

Calculation of Pension Entitlements in the Sample of Integrated Labour Market Biographies (SIAB)

Mona Pfister,
Svenja Lorenz,
Thomas Zwick

Bundesagentur für Arbeit

Calculation of Pension Entitlements in the Sample of Integrated Labour Market Biographies (SIAB)

Mona Pfister, Svenja Lorenz and Thomas Zwick (University of Wuerzburg)

Documentation version: DOI: 10.5164/IAB.FDZM.1801.en.v1

Die FDZ-Methodenreporte befassen sich mit den methodischen Aspekten der Daten des FDZ und helfen somit Nutzerinnen und Nutzern bei der Analyse der Daten. Nutzerinnen und Nutzer können hierzu in dieser Reihe zitationsfähig publizieren und stellen sich der öffentlichen Diskussion.

FDZ-*Methodenreporte* (FDZ method reports) deal with methodical aspects of FDZ data and help users in the analysis of these data. In addition, users can publish their results in a citable manner and present them for public discussion.

Content

Abstract

We describe a method to calculate individual pension entitlements (earning points, *Entgeltpunkte*) based on information available in conventional employment history data. Pension entitlements can be seen as an important driver of retirement decisions and of the labour market attachment of older people. Therefore, the calculation of pension entitlements using labour market history datasets improves the usability of these datasets for a broad range of topics around the decision to retire. In the first part of this report, we use a high-quality administrative biographical dataset with linked information on pension entitlements, covering a sample of almost all employees in Germany (Biographical Data of Selected Social Insurance Agencies in Germany, BASiD). Based on the BASiD, we explain which information is needed for the calculation of earning points and outline potential sources of error in this calculation.

In a next step, we implement our method to calculate earning points in a larger administrative dataset with only conventional employment history information (Sample of Integrated Labour Market Biographies, SIAB). We describe our calculations and assess possible sources of error by mimicking the SIAB data structure in the BASiD. The average deviation for annual earning points is only around 1%, and the error for the sum of the earning points is around 7%. Most of these errors can be explained by large observation gaps in individual pension contributions during phases of reduced employment such as parental leave, marginal employment or unemployment. There might be additional small errors for the implementation of public compensation payments during these periods as a result of specific legal rules for certain socio-economic groups and minor rounding errors during all periods. Finally, we calculate the earning points using the SIAB and show descriptive statistics for annual and total pension entitlements for the entire sample and selected subgroups. We conclude that our approach to calculating pension entitlements using conventional employment history data opens various new and important research options at least for employees without large employment gaps.

Zusammenfassung

Wir beschreiben eine Vorgehensweise zur Berechnung der individuellen Rentenansprüche (Entgeltpunkte) auf Basis von Datensätzen mit allgemeinen Informationen aus der Erwerbsbiografie. Die Höhe der Rentenansprüche kann als ein wichtiger Treiber der Renteneintrittsentscheidung und der Arbeitsmarktaktivität im Alter angesehen werden. Für viele Fragestellungen kann es demnach hilfreich sein, die Höhe der Entgeltpunkte pro Person in einem administrativen Datensatz über individuelle Arbeitsmarktkarrieren berechnen zu können, der keine spezifischen Angaben zu Rentenansprüchen enthält. Im ersten Teil des Reports nutzen wir einen hochqualitativen administrativen Datensatz mit Informationen zu individuellen Erwerbsverläufen und Rentenansprüchen einer Stichprobe fast aller abhängig Beschäftigten in Deutschland (Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland, BASiD). Wir erklären, welche Informationen für die Berechnung der Entgeltpunkte notwendig sind und skizzieren mögliche Fehlerquellen bei deren Berechnung.

Im zweiten Teil des Reports übertragen wir die vorher beschriebene Methode zur Berechnung der Entgeltpunkte auf einen umfangreicheren administrativen Datensatz mit lediglich konventionellen Informationen über Arbeitsmarktkarrieren (Stichprobe der Integrierten Arbeitsmarktbiografien, SIAB). Auch hier beschreiben wir unsere Berechnungen und die dabei auftretenden Fehler, indem wir die Datenumgebung des SIAB im BASiD nachbauen. Die durchschnittliche Abweichung der jährlichen Entgeltpunkte beträgt lediglich 1%, der Fehler bei der Berechnung der Summe der Entgeltpunkte liegt bei 7%. Ein Großteil der Abweichungen kann durch lückenhafte Rentenbeitragszahlungen während Phasen niedrig bezahlter Beschäftigung, wie Elternzeit, geringfügiger Beschäftigung oder Arbeitslosigkeit erklärt werden. Zusätzlich sind kleinere Fehler bei öffentlichen Extrazahlungen während dieser Perioden und aufgrund spezifischer Rechtsvorschriften für bestimmte sozioökonomische Gruppen nicht auszuschließen. Außerdem kommen kleine und nicht systematische Rundungsfehler hinzu. Abschließend berechnen wir die Entgeltpunkte mit dem SIAB und zeigen deskriptive Statistiken der jährlichen und kumulierten Rentenansprüche für das gesamte Sample und Untergruppen. Wir können somit zeigen, dass unser Verfahren zur Berechnung von Pensionsansprüchen zahlreiche neue und wichtige Forschungsansätze mit konventionellen Daten über Beschäftigungsbiografien zumindest für Beschäftigte ohne lange Erwerbslücken ermöglicht.

Keywords: calculation of pension entitlements, earning points, employment history, retirement decision, Germany

We would like to thank Dana Müller, Daniela Hochfellner, and Regina Herz for their helpful comments. Work on this report was partially financed by the German Research Foundation (DFG, grant number ZW172/3-1). This study uses the Biographical Data of Selected Insurance Agencies in Germany (BASiD) (Version 1951–2009) and the Sample of Integrated Labour Market Biographies (SIAB) (Version 1975–2014). Data access was provided initially via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently through remote data access.

1 Introduction

Research on retirement issues is gaining in importance for developed countries as the effects of demographic change become increasingly difficult to ignore. In particular, the ageing population and the reduction in cohort sizes are challenging the sustainability of the German pay-as-you-go pension system. The recent underfunding of the pension system has further elevated the need for measures designed, for example, to decrease pension entitlements, increase tax subsidies for pensions and increase older employees' labour market attachment. Improving the sustainability of the pension system has never been off of the policy agenda, but it now has a more important place in this agenda than ever before.

Retirement in Germany is possible at two statutory ages, the early retirement age (ERA) and the normal retirement age (NRA). Most studies interested in the determinants of retirement decisions take into account legal rules on retirement age and changes in these, as well as considering the influence of the social security system and the individual financial incentives it generates. Changes in pension entitlements during the last years of employment before retirement and the sum of pension entitlements seem to play an especially crucial role in the retirement decision and labour attachment in older age.[1] For example, in their seminal article, STOCK AND WISE (1990) propose an empirical method to identify individual financial incentives to retire at a certain age. Their model combines (mainly financial) features of the social security system with individual attributes that may simultaneously affect retirement decisions. More specifically, they identify additional pension entitlements from continuing work plus labour income in comparison to the accumulated pension entitlements as the main incentive mechanism for retirement. The validity of their approach has been empirically verified by several empirical papers and in several countries (DE LA CROIX ET AL. (2013), GRUBER AND WISE (1999), and FISHER AND KEUSCHNIGG (2010)). They showed that the pension system offers notable financial incentives to early retirement before the NRA. Their results were confirmed by BÖRSCH-SUPAN (1992), RIPHAHN AND SCHMIDT (1995), and BÖRSCH-SUPAN ET AL. (2004) for Germany. Accordingly, ARENT AND NAGL (2010) found a strong correlation between the financial arrangements of the pension system and the retirement entry age in 16 countries.

BERKEL AND BÖRSCH-SUPAN (2004), HANEL (2010), and ENGELS ET AL. (2017) pointed out that the German pension reform of 1992 increased the average retirement age by introducing deductions from pension entitlements for early retirement. ENGELS ET AL. (2017) showed that the decrease in early retirement options without reductions in old-age pension entitlements increased the retirement age among women (who were affected by 15 months), and the duration of employment by 15 months.[2] HANAPPI (2012) and MANOLI AND WEBER (2016) confirmed this mechanism and analysed the modifications to retirement benefits created by changes in policy rules (such as additional bonuses and deductions from pension entitlements)

[1] Cf. Stock and Wise (1990), p. 1158.
[2] Cf. Engels et al. (2017), p. 226.

for workers in Austria. They showed that financial incentives have an important influence on individuals' pension behaviour.[3]

In addition to the determinants of individuals' retirement age, some studies have analysed factors that promote working after the NRA. Not only the labour force participation rate, but also the number of individuals aged at least 65 years who still work, increased continuously.[4] There are different explanations for working after retirement proposed in the literature. Again, financial incentives play a crucial role for the identified mechanisms. A first important push factor can be a relatively low income level during an individual's working life, which may be translated into low pension entitlements. In this case, working after the NRA is necessary to maintain one's standard of living, as has been pointed out by BURKERT AND HOCHFELLNER (2017) and HOCHFELLNER (2014). In Germany, similarly to other countries such as Sweden and Denmark, the increased share of people working after retirement can also be attributed to financial incentives introduced through political reforms (e.g., percentage supplements to pension entitlements) (LARSEN AND PEDERSEN (2017)). Second, most employees start to work again in their previous companies after their retirement because they are used to the social contact they have with colleagues and to the enjoyment of work.[5] The initiative to work again after retirement therefore usually comes from the employees themselves. This behaviour can be observed especially among older employees with high salaries and high pension entitlements.[6]

Other literature that concentrates on differences among socio-economic groups in Germany discusses the impact of individual employment histories on pension entitlements. It shows that, above all, the influence of employment gaps caused by unemployment and maternity lead to large pension entitlement losses, despite public compensation payments (see FITZENBERGER ET AL. (2001), BOOCKMANN AND STEINER (2006), and POTRAFKE (2011) for Germany and NEUGSCHWENDER (2014) for five other European countries). The effects of employment interruptions on pension entitlements differ strongly by country and by social security system, however.[7] In addition, HONEKAMP AND SCHWARZE (2010), KLUTH AND GASCHE (2013), and RASNER (2014) show that low-skilled and part-time workers as well as women in Western Germany gain lower entitlements than do their respective comparison groups.[8]

Another group of studies looks at pension entitlements for specific groups of people over time. These articles, for example, identify differences in entitlements between men and women, and among employees with different workplaces, work arrangements and levels of education, as

[3]Cf. Hanappi (2012), pp. 32-33; Manoli and Weber (2016), p. 179.
[4]Cf. Steiner (2017), pp. 1-2.
[5]Cf. Burkert and Hochfellner (2017), p. 156.
[6]Cf. Lewicki (2014), pp. 118–120.
[7]Cf. Fitzenberger et al. (2001), pp. 13–14; Boockmann and Steiner (2006), pp. 1149–1151; Potrafke (2011), p. 234; Neugschwender (2014), p. 32.
[8]Cf. Honekamp and Schwarze (2010), p. 223; Kluth and Gasche (2013), pp. 45–47; Rasner (2014), p. 50.

well as across birth cohorts.[9] KORTMANN AND SCHATZ (1999), BIEBER AND STEGMANN (2000), SCHATZ ET AL. (2002), and HIMMELREICHER AND FROMMERT (2006), for instance, show that cohorts born after 1950 differ notably from older cohorts (especially those born before 1940). The average pension entitlement among men decreases over time in Western and Eastern Germany. Younger female cohorts, however, especially in Eastern Germany, show a moderate increase in pension entitlements.[10] A reason for the increase in women's pension entitlement levels is the incidence of part-time work; women increasingly work part-time instead of interrupting their employment for extended periods, and they can thus increase the working time during their careers. This higher labour market attachment enables younger women to accumulate more pensionable periods and acquire more pension entitlements.[11] GEYER AND STEINER (2014) analyse both the development of pension entitlements in the past and possible future pension entitlements in Germany. They show that young cohorts (born in 1960 onwards) in Eastern Germany and low-skilled employees have the lowest pension entitlements.[12] They point to high unemployment rates as a major cause of low pension entitlements for these groups.[13] They also find that the increasing employment participation of women has a positive effect on the development of their pension entitlements.[14]

The German pension system calculates pension entitlements roughly proportional to lifetime earnings. The replacement rate (*Ersatzrate*)[15] – monthly pension entitlements in comparison to a worker's last income – was approximately 51% in 2016.[16] Public pension benefits are the most important source of income (approximately 75%) for the majority of the population.[17] For this reason, pension entitlements have been used in many papers on the labour market attachment of older people. This information is, however, not available in most of the conventional datasets that are widely used for labour market research. This is one of the reasons why most articles on the labour market attachment of older employees use specific datasets that include pension entitlements, such as the Biographical Data of Selected Social Insurance Agencies in Germany (BASiD) or the Sample of Insured Persons and Their Insurance Accounts (*Versicherungskontenstichprobe*, VSKT), or construct pension

[9]Cf. Kortmann and Schatz (1999), p. 576; Bieber and Stegmann (2000), p. 368; Schatz et al. (2002), p. 234.
[10]Cf. Kortmann and Schatz (1999), p. 582; Bieber and Stegmann (2000), pp. 380–381; Schatz et al. (2002), p. 249.
[11]Cf. Kortmann and Schatz (1999), pp. 582–583; Bieber and Stegmann (2000), pp. 380–383; Schatz et al. (2002), p. 250.
[12]Cf. Geyer and Steiner (2014), pp. 182–184.
[13]Cf. Geyer and Steiner (2014), pp. 189–190.
[14]Cf. Geyer and Steiner (2014), p. 191.
[15]The replacement rate is not to be confused with the usual pension level (*Rentenniveau*), which compares the average individual pension entitlement with the average income of all German workers in the same year.
[16]Cf. OECD (2017), p. 109.
[17]In eastern Germany, the share of 97% is even much higher than in western Germany at 70%. Other income options during retirement are company pensions (approximately 10%), private pensions (approximately 9%) and public transfers for individuals whose income is below the social assistance level (cf. Geyer and Steiner (2014), p. 184; Deutsche Rentenversicherung Bund (2017), pp. 70–71; BMAS (2016), p. 14).

entitlements from income measures, for example, in the German Socioeconomic Panel (*Deutsches Sozio-ökonomisches Panel,* GSOEP). Survey datasets, however, usually are smaller, cover shorter observation periods, have not repeated observations of the people asked and therefore do not offer precise longitudinal information on key variables such as earnings and labour market states. These are reasons why empirical research in labour economics increasingly uses administrative labour market history datasets (CARD ET AL. (2010); GÜRTZGEN AND NOLTE (2017)). We argue that the BASiD and VSKT also have drawbacks that restrict the options for analysing relevant questions on retirement behaviour. For example, these datasets have relatively few observations in comparison to conventional administrative labour market history data, include only a limited list of variables beyond the labour market history of employees, and are representative only of selected birth cohorts.[18]

In this report, we present a method to calculate individual pension entitlements for employees in Germany using SIAB, a large dataset with conventional longitudinal administrative information on individual labour market biographies that is available for scientific research. We also use the BASiD to validate our procedure.[19] The BASiD data include full employment biographies of individuals, retirement entry, and, most importantly, the actual pension entitlements as well as information, for example, on labour market state necessary to calculate the entitlements.[20] In the first part of the report, we calculate earning points from year to year, as well as the sum of entitlements according to the rules of the German Social Code (*Sozialgesetzbuch,* SGB) with BASiD data that include employees with no long labour market gaps (BASiD Sample I). We show which information is necessary to obtain the yearly and the sum of all pension entitlements.

In the second part of this report, we exemplarily implement our method in a large conventional administrative dataset, the Sample of Integrated Labour Market Biographies (SIAB). The SIAB has the advantage with respect to the BASiD that it is representative of employees from more birth cohorts and more retirement years. First, we identify possible sources of error in our calculation of the pension entitlements. We mimic the SIAB data structure in BASiD by excluding observations before 1975 (BASiD Sample II). The quality is assessed by comparing the results of our calculation method in the BASiD Sample II with the actual administrative earning points. Afterwards, we perform the calculations in the full SIAB sample and present descriptive statistics on the resulting earning points.[21]

We show how errors in our calculation method can be decreased by gradually eliminating particular subgroups from our BASiD Samples I and II. We start with full samples and restrict them, in a first step, to individuals who had worked only in Western Germany because labour

[18]BASiD is representative for people who retire in the year 2007. This means that mainly the birth cohorts 1940–1942 are represented in the dataset and we therefore concentrate on these three birth cohorts (cf. Burkert and Hochfellner (2017), p. 152).
[19]We use a customized add-on to BASiD and SIAB, namely the day of birth.
[20]Cf. Hochfellner et al. (2011), p. 11.
[21]See Appendix B for a detailed description of the characteristics of the complete BASiD, BASiD Sample I, and BASiD Sample II.

market histories of employees in Eastern Germany are only observable since 1990. In the next step, people with large employment gaps in their work histories (more than 5 years for men and more than 10 years for women) are excluded. In the last step, all women are deleted because they have greater gaps in their employment histories than do their male counterparts. Such gaps are primarily caused by maternity, parental leave, caregiving on a non-commercial basis, and marginal employment without contributions to the statutory retirement insurance. These are periods during which some people pay voluntary pension entitlement contributions or (to a smaller extent) receive public pension entitlements that are problematic to calculate because the relevant legal rules are exceedingly difficult or even impossible to reconstruct.

Two problems may arise when using our calculation method. First, errors may occur when applying the calculation. These could be rounding errors or inaccuracies in the calculation of the earning points for individuals with elusive employment histories. Information gaps such as periods of illness or time spent out of labour force may cause errors in the calculation of earning points. Periods with two or more simultaneous pension-relevant activities or reduced employment hours (for example during parental leave, unemployment, marginal employment, or, especially, voluntary individual or employer pension insurance payments) are mainly associated with deviations in the calculation of pension entitlements. We can, however, show that the measurement errors associated with these information gaps are negligible for employee groups our report covers. Second, in the SIAB, we have information gaps in comparison with the BASiD. For example, we cannot observe complete employment histories for older cohorts because the SIAB starts in 1975 (the BASiD includes information on labour market histories since 1951). In addition, some variables necessary for the classification of the pension type or the calculation of the pension entitlement for specific groups (such as pensionable periods, the identification of severe disability, the number of children or periods of maternity protection) are not directly available in the SIAB. Furthermore, we do not have information about statutory or factual retirement dates (and the resulting retirement ages). With information from the individual employment history data and the legal pension rules, we can identify the earliest statutory retirement dates (ERA and NRA) each individual is eligible for. These dates determine the deductions and supplements associated with each factual exit of the labour market date. We then demonstrate with a sensitivity analysis that, measurement errors in the pension entitlement calculation caused by information gaps on retirement age are negligible for key employee groups.

The remainder of this article is organised as follows. We start in Section 2 with an overview of the German pension system and the relevant pension reforms for the calculation of the pension entitlements. In Section 3, we describe our two datasets. Section 4 examines the quality of our calculations. Section 5 presents sources of and reasons for errors in the calculation of pension entitlements and identifies socio-economic groups for whom the estimates exhibit a high degree of accuracy or bias. A discussion of the results follows in Section 6. We conclude in Section 7.

2 Institutional Background: The German Pension System and Its Changes

Introduced in 1889 as part of the Bismarckian social insurance system, the German public pension system *(gesetzliche Rentenversicherung)* has existed for more than 125 years and is one of the oldest pension systems in the world.[22] It is one of the pillars of the social security system, guaranteeing social security for people exiting the labour force at an old age and in cases of occupational disability, and providing financial insurance for the surviving dependents in the event of the death of the sole earner in a household.[23] For the purpose of our analysis, we concentrate on the old-age pension system and its changes for employees who worked from 1951 to 2007 and started retirement in 2000 or later.[24]

Legal pension entitlement rules

To calculate the pension benefits, we demonstrate that it is sufficient to use standard information obtained from employment histories of individuals in the dataset and the information from the German SGB. Pension benefits are calculated according to the following formula:

$$PB_{T+s} = \left(\sum_{t=1}^{T} EP_t\right) \times PT_T \times AF_T \times CPV_{T+s},\ s = 0, 1, \ldots, S,\quad t = 1, \ldots, T.$$

PB is the individual monthly pension benefit, ***EP*** the earning points, ***PT*** the pension type factor, ***AF*** the age factor, ***CPV*** the current pension value, **t** the years of contribution before entering retirement (***T***) and ***s*** the years after retirement until death at (***S***). The monetary pension entitlement is a product of the sum of the earning points (*Entgeltpunkte*) and the annually adjusted current pension value (*aktueller Rentenwert*), differentiated between Eastern and Western Germany. To calculate the earning points earned per year, one has to sum the individual gross earnings per year and relate them to the average gross earnings of all German employees for the respective year reported in official statistics (see Table 1).[25] The sum of individual gross earnings per year is only taken into account for public pension entitlements up to the so-called pension contribution limit (*Beitragsbemessungsgrenze*). Therefore, the pension contribution limit represents the maximum of the earning points to be achieved per year. In the last step, the sum of individual earning points is multiplied by the pension type factor (*Rentenartfaktor*) and the age factor (*Zugangsfaktor*).

[22]Cf. Deutsche Rentenversicherung Bund (2014), p. 82; see Börsch-Supan and Schnabel (1999) and Börsch-Supan and Wilke (2004) for a more detailed description of the German pension system.
[23]Cf. Deutsche Rentenversicherung Bund (2016b).
[24]The period 1951–2007 is the maximum observation period on pension information available in the BASiD; see Chapter 3.
[25]In case of periods of several simultaneous jobs, the earnings of all jobs have to be summed. Special bonuses (Christmas allowance; vacation allowance) are also part of the annual gross earnings and are therefore converted into earning points.

First, we need to understand the rules of the German pension insurance system that determine the largest part of pension entitlements and the changes in these entitlements. The rules are explained in the Sixth Book of the German SGB (*SGB VI*)[26]:

- NRA by birth cohort;
- ERA by birth cohort;
- requirements for the different pension types;
- pension contribution limit for the calculation of the individual earning points per year;
- average gross earnings of all German employees per year (this average value has to be set in relationship to the individual gross earnings);
- conversion factors for the values for Eastern German employees;
- pension type factor;
- age factor, which includes deductions from and supplements to earning point sums for those who retire earlier or later than the NRA[27]; and
- adjustments to the calculation of the earning points because of parental leave, care-giving periods, unemployment and low income.

To calculate pension entitlements, we have to know the actual retirement date and the NRA as determined by the law for every old-age pension type.[28] The pension law also stipulates the pension contribution limit, at which the maximum entitlement in earning points per year is obtained, and the relationship between earnings and earning points. Table 1 shows, for example, that the maximum number of earning points that could be obtained in 2007 was 2.1034. This maximum was reached at the pension contribution limit corresponding to a yearly income of €63,000.00 in Western Germany and €54,600.00 in Eastern Germany. Income beyond the pension contribution limit does not contribute to pension entitlements. A person working in Western Germany with a gross income of €29,951.00 in 2007, which is exactly the average earnings of all employees in Germany, accordingly obtained one earning point. The lowest number of maximum points obtainable per year was in 1974, at 1.4720 earning points (which corresponded to a yearly income of €15,338.76). The pension contribution limit and the average earnings per year for Eastern Germany are products of the values from the Western German relation between earnings and entitlements and the so-called conversion factor. Thus, Eastern and Western German employees are able to reach the same maximum earning points per year.

[26]Regulations of pension insurance can be found in the Sixth Book of the *SGB*. Since 1992, the *SGB* has combined the employees' '*Angestelltenversicherungsgesetz*', the '*Reichsversicherungsordnung*' for workers and the '*Reichsknappschaftsgesetz*' for insured persons in the miners' association.
[27]Retirement at the NRA causes no deduction in pension entitlements, so the age factor is one.
[28]We only calculate the pension entitlements for already retired individuals in the dataset because most applications are based on the sum of pension entitlements collected. Individuals without retirement date information are deleted in both datasets (see Chapter 3).

Table 1. Values for the pension calculation from the German Social Code (SGB), 1951–2014

Year	Conversion factor (Eastern into Western German entitlements)	Pension contribution limit, Western Germany (in €)	Pension contribution limit, Eastern Germany (in €)	Average gross earnings, Western Germany (in €)	Average gross earnings, Eastern Germany (in €)	Maximum earning points
1951		3681.23		1829.91		2.0117
1952 until 31 August		3681.39		1969.50		1.8692
1952 from 1 September		4601.54		1969.50		2.3364
1953		4601.63		2076.36		2.2162
1954		4601.52		2164.81		2.1256
1955		4601.65		2325.36		1.9789
1956		4601.71		2476.70		1.8580
1957		4601.63		2578.44		1.7847
1958		4601.63		2725.19		1.6886
1959		4908.40		2864.26		1.7137
1960		5215.18		3119.39		1.6719
1961		5521.95		3437.42		1.6064
1962		5828.73		3746.75		1.5557
1963		6135.50		3975.29		1.5434
1964		6749.05		4329.11		1.5590
1965		7362.60		4718.71		1.5603
1966		7976.15		5058.21		1.5769
1967		8589.70		5224.89		1.6440
1968		9816.80		5543.43		1.7709
1969		10,430.35		6053.18		1.7231
1970		11,043.90		6822.17		1.6188
1971		11,657.45		7634.10		1.5270
1972		12,884.56		8351.95		1.5427
1973		14,111.66		9354.08		1.5086
1974		15,338.76		10,420.64		1.4720
1975		17,179.41		11,150.25		1.5407
1976		19,020.06		11,931.00		1.5942
1977		20,860.71		12,754.18		1.6356
1978		22,701.36		13,417.32		1.6919
1979		24,542.01		14,155.12		1.7338
1980		25,769.11		15,075.44		1.7093
1981		26,996.21		15,798.92		1.7087
1982		28,836.86		16,462.58		1.7517
1983		30,677.51		17,022.44		1.8022
1984		31,904.61		17,533.22		1.8197
1985		33,131.71		18,041.45		1.8364
1986		34,358.81		18,727.09		1.8347
1987		34,972.36		19,289.00		1.8131
1988		36,813.02		19,887.21		1.8511
1989		37,426.57		20,483.89		1.8271
1990 from 1 July	2.3473	38,653.67	16,565.86	21,446.65	9136.73	1.8023
1991 until 30 June	1.7235	39,880.77	18,406.51	22,712.10	13,177.89	1.7559
1991 from 1 July	1.7235	39,880.77	20,860.71	22,712.10	13,177.89	1.7559

(continued)

(Table 1 continued)

1992	1.4393	41,721.42	29,450.41	23,938.69	16,632.17	1.7428
1993	1.3197	44,175.62	32,518.16	24,633.02	18,665.62	1.7933
1994	1.2687	46,629.82	36,199.47	25,125.91	19,804.45	1.8558
1995	1.2317	47,856.92	39,267.22	25,904.60	21,031.58	1.8474
1996	1.2209	49,084.02	41,721.42	26,422.54	21,641.86	1.8577
1997	1.2089	50,311.12	43,562.07	26,660.29	22,053.35	1.8871
1998	1.2113	51,538.22	42,948.52	27,060.12	22,339.74	1.9046
1999	1.2054	52,151.77	44,175.62	27,357.69	22,695.95	1.9063
2000	1.2030	52,765.32	43,562.07	27,740.65	23,059.56	1.9021
2001	1.2003	53,378.87	44,789.17	28,231.49	23,520.36	1.8908
2002	1.1972	54,000.00	45,000.00	28,626.00	23,910.79	1.8864
2003	1.1943	61,200.00	51,000.00	28,938.00	24,230.09	2.1149
2004	1.1932	61,800.00	52,200.00	29,060.00	24,354.68	2.1266
2005	1.1827	62,400.00	52,800.00	29,202.00	24,690.96	2.1368
2006	1.1827	63,000.00	52,800.00	29,494.00	24,937.85	2.1360
2007	1.1841	63,000.00	54,600.00	29,951.00	25,294.32	2.1034
2008	1.1857	63,600.00	54,000.00	30,625.00	25,828.62	2.0767
2009	1.1712	64,800.00	54,600.00	30,506.00	26,046.79	2.1242
2010	1.1726	66,000.00	55,800.00	31,144.00	26,528.11	2.1192
2011	1.1740	66,000.00	57,600.00	32,100.00	27,342.42	2.0561
2012	1.1785	67,200.00	57,600.00	33,002.00	28,003.39	2.0362
2013	1.1762	69,600.00	58,800.00	33,659.00	28,616.73	2.0678
2014	1.1665	71,400.00	60,000.00	34,514.00	29,587.66	2.0687

(Source: Deutsche Rentenversicherung Bund (2016c), pp. 258–272)

Pathways to retirement and their changes

In this report, we restrict our sample to workers with an old-age pension in Germany. Therefore, the pension type factor is always one.[29] We exclude the old-age pension for miners, pensions for surviving dependants and reduced earnings capacity pensions and the corresponding people as far as we can identify them because we do not have the information needed for the specific construction of these pension entitlements in conventional administrative datasets. Thus, we take into account five types[30] of old-age pensions in the German pension system:

1 the regular old-age pension;
2 the old-age pension for long-term insured employees;
3 the old-age pension for the unemployed or those under a progressive retirement plan;
4 the old-age pension for women; and
5 the old-age pension for severely disabled individuals.[31]

In the following, we sketch the development of the ERA and NRA for the relevant pension types. The statutory pension dates not only determine at which age an individual can retire at the earliest but they also imply deductions or supplements that are relevant for pension

[29]Cf. Arent and Nagl (2010), p. 421; Rüb and Lamping (2010), p. 155; Social Security Administration (2014), p. 115.
[30]The pension for persons with an exceptionally long insurance record (*Rente für besonders langjährig Versicherte*) is not included in our report because this pension type has only existed since 2012.
[31]Cf. Social Security Administration (2014), pp. 114–121.

entitlements. Until 1972, the pension system included three different old-age pension types, and retirement was mandatory at age 65 for men and at age 60 for women and unemployed individuals. As shown in Table 2, the pension reform in 1972 provided for the first time the possibility for long-term insured employees to retire after their 63rd birthday when they had more than 35 years in pensionable periods. The reform also allows women to retire at age 60 when they have at least 15 years of pension contributions and more than 10 contribution years after their 40th birthday. Unemployed people can also retire when they are at least 60 years old. Individuals with a normal old-age pension have to be 65 years old. To be entitled to an old-age pension, all workers have to fulfil at least 15 years of contributions (*Wartezeit*). Early retirement is not possible with the regular old-age pension, even with deductions.[32]

As a reaction to the ageing of the German population and the shift in the dependency rate between employees and people in retirement, pension reforms starting in 1992 increased the NRA gradually to 65 years for birth cohorts of 1937 and later, except for people entitled to pensions for the severely disabled.[33] The increase in NRA, however, took place very differently, depending on the pension type.[34] Therefore, the attractiveness of the pension types differs from period to period.[35] In 1984, the minimum number of contribution years necessary for eligibility was reduced to five.[36]

The paragraph § 236 SGB VI allows **long-term insured employees** born before 1 January 1937 to retire after their 63rd birthday. In addition, they have to have at least 35 contribution years to get the full pension entitlement. The increase of the NRA to 65 years is accomplished as follows. The pensionable age is 63 years and one month for those born in January 1937. The pensionable age is increased with every subsequent month of birth from January 1937 onwards. As a result, employees born in December 1938 have to be 65 years old to retire. The ERA is constant at 63 years.[37]

The old-age pension described in § 237 SGB VI allows **unemployed people and those under a progressive retirement plan**[38] to retire when they are 60 years old for birth cohorts of 1936 and later.[39] The following conditions have to be fulfilled to be eligible for this pension type: First,

[32]Cf. Deutsche Rentenversicherung Bund (2014), p. 54; Börsch-Supan et al. (2015), p. 811.
[33]Cf. Deutsche Rentenversicherung Bund (2014), p. 68.
[34]In a companion paper, we describe a strategy for the identification of the pensionable periods in conventional labour market history data and we determine whether an individual is eligible to obtain one of the old-age pension types. Eligibility for a pension types then allows us to identify statutory retirement dates (ERA and NRA) and individual paths of older employees out of the labour market. In addition, we show an extended list of the increases in the NRA and ERA by pension type, birth cohort, and birth month (cf. Lorenz et al. (2018)).
[35]In this chapter, we describe the 1936–1947 birth cohorts because we can observe the employment histories only until 2007, and the earliest possible retirement age is 60 years.
[36]Cf. Deutsche Rentenversicherung Bund (2014), p. 54; Börsch-Supan et al. (2015), p. 811.
[37]Cf. SGB (n. d.), § 236.
[38]'Under a progressive retirement plan' means working part-time in old age. This is possible for workers aged 55 years or older. The programme allows older employees to work a reduced number of hours. Income losses are partially replaced by a government subsidy. The programme was terminated in 2009. More detailed explanations can be found in AltersTZG (n. y.), §§ 1–16.
[39]Cf. SGB (n. d.), § 237.

all individuals must have 15 contribution years. Second, they have to be unemployed when they enter retirement and must have been unemployed for at least 52 weeks in total after reaching 58.5 years of age. Alternatively, individuals have to have been in partial retirement for 24 months and have at least eight years in pensionable periods during the 10 most recent years before retirement.[40] Starting with 60 years and one month for the January 1937 birth cohort, the NRA is increased by one month with every subsequent month of birth. As a consequence, individuals born in December 1941 or later have to be 65 years old to be eligible for the pension after unemployment or partial retirement. The ERA is 60 years for birth cohorts of January 1946 or earlier. Thereafter, it increases by one month with every subsequent month of birth. As a result, the ERA for birth cohorts after December 1948 is 63 years.[41]

Table 2. NRA by old-age pension type in Germany

Pension type	Retirement age[a]				Contribution years necessary for eligibility	Additional conditions
	1972 reform	1992 reform	1999 reform	2007 reform[b]		
Regular old-age	65	65	65	67	5[c]	
Long-term insured employees	63	65	65	67	35	
Unemployed or under a progressive retirement plan	60	65	-	-	15	Born before 1952; at least 52 weeks in unemployment after reaching 58.5 years of age or partial retirement for 24 months before retirement
Women	60	65	-	-	15	Born before 1952; 10 contribution years after the 40th birthday
Severely disabled	60	60	63	65	35	Disability degree of at least 50% or occupationally disabled or incapable of employment pursuant to applicable law from 31.10.2000

Notes: [a]Official normal retirement age after increase of age limit in the respective reform.
[b]For the sake of completeness, the 2007 reform is included, but this reform is not applicable to our dataset.
[c]Retirement entry in 1984 or earlier (not applicable to our dataset): 15 years of contribution.
Source: own illustration based on Deutsche Rentenversicherung Bund (2015b), pp. 1–234; Social Security Administration (2014), pp. 114–121 as cited in Börsch-Supan et al. (2004), p. 292.

As the most attractive type of retirement, the **pension for women** has the lowest NRA, at 60 years for most observable cohorts. According to § 237a SGB VI, the NRA for women beginning

[40]The time frame of 10 contribution years is extended in cases of credited periods or child-raising periods (cf. SGB (n. d.), § 237).
[41]Cf. SGB (n. d.), § 237.

with the 1940 birth cohort is to be gradually increased to 65 years, with retirement starting in 2000. The pensionable age is 60 years and one month for those born in January 1940. It is increased with every subsequent month after January 1940. Beginning with the cohort born in December 1944, all women have to be 65 years old to receive a pension.[42] We have to examine eligibility using birth months and contribution type by checking whether women have 15 contribution years in total and more than 10 years in pensionable periods after their 40th birthday.[43] The ERA is constant at age 60.

The pension reform in 1991 further postponed the retirement age from 60 to 63 years for old-age **pensions for those with severe disability**. For severely disabled employees retiring before 1999, § 236a SGB VI allows retirement at age 60 in case of a 35-year contribution period.[44] To be able to allocate this old-age pension type, severe disability when entering the retirement has to be observable in the dataset. For retirement after 1998, severely disabled individuals born in January 1941 have to be 60 years and one month old, with one additional month for every subsequent month of birth. Therefore, individuals born in December 1943 have to be 63 years old to retire according to this law. The ERA is constant at 60 years.[45]

We assume that the financial incentives of the pension system depend on the best type of pension available for the individual, irrespective of whether this pension type is actually chosen or not. According to SEIBOLD (2017) and LORENZ ET AL. (2018), the best (or dominant) pension type offers the earliest possible retirement age with respect to ERA and NRA.

Adjustments to the earning point calculation

The calculation of earning points is strictly based on the formula of § 70 SGB VI, explained on page 10. In the following paragraphs, we describe the four adjustments to the earning points over the period 1951–2014 (see Table 3 for an overview of the variables) necessary to construct pension entitlements precisely.

First, until July 1978, the Federal Employment Agency paid contributions to the pension insurance for **periods of unemployment** (with or without benefits). For this period, the earning points have to be calculated with 80% of the limited overall assessment procedure (*Gesamtleistungsbewertung*).[46] This rule also applies from 1983 to 1991. During this time, receiving unemployment benefits (*Arbeitslosengeld*) or unemployment assistance *(Arbeitslosenhilfe*) is required as a condition for the Federal Employment Agency to pay the contributions.[47] From July 1978 until the end of 1982, the earning points in unemployment

[42]Cf. SGB (n. d.), § 237a.
[43]Cf. SGB (n. d.), § 237a.
[44]Cf. SGB (n. d.), § 236a.
[45]Cf. SGB (n. d.), § 236a.
[46]The paragraph § 71 describes the calculation of earning points for non-contribution and reduced contribution periods according to the overall assessment procedure. This procedure ensures a stronger dependency of the individual pension level on the amount and density of the contributions to the pension insurance (cf. Beye (2009), pp. 40–41; SGB (n. d.), § 71).
[47]Unemployment benefits are paid to all individuals who are registered as unemployed with the Federal Employment Agency and who have been subject to social security contributions for at least 12 months

periods are calculated as 100% of the former gross income and they are regarded as full contribution periods. From 1992 to the end of 2004, the earning points are calculated as 80% of the overall assessment procedure. This also applies for the period 2005–2010. However, there is an exception to consider: Periods with unemployment assistance are not considered with earning points. Periods with unemployment benefit II (*Arbeitslosengeld II*) are taken into account with the earning points resulting from the monthly contribution of €400 for 2005 and 2006 and the monthly contribution of €205 for 2007–2010. Dividing the monthly contribution by the applicable contribution rate of 19.5% in 2005 and 2006 or 19.9% in 2007–2010, we get an amount of money that can be used in the same way as labour income for the calculation of earning points. From 2011 to the end of 2014, the earning points are calculated as 80% of the overall assessment procedure again. Periods with unemployment assistance and periods with unemployment benefit II are not considered for the accumulation of earning points, however. The calculation of additional earning points accumulated during unemployment is complex and has changed several times over the years and with the reforms. We therefore do not take into account these additional rules but assess later whether we incur errors in the calculation of pension entitlements.

Second, individuals receive additional earning points during **parental leave and care-giving periods**.[48] According to the pension entry year and the birth years of their children, the additional earning points range from 0.0625 to 0.0833 per month. Pensions starting before 1 July 1998 allow 0.0625 additional earning points per month during parental leave or care-giving periods. Individuals paying pension contributions that generate more earning points than the value mentioned above do not receive the additional earning points. In parental leave or care-giving periods, retirement entry from 1 July 1998 to the end of the year 2001 allows additional earning points of 0.0833 per month. Having an income during parental leave and care-giving periods after 1991, individuals additionally get half of the earning points associated with their monthly income when retiring after 2001. The additional earning points may not exceed 0.0278 points per month. Moreover, the sum of original and additional earning points is limited to 0.0833 earning points per month, and individuals must fulfil 25 years of pensionable periods.[49] The same rules apply for individuals with two or more children; the only difference is that additional earning points of 0.0278 are granted to these people, regardless

in the last 24 months. The benefit period is 12 months for persons younger than 50 years old and up to 24 months for those over 50 (depending on the duration of the previous insured employment). Until 2004, unemployment assistance was paid for jobseekers immediately after the termination of unemployment benefits (for one year each, with the option of extending it several times). Unemployment benefit II (*Hartz IV*) replaced unemployment assistance in 2005. It is considered as a basic subsistence guarantee, and neither unemployment nor a previous unemployment benefit is necessary. It can also be used as a supplement to other income or benefits if that income is not sufficient to maintain livelihood (cf. Deutsche Rentenversicherung Bund (2015a), pp. 4–6; Geyer and Welteke (2017), pp. 5–6).

[48]The children dependent on care have to be younger than 18 years old.

[49]Cf. SGB (n. d.), § 70 (Artikel 1 des Gesetzes v. 18. Dezember 1989 und v. 01. Januar 2002); Beye (2009), pp. 37–41; Social Security Administration (2014), p. 115; Deutsche Rentenversicherung Bund (2016c), p. 308.

of employment income (with the same maximum of 0.0833 earning points from original and additional earning points per month).

Furthermore, individuals are allowed to pay pension contributions voluntarily and independently of prior income. However, the total of these additional earning points may not exceed the maximum values that can be earned per month (the pension contribution limit).

Third, after including all pension regulations in the calculation of the earning points per period, the sum of an individual's earning points must be adjusted in several cases. Until 31 December 1991, the sum of the earning points can be increased for months with a **low income,** generating less than 0.0625 earning points, on average. One condition is the collection of at least 35 years of pensionable periods and full pension contributions in the months with low income. These individuals also get half of the obtained earning points for their monthly income until the average of the earning points per month reaches 0.0625.[50]

Fourth, a modification made by the 1992 reform is the implementation of **pension deductions and pension supplements** when people retire earlier or later than the NRA.[51] German law codifies the beginning and the amount of the modification for every deviating month. Early retirement causes a 0.3% permanent deduction per month before the NRA, and there is a 0.5% supplement per month for working after the NRA.[52]

Variables necessary to construct pension entitlements from labour market history data

For an accurate calculation and to reproduce changes in employment status during the year or within a month, the data should be longitudinal and have daily accuracy. To rebuild pension entitlements, we require the following information. 'Daily wage' and the 'workplace in East or West Germany' is necessary to calculate earning points according to the formula mentioned on page 10. 'Gender' information is needed to examine the possibility of retirement for women, and 'date of birth' (at least the month and year of birth) is required to calculate statutory retirement dates. In addition, the validity of many pension rules and their protection of legitimate expectation *(Vertrauensschutzregel)* depend on the birth year.[53] In our dataset, the birth date is recorded on a daily basis; we can therefore determine the 'retirement date' to the day.[54] In LORENZ ET AL. (2018) we describe how to identify the pensionable periods in order to assign whether an individual is eligible for a certain 'type of old-age pension' and to determine the respective ERA and NRA. When we know the best pension type for an individual, we can adjust the age factor as described above. In case the retirement date is not available, LORENZ

[50]Cf. SGB (n. d.), § 262 (Artikel 1 des Gesetzes v. 18.Dezember 1989); Beye (2009), pp. 39–40; Social Security Administration (2014), pp. 114–115.
[51]Cf. SGB (n. d.), § 77.
[52]Cf. Börsch-Supan et al. (2004), pp. 291–295; Deutsche Rentenversicherung Bund (2014), p. 68.
[53]See Lorenz et al. (2018) for an overview of all protection of legitimate expectations per pension type.
[54]Individuals who were born on the 2nd–31th day of the month retire on the first day of the month after the birthday when they reach the NRA. If a person was born on the first day of the month, he or she may already retire on their birthday rather than a month later. If we have the date of birth on a monthly basis, we end up with only minimal deviations of a maximum of one month.

ET AL. (2018) show a method for how to approximate it on the basis of the last labour market activity period.

Table 3. List of variables needed in the dataset for the calculation of earning points[55]

- Daily wage
- Workplace in East or West Germany
- Gender
- Date of birth (at least month and year)
- Date of retirement
- Pension type
- Category of persons
  - Self-employed, civil servant, miner, seaman
  - Severe disability, occupational disability
- Current labour market state
  - School education
  - Employment subject to social security contributions
  - Full-time employment, marginal employment[a]
  - Receiving unemployment benefits/assistance (*Arbeitslosengeld ALG, ALG II, Arbeitslosenhilfe ALHI*)
  - Seeking work (registered or unregistered at the Federal Employment Agency)
  - Measures of job trainings and employment-related variables (*Maßnahmen zur Förderung der Berufsausbildung, Beschäftigung begleitende/schaffende Maßnahmen*)
  - Parental leave, care-giving periods
  - Voluntary insurance
  - Partial retirement
- Information on children

Notes: [a]'Marginal employment' (*geringfügige Beschäftigung*) involves employees with a yearly income below certain thresholds determined by the law (e.g., the threshold is €4800 in 2007; see Table A1 in Appendix A for all threshold values.

The 'category of persons' identifies individuals in BASiD who are severely disabled or occupationally disabled and therefore have the possibility of obtaining a pension because of severe disability.[56] In addition, we can identify people we exclude from the dataset: self-employed people, civil servants, miners, seamen and people who received a pension for surviving dependents or a reduced earnings capacity pension.[57] The 'current labour market state' and information on children are required for adjusting the calculated earning points per period and the sum of the earning points. Apart from BASiD, other administrative data do not include all of the listed variables. For example, in the SIAB, the 'current labour market state' (e.g., voluntary insurance) and 'information on children' are not available, and earning points related to this information are therefore missing. In Sections 5 and 6, we assess the deviations caused by this missing information.

[55]The following variables are used in SIAB: tentgelt, ao_bula, frau, gebjahr (we use gebdat via CADAL), erwstat, teilzeit, grund. The variable kind is not used because its quality is too low.

[56]If severe disability cannot be observed, the sum of the pension entitlements of these persons may be underestimated because of incorrect deductions. However, the severely disabled are only a small group of retirees (approximately 10% in our dataset and approximately 7% according to official reports of the German pension insurance). Therefore, not being able to identify these individuals will have minor effects, and no effect on the calculation of annual earning points. Severe disability only affects the retirement entry age and thus the possible adjustment of the sum of the earning points in cases of early or late retirement (cf. Deutsche Rentenversicherung Bund (2016a), p. 2).

[57]In most datasets, the pension for surviving dependents or the reduced earnings capacity pension is not observable. We assume that these pension types have no effect on the analysis.

3 Data Description and Sample Selection

This report is based mainly on the BASiD 5109 data provided by the Research Data Centre (FDZ) of the German Pension Insurance Agency (FDZ-RV) and by the Research Data Centre of the Federal Employment Agency in the Institute for Employment Research (FDZ BA/IAB). BASiD connects two longitudinal biographical data sources from different social security institutions. One of these data sources is the pension entitlement information from the VSKT from 1951 until the end of 2007.[58] In addition, selected individual data from the Federal Employment Agency are added (e.g., professional[59] and labour market status information, location and place of employment).[60] The individuals in the dataset represent a 1% representative sample of all persons aged 15–67 years who were insured by the German pension insurance on 31 December 2007.[61] BASiD covers employment histories from 1951 to 2009. This means that, at least for older birth cohorts, the complete employment history can be observed. BASiD contains all information necessary to study pension entitlements, including, among other information, the exact number of earning points per observation and the total sum of earning points per individual. In addition, all components of the sum of the pension entitlements, such as additional earning points during childcare, are available. The oldest people in the dataset are those who were employed subject to social security contributions for the first time in 1951 at the age of 15 years and who retired when aged approximately 63–65 years in 1999–2001. In co-operation with the FDZ BA/IAB and based on publicly available datasets, we use a unique individual BASiD dataset. More specifically, the co-operation in the framework of the Custom Shaped Administrative Data for the Analysis of Labour Market (CADAL) project allows us to include the day of birth.[62] This information enables us to identify the precise date of the statutory retirement entry and, consequently, the exact difference between the earliest statutory and the actual retirement date. Data from the VSKT and from the Federal Employment Agency are administrative and therefore do not have non-response or reporting error, which may confound survey-based analyses.[63] However, we do not have any information about activities during employment gaps or reduced employment hours that do not entitle someone to a pension. HOCHFELLNER ET AL. (2011) provide a detailed description of the dataset.

The BASiD includes data on 568,468 individuals.[64] For the calculation of the earning points and the comparison of the approximated entitlements with the true entitlements, we restrict the

[58]The data on pension entitlements in 2008 and 2009 are not the actual data for these years but rather an imputation of the information from 2007 8cf. Hochfellner et al. (2009), pp. 75–779. That is why we concentrate our analysis on the years 2007 and earlier.
[59]The education variable shows missing values because this variable is provided only for statistical purposes. We impute missing values using the method of Fitzenberger et al. (2005).
[60]Cf. Hochfellner et al. (2011), pp. 10–11.
[61]Cf. Hochfellner et al. (2011), p. 22.
[62]The CADAL project provided customised data for participants of the German Research Foundation (DFG) priority programme number 1764.
[63]Cf. Hochfellner et al. (2009), p. 75.
[64]Cf. Hochfellner et al. (2011), p. 14.

BASiD as follows. After deleting persons who did not retire before the end of 2007, self-employed people, civil servants, miners and persons with less than five contribution years (these individuals have no pension claim; see Table 2), a total of 10,883 individuals remains (4676 men, 6207 women).[65] For the evaluation of the calculation and the application in the SIAB, we create two different samples in BASiD, the BASiD Sample I that excludes observations with long gaps in the labour market history and the BASiD Sample II that mimics the SIAB environment and excludes observations before 1975.[66]

BASiD Sample I

The BASiD Sample I includes all persons who were employed for at least half of their labour market observation period and persons who did not insure themselves completely voluntarily. We observe 7269 individuals (3346 men, 3923 women) without extremely long spells outside the labour market from 1957 to 2007, corresponding to birth cohorts of 1940–1942.[67] This sample restriction includes retirements in the years 2000–2007. The average retirement age of 62.27 years of our restricted sample is very close to the average retirement age of 62.83 years for the 1940–1942 birth cohorts.[68]

Every individual in the sample has at least one observation on pension entitlements per year. Therefore, we can calculate the yearly individual deviation between our calculation and the BASiD earning point information, as well as the entire deviation per birth cohort, year or other subsamples.

BASiD Sample II

For the earning point calculation in SIAB, we have to apply additional restrictions. We replicate the SIAB data structure in the BASiD and use Sample II to evaluate the calculation method and to compare the results of the simulated SIAB data structure with the results of the actual values from BASiD. In addition to the restrictions of BASiD Sample I, we limit our observation period in BASiD Sample II to the years 1975–2007 in analogy to the left-censoring in SIAB 7514. After these reductions, we have 4204 individuals (2201 men, 2003 women) in the 1940–1942 birth cohorts. The BASiD Sample II again includes retirements in the years 2000–2007. Average retirement age is 62.22 years.[69]

In contrast to the BASiD, the SIAB has missing information on two variables that are key for our purposes: the retirement date and the pension type. We try to approximate this missing information using variables in the SIAB: More specifically, we use age at the end of the last

[65]Most individuals (493,419 persons) did not retire until 2007.
[66]See Table B1 in Appendix B for a detailed description of the characteristics for the three BASiD samples: BASiD, BASiD Sample I and BASiD Sample II.
[67]As we only have a small number of observations in the years 1951–1956 and missing pension information for the last two years (2008 and 2009), we concentrate on the years 1957–2007 in our descriptive analysis.
[68]Cf. Deutsche Rentenversicherung Bund (2016c), p. 142: See Table C in Appendix C for further sample characteristics.
[69]See column 5 in Table C1 and C3 in Appendix C for further sample characteristics.

employment activity as an approximation for the retirement age.[70] With further information from the existing individual employment history data, we can determine the eligibility for different pension types, assign the type of pension with the earliest option to retire, and accordingly select the statutory retirement date (ERA and NRA). Conditional on a number of adjustments to the dataset, the accuracy of the estimation of retirement age is very good.[71] Information on the number of children and their dates of birth is also unavailable in the SIAB. This information cannot be estimated for the majority of women because the observation period starts at age 32 at the earliest.[72] The pension for severely disabled people can also not be identified, as no information on the degree of disability is available.[73] Deviations caused by the lack of this information are discussed and evaluated in the following sections.

SIAB

The SIAB 7514 is a 2% representative sample of the population of the Integrated Employment Biographies of the IAB. The SIAB is an administrative dataset and covers the employment histories of 1,757,925 individuals. It contains both individual data and establishment information. A detailed description of the dataset can be found in ANTONI ET AL. (2016). With the SIAB, we also benefit from the collaboration of the CADAL project, which allows us to include the exact day of birth as a customized add-on.

When imposing the same sample restrictions as in our Sample II, we observe 73,958 individuals (45,736 men, 28,222 women) across substantially more birth cohorts (1936–1948)[74] and a longer observation period (1975–2014) than in the BASiD. This results in the retirement years of 1996–2014, with an average retirement age of 64.18 years.

[70]See also Seibold (2017), p. 81–82 for a detailed description of the approach for the 1932–1949 birth cohorts, as well as Brussig (2015), p.5.
[71]See Lorenz et al. (2018) for a detailed description of the sample restrictions necessary to obtain this result: at least one employment period subject to social security contributions before the 55th birthday; last labour market activity after the 59th birthday; first labour market activity before or at the age of 41; last labour market observation is not an information about the person's death; the status of compulsory contributions is known in the last ten years before leaving the labour market.
[72]Detailed methods to identify mothers by identifying family-related breaks in administrative datasets can be found in Müller and Strauch (2017) and Schönberg (2009). However, we cannot apply these methods because women in our samples are observable at the earliest from age 32.
[73]In this case, the variable 'Reason of cancellation/notification/termination (grund)' cannot help because the proportion of valid values (degree of completeness) is too low.
[74]We start with the 1936 birth cohort because we have to observe the years after the age of 40 (observation period starts in 1975) for women who are in principle eligible for the pension for women. Women must have more than 10 years in pensionable periods after their 40th birthday to be eligible for this pension type. See also Table 8 on page 35 and Table in Appendix E for further sample characteristics.

4 Calculation of Retirement Benefits

BASiD Sample I

In this section, we present the quality of the earning point recalculation in BASiD. Table 4 shows the average deviations per year and the average deviations of the sum of the earning points for different subgroups. Differences between the BASiD earning points and the calculated values can have either positive or negative signs. Positive deviations indicate an underestimation of the earning points, and negative deviations indicate an overestimation. The average annual deviation for the entire sample is 0.048 earning points; this means, on average, 5.99%, or an underestimation of €1.25 of pension entitlements per month.[75] The average deviation of true and approximated pension entitlement sums in Sample I is 1.83 earning points, which is a difference of 6.17% or €48.00 in total.

When we split the deviations by employee groups, we find the following pattern: The annual deviations for men are 0.032 earning points. They are lower in comparison with those for women (0.063 earning points). These values correspond to differences of €0.85 and €1.66 in monthly pension entitlements, respectively. The average deviations in the total amount of pension entitlements are even smaller for men (0.68) in comparison with 2.81 earning points for women, corresponding to differences of €17.81 and €73.76, respectively. Similarly, Western German workers have lower annual deviations than do Eastern German employees (0.048 and 0.059 earning points, respectively). A relatively high average annual deviation is observed for male employees in Eastern Germany (0.092 earning points) and for unemployed individuals (0.18 earning points). With 0.072 earning points, people with children show higher deviations than do those without children (0.030 earning points).[76] A period of marginal employment increases the deviations to an average of 0.075 earning points.

The largest deviation (0.18 earning points) can be found for unemployed people. At this point, it can already be assumed that gaps in the employment history lead to the largest deviations in the recalculation of earning points. One explanation might be income (e.g., marginal employment) or pension contributions (e.g., voluntary contributions) without notification to the social insurance; another reason might be undocumented information about individual or public pension contributions in the dataset. In such cases, the recalculation is impossible.

[75]This calculation is based on the current pension value as of July 2007 (€26.27). The annual deviations in the previous years are smaller than the listed values because the pension values are smaller. The percentage deviations given in the following sentences are average percentage deviations per subgroup (see Table 4).

[76]This can affect both men and women, but women are the majority.

Table 4. Descriptive statistics on average deviations of pension entitlements (in earning points), BASiD Sample I

	per year					per cohort		
Group	**mean**	**sd**	**min**	**max**	**N**	**mean**	**sd**	**N**
Men	0.0323	0.1954	0.0000	2.0233	135,316	0.6779	7.0281	3346
Women	0.0633	0.2257	0.0000	2.1149	134,038	2.8076	5.3651	3923
West	0.0477	0.2114	0.0000	2.1149	268,720	1.7659	6.1745	7152
Men	0.0322	0.1951	0.0000	2.0233	135,015	0.6354	6.9448	3288
Women	0.0634	0.2255	0.0000	2.1149	133,705	2.7279	5.2463	3864
East	0.0594	0.3009	0.0000	1.6323	634	5.5779	10.1706	117
Men	0.0921	0.3045	0.0000	1.6215	301	3.0865	10.5526	58
Women	0.0299	0.2950	0.0000	1.6323	333	8.0272	9.2269	59
Number of children								
0	0.0299	0.1904	0.0000	2.0233	155,980	0.7807	6.7041	3957
≥ 1	0.0723	0.2356	0.0000	2.1149	113,374	3.0777	5.4678	3312
Employed full-time	0.0154	0.1616	0.0000	2.1149	117,265	1.4928	6.2992	5211
Men	0.0168	0.1644	0.0000	2.0233	73,339	0.7462	6.8942	3030
Women	0.0130	0.1566	0.0000	2.1149	43,926	2.4957	5.2019	2181
Marginally employed	0.0745	0.2679	0.0000	2.1149	12,144	2.3684	7.6236	1510
Unem-ployed	0.1841	0.4075	0.0000	1.7788	14,633	2.6561	6.6211	2521
Large gap in employ-ment his-tory	0.0467	0.2556	0.0000	2.0934	56,750	3.0569	5.2843	2183
Total	0.0477	0.2116	0.0000	2.1149	269,354	1.8273	6.2763	7269

Notes: The table reports the average deviation in earning points between the true BASiD values and the approximated earning points calculated using our method. 'West' means having worked only in Western Germany, which corresponds to the clear majority of the people. 'East' means also having workplace observations in Eastern Germany. 'Large gap in employment history' means more than 5 unobservable years for men and more than 10 unobservable years for women. Average deviations per cohort are deviations in the sum of the individual earning points over the observable career. N is the number of observations per group for the left hand side of the table and the number of individuals for the right hand side of the table.
Data: BASiD 5109, Sample I

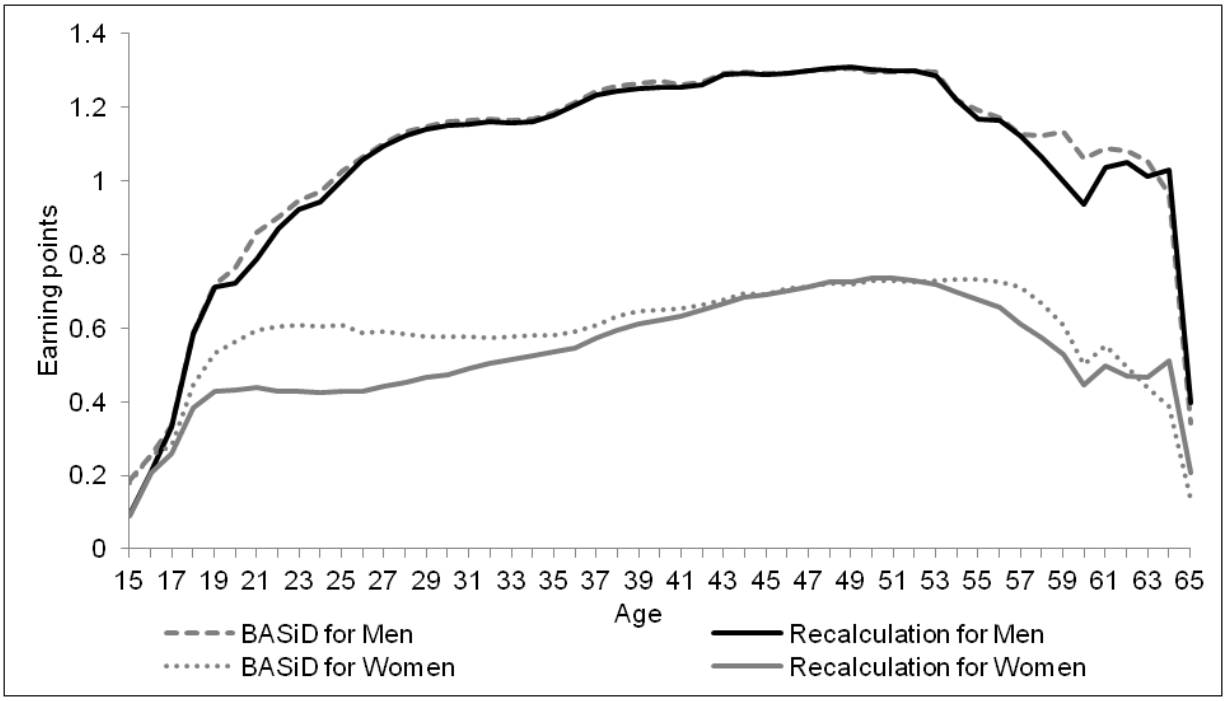
Looking at the average deviations in the sum of the earning points, the largest deviations emerge in East Germany (5.58 earning points), especially for women (8.027 earning points). One reason for this pattern is that Eastern German pension entitlements are observed only since 1990 and therefore there are huge observation gaps in the labour market history. As a result, Eastern Germans, especially women, have very large deviations in the sum of the earning points. People in full-time employment accordingly exhibit the smallest deviations, with an average of 0.015 earning points per year and 1.49 earning points in total; for men, the deviation in total earning points is only 0.77.

The maximum deviations correspond to the average maximum earning points that can be achieved per year.[77] This shows that there are individuals with a recalculated value of zero earning points, although they actually received the maximum earning points per year. We find this pattern when the income variable is recorded as €0 for someone who gets the maximum earning points per year because of, for example, voluntary contributions to the pension insurance. In our data preparation, we exclude individuals with only voluntary insurance (for example entrepreneurs or public employees (*Beamte*)). There are, however, still observations that cover earning points from voluntary insurance (for example, in addition to regular earning points from employment, special payments right before retirement to compensate for pension deductions or, particularly, during parental leave), although there is no claim that can be derived from the pension insurance data. These cases cannot be identified in the dataset because we only see if someone contributes voluntarily, but not the amount of voluntary payments. We find, however, that voluntary payments are responsible for the maximum earning point deviation.[78] Additional errors can result from periods in which the 'daily wage' variable is poorly recorded, mainly in times of non-regular employment subject to social security contributions. In addition, some information from employment spells is missing, resulting in a few observations with earning points, although there is no claim that can be derived from the pension insurance data.[79] The minimum of zero earning points is obtained if both BASiD and the recalculation show zero earning points for a particular year. This is only the case when income is €0 and there are no new earning points earned because, for example, people are engaged in parental care or care-giving periods, are unemployed or are out of the labour market for the entire year. We never find exactly zero deviations between the predicted and actual pension entitlements for people who earned pension entitlements or have positive earnings. The (usually very small) differences can therefore be attributed to rounding errors.

Figure 1 provides an overview of the development of both, the average earning points in the BASiD and the approximated values over the career span. An interesting feature is the inverted U-shape, which broadly resembles the age–earnings profile for workers aged from 15 to 65 years.[80] The inverted U-shaped development is particularly evident for men. Women show a flat development in pension entitlements between approximately ages 19 and 35. This is the main period of child birth.[81] Consequently, a large proportion of women are out of the labour market for a limited period of time or work less during their years with little children and accordingly receive fewer earning points.

[77]Cf. Table 1.
[78]Approximately 2.5%–3.0% of active insured persons are completely voluntarily insured. This percentage was higher in the 1960s–1980s (which we observe in the BASiD) than later. The decline is especially strong among women (cf. Deutsche Rentenversicherung Bund (2015c), pp. 15–17; SGB § 187a).
[79]Cf. Hochfellner et al. (2011), pp. 15–21.
[80]Cf. Figure in Appendix C.
[81]Cf. Figure C in Appendix C.

Figure 1. Development of the average earning points by age, BASiD Sample I



(Source: own illustration based on the BASiD 5109, Sample I)

Figure 1 makes it clear that there are few and seemingly non-systematic differences between the average predicted and actual pension entitlements. We also see that the approximated average earning points are consistently underestimated (also compare Table 4).

In Figure 2, we plot the age profile of the average deviation in earning points per year again using the BASiD Sample I. The deviations correspond to the gaps between the two lines from BASiD and the recalculation in Figure 1. Deviations occur particularly in individuals' first and last years on the labour market. Women show greater and longer-lasting deviations because of maternity leave and because of their low labour market attachment between the ages of 19 and 35. The deviations for women then decrease markedly, tending towards zero. Only in the last years before retirement entry deviations rise again, more so for men than for women.

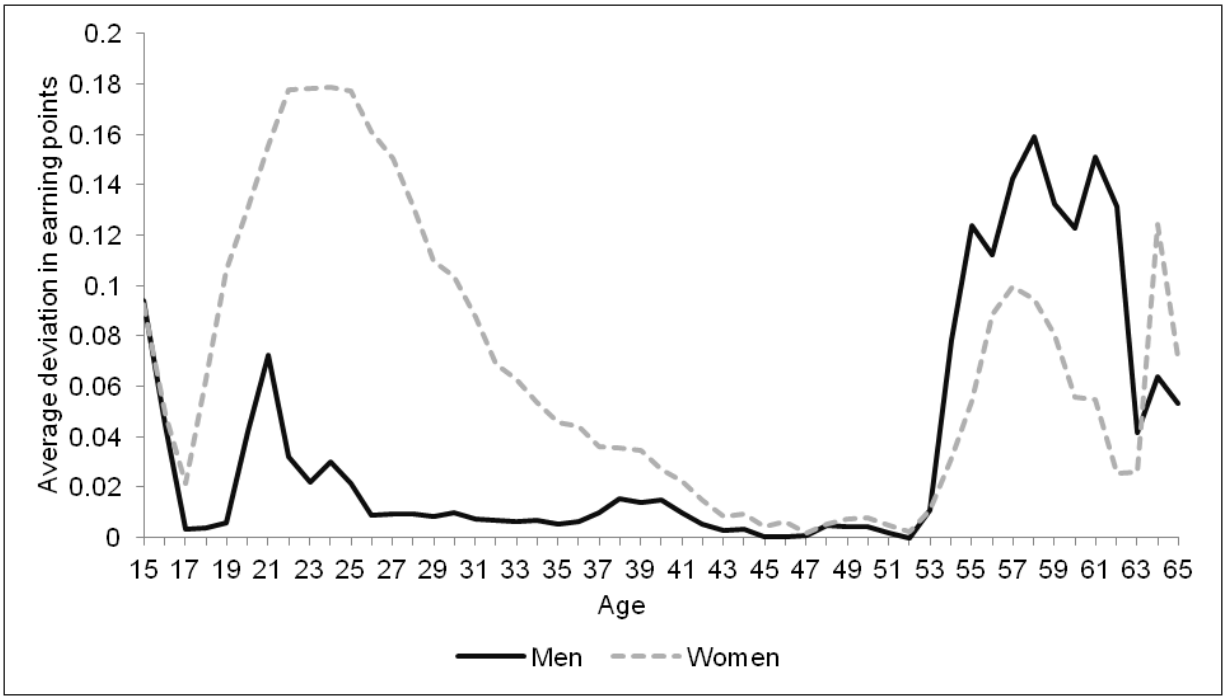
Unemployment as a bridging solution between regular employment and retirement is a common phenomenon. As shown by GROGGER AND WUNSCH (2012), ENGELS ET AL. (2017) and GEYER AND WELTEKE (2017), unemployed individuals usually use the full unemployment benefit and the full benefit period of 12–24 months before their pension starts. There are strong institutional incentives to bridge employment and retirement because, for example, job-search requirements for older people who are unemployed are very low.[82] We also find a high incidence of bridge unemployment before retirement in our sample, also compare Figure C3 in Appendix C.[83]

[82]See Footnote 47 for a description of the unemployment benefit system in Germany.
[83]Table C1 in Appendix C shows the high number of persons with a bridging solution (1619 of the 2840 persons, i.e., more than half of the people with an ERA). It is mainly older males interested in early

To assess the precision of our prediction error, in Figure D2 (Appendix D), we add the confidence intervals for the average deviations by age. It is important to note that the variance of deviations is very small. Deviations therefore do not differ enormously between individuals within each age group.

To summarise, our initial descriptive overview suggests that our calculations works well for a wide range of socio-demographic characteristics. To the extent that prediction errors occur, we find them to be concentrated in specific age groups and employment states.

Figure 2. Development of the average deviation in earning points by age, BASiD Sample I



(Source: own illustration based on the BASiD 5109, Sample I)

BASiD Sample I: Sensitivity analysis

In this section, we analyse how the quality of the earning point calculation improves when we increasingly restrict our sample to selected socio-economic groups. The results, summarised in Table 5, are again based on BASiD Sample I. Column 1 repeats the results from the section above (full Sample I), column 2 contains the results after eliminating Eastern German workers[84], column 3 additionally excludes persons with large labour market gaps (that is, more than 5 unobservable years for men and more than 10 unobservable years for women), column 4, in addition, drops women.

retirement who choose unemployment as bridge until the ERA. Reasons for this observation are given in Appendix B.

[84]We only include individuals who do not have a single employment observation in Eastern Germany. Therefore, the few people who worked in both Eastern and Western Germany are excluded as well.

Table 5. Deviation of the earning point calculation: Sensitivity analysis, BASiD Sample I and Sample II

	Sample I	Without East Germany	Without large labour market gaps	Without women	Sample II	Without East Germany	Without large labour market gaps	Without women
	1	2	3	4	5	6	7	8
Average EP per year	0.8661	0.8657	0.9253	1.1517	1.0502	1.0511	1.0668	1.3262
Average sum of EP	33.1371	33.1113	39.9420	51.5645	30.0861	30.1081	31.0035	38.9234
Average deviation of EP per year in %	5.9905	5.7636	5.6949	3.2038	1.1209	0.8354	1.0079	0.3447
Average deviation of sum of EP in %	6.1695	6.1648	6.3559	1.0007	8.1822	7.4983	6.9854	9.2416
Average deviation of EP per year in EP	0.0477	0.0457	0.0455	0.0342	0.0308	0.0287	0.0296	0.0324
Average deviation of sum of EP in EP	1.8273	1.9791	2.2846	0.7642	2.4612	2.2571	2.3587	3.4277
N	7269	7152	4998	2501	4204	4132	3951	2110

Notes: Column 1 represents the Sample I without further limitations. Columns 2–4 add step-by-step limitations, with each subsequent column containing the previous limitations. Column 5 represents the Sample II without further limitations. Columns 6–8 add step-by-step limitations, with each subsequent column containing the previous limitations. EP means earning points and N is the number of individuals per group.
Data: BASiD 5109, Samples I and II

The average deviation of earning points per year decreases with each restriction of the sample. In our preferred restriction (column 3: the sample with Western German workers without large employment gaps), we find an average deviation of 5.69% (equivalent to 0.046 earning points, or €1.20 per year). The deviation of the sum of the earning points is very similar to the deviation of the yearly change in earning points. With a 6.17% average deviation between the true and approximated pension entitlement sums in BASiD Sample I, we get a difference of 1.83 earning points, or €48.00 monthly pension earnings, in total.[85] In column 3, we find a deviation of 6.36% (corresponding to 2.28 earning points, or €60.02). In column 4, the deviations in the sum of the earning points are very small because women are excluded (1.0007%, or 0.76 earning points or €20.08 in total).

[85]This calculation is again based on the pension value in 2007.

In sum, these results show that the largest errors are associated with having an irregular working life. We next examine the sources of the calculation errors in more detail.

BASiD Sample II

In this section, we present the results of our earning point calculation for BASiD Sample II that mimics the SIAB data structure. Besides the reduction in the observation period to 1975-2007, we also remove variables that are not available in the SIAB: pension type and retirement entry date, number and dates of birth of children, identification of those entitled to a pension for the severely disabled. As a consequence, both the number of earning points and the deviations differ, compared with BASiD Sample I.

Table 6. Descriptive statistics on average deviations of pension entitlements (in earning points), BASiD Sample II

	per year					per cohort		
Group	**mean**	**sd**	**min**	**max**	**N**	**mean**	**sd**	**N**
Men	0.0328	0.2257	0.0000	2.1227	63,877	3.5134	5.3594	2201
Women	0.0284	0.1862	0.0000	2.1149	51,850	1.3051	4.8012	2003
West	0.0308	0.2078	0.0000	2.1227	115,212	2.4240	5.1568	4132
Men	0.0327	0.2243	0.0000	2.1227	63,607	3.5077	5.2631	2166
Women	0.0284	0.1855	0.0000	2.1149	51,605	1.2300	4.7607	1966
East	0.0466	0.3875	0.0000	1.6323	515	4.5997	7.7506	72
Men	0.0489	0.4570	0.0000	1.6215	270	3.8670	9.7239	35
Women	0.0441	0.2932	0.0000	1.6323	245	5.2933	5.3062	37
Employed full-time	0.0140	0.1678	0.0000	2.1227	83,899	2.7050	5.2046	3252
Men	0.0151	0.1744	0.0000	2.1227	56,080	3.5645	5.2230	2126
Women	0.0118	0.1538	0.0000	2.1149	27,819	1.0813	4.7659	1126
Marginally employed	0.0169	0.2439	0.0000	2.1227	2572	3.6131	3.8276	590
Unem-ployed	0.2702	0.4539	0.0000	1.7527	6133	4.6500	3.9235	1489
Large gap in employ-ment his-tory	0.0028	0.2742	0.0000	1.5314	3390	2.9853	5.2245	184
Total	0.0308	0.2081	0.0000	2.1227	120,303	2.4612	5.2184	4204

Notes: The table reports the average deviation in earning points between the true BASiD values and the approximated earning points calculated using our method for the transferred SIAB data structure in the BASiD (Sample II). 'West' means having worked only in Western Germany, which corresponds to the clear majority of the people. 'East' means also having workplace observations in Eastern Germany. 'Large gap in employment history' means more than 5 unobservable years for men and more than 10 unobservable years for women. Average deviations per cohort are deviations in the sum of the individual earning points. N is the number of observations per group for the left side of the table and the number of individuals for the right side of the table.
Data: BASiD 5109, Sample II

Table 6 shows that the annual average deviation for all persons is about 0.031, corresponding to 1.12%, or €0.81 monthly pension income in total, whereas the deviation of the sum of earning points equals 2.46 on average (8.18%, or €64.66). In this sample, men have higher deviation values than do women: 0.033, or 3.51 earning points vs. 0.028, or 1.31 earning points, respectively. As in BASiD Sample I, employees in Eastern Germany also show greater deviations in the sum of earning points than do those in Western Germany (4.56 vs. 2.42). The largest deviations are again for persons with unemployment spells (0.27 annually or 4.65 earning points in the sum). The smallest differences in the sum of the earning points are obtained for full-time employed women, at 1.081 earning points (corresponding to a deviation of 3.59%, or €28.41 in total).

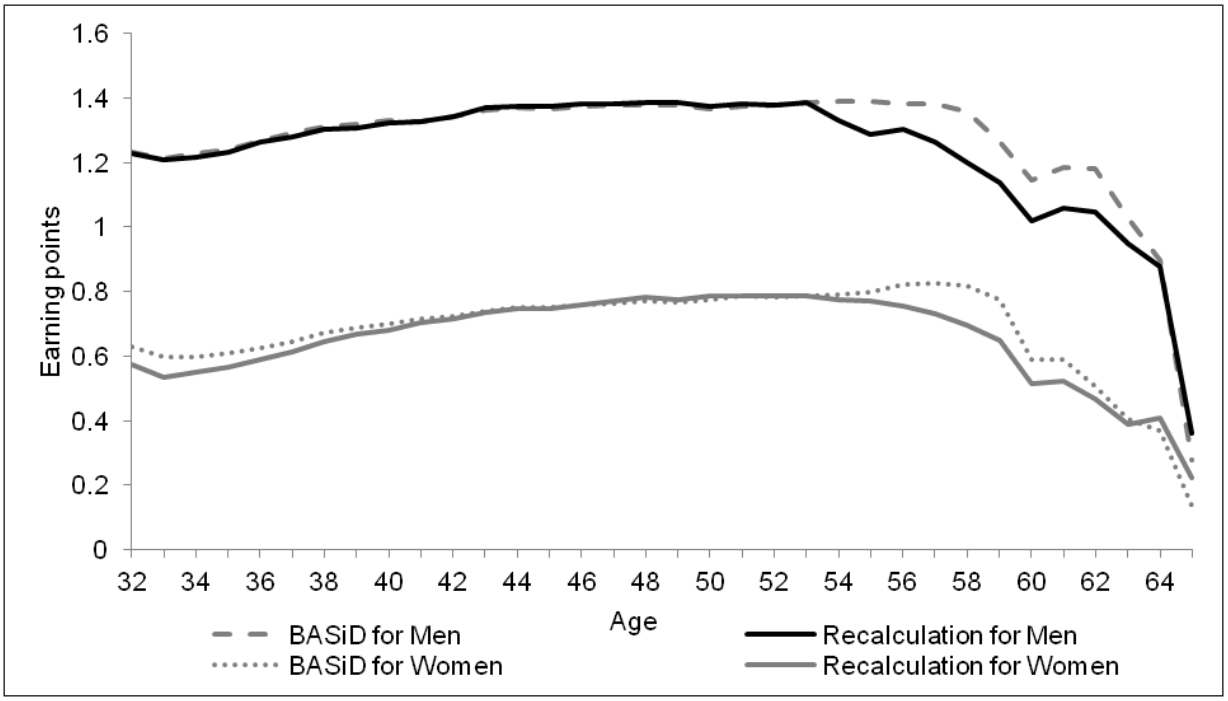
The maximum deviation per year corresponds, as in the BASiD Sample I, to the maximum number of earning points obtainable per year (the pension contribution limit). This maximum deviation again applies to cases for which income is missing. The minimum deviation is zero and applies only if there are no pension claims in a year.

Figure 3 shows the development of annual earning points by age for men and women. In Figure 4, we plot again the age profile of the average deviation in earning points per year in BASiD Sample II. The patterns are very similar to those seen in BASiD Sample I (Figure 1 and Figure 2). Note that the observations begin only at age 32.

In particular, we find that annual earning points are consistently underestimated. Again, we find that unemployment becomes more prevalent in the last years on the labour market (partly because it is used as bridge between employment and retirement)[86], and the calculation of retirement entitlements during unemployment is more problematic because pension rules are complex, statutory provision change frequently and we do not have all information needed to replicate the legal provision rules. The variance of the deviation of the annual earning points in the BASiD is very small and increases slightly with the deviation (see Figure in Appendix D). It should be noted again that gaps in the employment history can be seen as the largest source of calculation error.

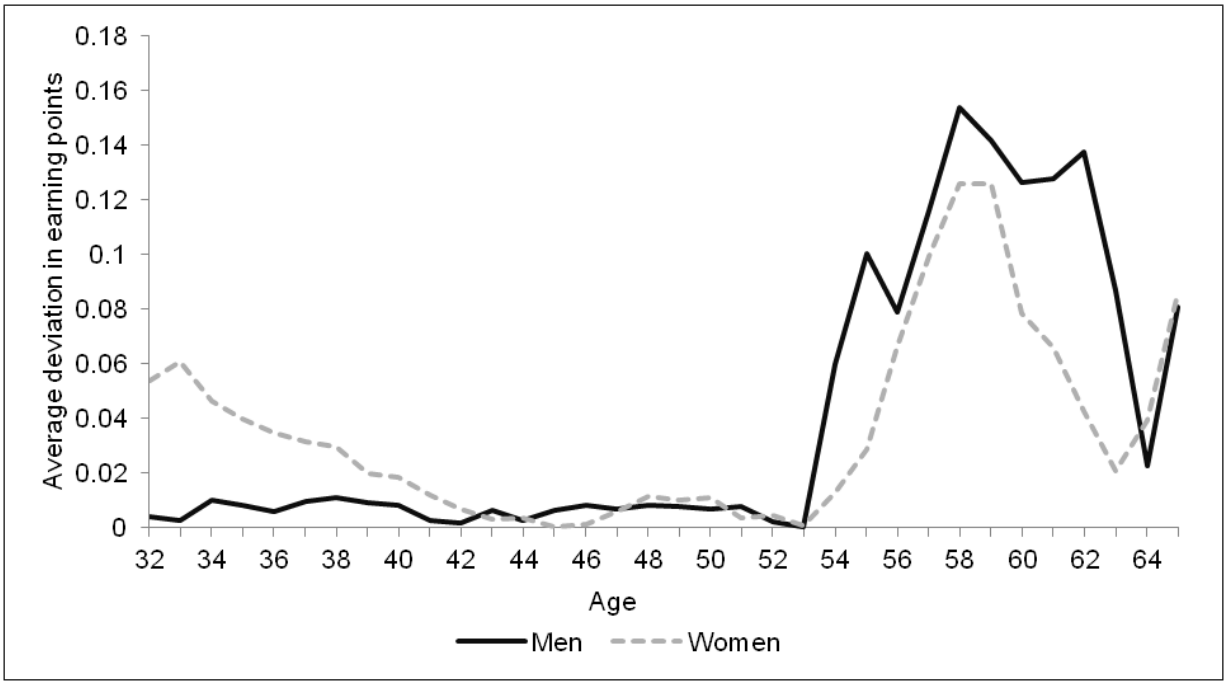
[86]Table C1 shows the high number of persons with a bridging solution including unemployment periods (1372 of the 2571 persons, i.e., approximately half of the people with an ERA). Figure C in Appendix C also shows the number of unemployment observations by age and FigureD1 in Appendix D shows the development of the average deviations by age for individuals without unemployment.

Figure 3. Development of the average earning points by age, BASiD Sample II



(Source: own illustration based on the BASiD 5109, Sample II)

Figure 4. Development of the average deviation in earning points by age, BASiD Sample II



(Source: own illustration based on the BASiD 5109, Sample II)

BASiD Sample II: Sensitivity analysis

In this section, we show again that the quality of the earning point calculation procedure improves when we restrict our sample to certain socio-economic groups, in line with the

restrictions from BASiD Sample I. The results based on BASiD Sample II can be found in Table 5. We see that increasing the limitations and excluding specific employee groups increases the annual average earning points and the sum of the average earning points. In addition, the deviations decrease from column to column. It seems, therefore, that people with low incomes have higher calculation errors and also leave the labour market earlier than the NRA.[87] In our preferred sample (only Western Germany, without large labour market gaps), the average annual deviation is 0.030 earning points, which corresponds to 1.008%, or €0.78. The deviation in the sum of the earning points is 2.36 earning points (which corresponds to 6.99%, or €61.96) in our preferred sample.

Although we are not able to observe all of the variables necessary to calculate the pension entitlements precisely in BASiD Sample II, the average deviations per year decrease compared with BASiD Sample I. Having less error obtained at the beginning of an individual's working life seems to compensate for the errors caused by the missing variables. The average level of earning points per year is higher in BASiD Sample II because the low-income years at the beginning of the career are not observed.[88] However, the amount of the sum of the earning points decreases in all columns in BASiD Sample II compared with BASiD Sample I and we also get a larger deviation in the sum of the earning points on average. Another driver of deviations in BASiD Sample II in comparison with BASiD Sample I is the lack of information on being entitled for the pension for the severely disabled. This missing information leads to an underestimation of entitlement points because we assume a deduction in the earning points during early retirement periods for these people although they can retire without deductions. Moreover, given missing information on children, we are unable to estimate earning point bonuses because of childcare, thus inducing a further downward bias in the calculated earning points. In Chapter 5, we examine the impact of the missing information in the SIAB more precisely.

We would like to emphasise again that the deviations in BASiD Sample II are mainly incurred by specific groups of people. In our preferred sample in column 7, the deviations are very small and, for the most part, they may be attributed to rounding errors. Errors could also arise from the missing income observations in the dataset, especially during time spent in active labour market policy measures or unemployment, working simultaneously in several jobs or making

[87]The number of people with an ERA is substantially higher than the number of people with an NRA or those who retire after the NRA (e.g., in column 8, 1636 of 2110 people [77.5%] have an ERA in comparison with 399 people with an NRA) (see Table C1 in Appendix C).

[88]The missing earning points from employment before age 32 can be approximated in accordance with the approach of Börsch-Supan et al. (2004), p. 300: They estimate the average earning points among all non-retired and full- or part-time workers in the sample who have a positive wage. They use a fixed-effects model for the earning points. This estimation is most efficient for earnings data because the fixed effects absorb all constant covariates (e.g., education, gender and marital status).This procedure would bring the amount of the sum of the pension entitlements close to the actual value in BASiD Sample I. We refrain from making such an approximation, as it is our purpose to show that our method works well for the existing observation years and for the available cohorts.

unobservable voluntary pension entitlement contributions.[89] The deviations in the sum of the earning points can largely be attributed to the missing observations of East Germans, the lack of identification of the severely disabled and the lack of information on children. Despite the unobservability of some variables in the SIAB in comparison with the BASiD, the calculations work well.

SIAB

In this section, we present the results of the earning point calculation for the SIAB dataset. Figure 5 shows the average earning points by age. The development is quite similar to the calculations using the BASiD Sample II.[90] However, the descriptive statistics of the SIAB sample cannot be directly compared with the BASiD Sample II results because the datasets are sampled differently. The sum of the earnings points is lower in the SIAB sample because the observations of individuals start with age 26 for the 1936–1948 birth cohorts instead of age 32 as in the BASiD Sample II. On average, individuals have 1.044 earning points annually and 28.44 earning points in total. Figure 5 shows the typical decline in average earning points in the last few years before retirement caused by the lower labour market attachment of this group. Women remain clearly below men in terms of the level of pension entitlements; women's entitlements only approach those of men towards the end of their career.

Figure 5. Development of average earning points by age, SIAB

(Source: own illustration based on the SIAB 7514)

[89]The information on the actual earning points comes from the VSKT. When preparing the BASiD dataset, some labour market information (typically regarding part-time or marginal employment) may be provided on parallel stints during a single period of time. The earning point information then contains the sum of the earning points generated during this period; however, not all labour market conditions are displayed. We therefore may not be able to account for part of the pension-relevant activities in our recalculation (cf. Stegmann (2006), p. 23; Himmelreicher and Stegmann (2008), p. 654).
[90]See Figure and Figure.

Table 7. Descriptive statistics on average pension entitlements (in earning points), SIAB

	per year					per cohort		
Group	**mean**	**sd**	**min**	**max**	**N**	**mean**	**sd**	**N**
Men	1.2187	0.4670	2.1368	0.0000	1,271,802	33.6986	13.9671	45,736
Women	0.7361	0.4363	2.1368	0.0000	688,854	19.9083	11.7403	28,222
West	1.0489	0.5149	2.1368	0.0000	1,836,957	30.9073	14.2233	62,707
Men	1.2232	0.4673	2.1368	0.0000	1,209,673	36.3434	12.4722	39,772
Women	0.7126	0.4284	2.1368	0.0000	627,284	21.4805	11.9869	22,935
East	0.9852	0.4968	2.1368	0.0000	164,664	14.6638	9.0836	11,251
Men	1.0600	0.5049	2.1368	0.0000	86,475	16.0607	10.0455	5964
Women	0.9025	0.4740	2.1368	0.0000	78,189	13.0880	7.5556	5287
Marginally employed	0.1128	0.1087	2.0767	0.0000	89,433	28.0788	14.9281	23,319
Unemployed	0.4174	0.3426	2.1368	0.0000	48,447	24.6910	13.0418	34,829
Large gap in employment history	0.5144	0.3890	2.1368	0	39,546	11.9875	7.2564	2105
Total	1.0436	0.5137	2.1368	0.0000	1,960,656	28.4363	14.7686	73,958

Notes: The table reports the average earning points per year and the average sum of the earning points per individual in the SIAB. ‘West’ means having worked only in Western Germany, which corresponds to the clear majority of the people. ‘East’ means also having workplace observations in Eastern Germany. ‘Large gap in employment history’ means more than 5 unobservable years for men and more than 10 unobservable years for women. Average earning points per cohort are the sum of the individual earning points. N is the number of observations per group for the left side of the table and the number of individuals for the right side of the table.
Data: SIAB 7514

Table 7 shows that women receive 0.74 earning points per year, on average, and men receive 1.22 earning points. The average values for employees in Eastern Germany are lower than for those for workers in Western Germany (0.99 yearly and 14.66 earning points in total vs. 1.05 and 30.91 earning points, respectively). Women in Western Germany, on average, receive fewer earning points per year than do men and women in Eastern Germany (0.71 earning points vs. 1.06 and 0.90 earning points, respectively). However, women in Eastern Germany have fewer total earning points than men in Eastern Germany. The main reason for this pattern is that people in Eastern Germany can only be observed since 1990. Male employees in Western Germany receive the most earning points, on average, both annually (1.22 earning points) and in total (36.34 earning points). The lowest values are obtained for marginally employed individuals (0.11 earning points annually) and for persons with large gaps in the sum of the earning points (11.99 earning points in total).

Table 8 shows an average age of 64.18 years at the last labour market observation in the SIAB sample. When the SIAB sample is sequentially limited using the BASiD sample restrictions, the average labour market exit age is stable around 64.00 years, but the average earning points per year increase. In our preferred sample in column 3, which excludes the errors incurred by the limited observation period in Eastern Germany and for persons with large

employment gaps, we obtain an average of 1.05 earning points per year and 31.01 earning points in total.

Table 8. Descriptive statistics for the SIAB

	SIAB	Without East Germany	Without large labour market gaps	Without women
	1	2	3	4
Average age at last labour market attachment	64.18	64.31	64.09	63.97
Average earning points per year	1.0436	1.0505	1.0514	1.2416
Average sum of earning points	28.4363	31.0397	31.0068	36.9347
N	73,958	61,562	58,320	35,803

Notes: Column 1 represents the SIAB sample without further limitations. Columns 2–4 add step-by-step limitations, with each subsequent column containing the previous limitations. N is the number of individuals per group.
Data: SIAB 7514

5 Possible Measurement Errors

A key finding of our analysis is that deviations in the recalculation of pension entitlements can be reduced by restricting the sample to persons without large employment gaps. However, two other sources of error might arise even in this restricted sample: deviations caused by rounding errors and minor accumulations in BASiD Sample I, and deviations caused by missing data on relevant variables in conventional administrative datasets in BASiD Sample II. To evaluate other potential sources of deviations, we regress the percentage deviation between the actual and calculated earning points on a full set of age and year fixed effects and dummies for gender, East Germany, the presence of one or more children and employment status (unemployment/marginal employment). We do not want to interpret the coefficients; rather, we are interested in their significance levels.

BASiD Sample I

In Table 9, BASiD Sample I shows statistically significant deviations for individuals with employment gaps. Employees with children and unemployment spells also have significantly higher deviations in the full BASiD Sample I. If one removes persons with large employment gaps, children, unemployment and also marginal employment periods still generate significantly higher deviations. The R^2 is very small (0.055 at most, depending on the limitations). This is an important result because it shows that earning point deviations are mainly unsystematic and caused by rounding errors.

Table 9. Estimation of the percentage deviation from the BASiD earning points per year, BASiD Sample I

	Sample I		Without East Germany		Without large labour market gaps		Without women	
	1		2		3		4	
Variable	**Coefficient**	**P > |t|**	**Coefficient**	**P > |t|**	**Coefficient**	**P > |t|**	**Coefficient**	**P > |t|**
Female	0.0043246	0.356	0.0032295	0.494	0.0034623	0.398		
East	0.0362722	0.187						
Children	0.0842130	0.000	0.0834106	0.000	0.0722471	0.000	0.0133302	0.301
Unemployment	0.0465982	0.000	0.0440364	0.000	0.1464284	0.000	0.1061393	0.000
Marginal employment	−0.009145	0.210	−0.018636	0.012	0.0344441	0.000	0.0567861	0.000
Year dummies	Yes		Yes		Yes		Yes	
Age dummies	Yes		Yes		Yes		Yes	
Number of observations	250,809		246,562		199,726		110,242	
R^2	0.0545		0.0545		0.0391		0.0422	

Notes: All variables are dummy variables. Negative signs mean overestimation of earning points and positive signs mean underestimation of earning points. Column 1 represents the BASiD Sample I without further limitations. Columns 2–4 add step-by-step limitations, with each subsequent column containing the previous limitations.
Data: BASiD 5109, Sample I

BASiD Sample II

As we cannot calculate deviations in the SIAB directly, we further check for possible variables that have statistically significant deviations in the SIAB data structure using BASiD Sample II. In Table 10, BASiD Sample II shows significant deviations in the calculation of annual earning points for those with unemployment or marginal employment. Unemployment and marginal employment remain significant under all tested restrictions (even with shorter durations). In contrast to Sample I, children do not cause significant deviations.[91] Even with further robustness checks, including dummies for the numbers of children and maternity protection spells, children are not associated with significant deviation differences.[92] The significant deviation in unemployment and marginal employment seems to be driven by unobservable income information (either because the information is missing or because there are parallel employment periods with only one pension-relevant activity included in the dataset), voluntary pension contributions and small errors in the implementation of public compensation payments

[91]Children are not directly observable in the SIAB data structure. In the BASiD, however, we can see who has a child and, therefore, attribute the deviations to these persons.
[92]See Table in Appendix D.

during these periods as a result of the complex legal rules.[93] Again, it should be emphasised that the R^2 is very small (smaller than 1% in all columns for BASiD Sample II). In addition, the significant measurement errors caused by gaps in the employment history create only small earning point deviations. Therefore, our calculation seems to work well for most employment groups.

Table 10. Estimation of the percentage deviation from the BASiD earning points per year, BASiD Sample II

	Sample II		Without East Germany		Without large labour market gaps		Without women	
	5		6		7		8	
Variable	**Coefficient**	**P > |t|**	**Coefficient**	**P > |t|**	**Coefficient**	**P > |t|**	**Coefficient**	**P > |t|**
Female	0.0096492	0.663	0.0095807	0.670	0.0076574	0.736		
East	0.0392375	0.667						
Children	0.0127044	0.575	0.0113356	0.623	0.0133188	0.568	0.0097896	0.904
Unemployment	0.1620273	0.000	0.1621582	0.000	0.1910741	0.000	0.1738817	0.001
Marginal employment	0.4463713	0.000	−0.529747	0.000	−0.610084	0.000	−0.422641	0.000
Year dummies	Yes		Yes		Yes		Yes	
Age dummies	Yes		Yes		Yes		Yes	
Number of observations	111,397		109,472		106,641		60,609	
R^2	0.0024		0.0026		0.0026		0.0078	

Notes: All variables are dummy variables. Negative signs mean overestimation of earning points and positive signs mean underestimation of earning points. Column 5 represents Sample II without further limitations. Columns 6–8 add step-by-step limitations, with each subsequent column containing the previous limitations.
Data: BASiD 5109, Sample II

In the analysis of the recalculation errors for the sum of the earning points in Table 11, we see statistically significant deviations between men and women. These differences can be traced back to women working in Eastern Germany (see Table 6). For the sum of the earning points, we obtain significant deviations for the dummy for children. We cannot take into account additional earning points as a consequence of childcare because the information on the number of children and their birthdates is missing. However, robustness checks show no significant deviations if we add dummies for numbers of children.[94] Marginal employment and the missing identification of severe disability also lead to slightly significant deviations in the

[93]There are a number of small socio-economic groups (e.g., craftsmen) that enjoy a variety of additional special pension rules. For these groups we might not be able to consider all of the legal rules adequately.
[94]See Table in Appendix D.

sum of pension entitlements, whereas unemployment does not seem to cause any significant deviation. The R^2 is smaller than 1% for all specifications, suggesting that, for the most part, our deviations are due to small and unsystematic rounding errors.

Table 11. Estimation of the percentage deviation from the BASiD sum of the earning points, BASiD Sample II

	Sample II		Without East Germany		Without large labour market gaps		Without women	
	5		6		7		8	
Variable	**Coefficient**	**P > |t|**	**Coefficient**	**P > |t|**	**Coefficient**	**P > |t|**	**Coefficient**	**P > |t|**
Female	−2.377011	0.000	−2.428145	0.000	−2.474826	0.000		
East	−0.600866	0.113						
Children	2.684470	0.080	2.711471	0.090	2.753560	0.000	0.2946978	0.637
Unemployment	−0.103959	0.395	−0.043417	0.726	−0.712962	0.578	−1.283842	0.000
Marginal employment	3.050566	0.000	2.776667	0.000	2.674359	0.000	3.707538	0.000
Severe disability	0.4143787	0.009	0.4190884	0.009	0.4110128	0.013	0.2898186	0.185
Cohort dummies	Yes		Yes		Yes		Yes	
Number of observations	242,908		237,156		229,174		119,596	
R^2	0.0017		0.0015		0.0015		0.0023	

Notes: All variables are dummy variables. Negative signs mean overestimation of earning points and positive signs mean underestimation of earning points. Column 5 represents the Sample II without further limitations. Columns 6–8 add step-by-step limitations, with each subsequent column containing the previous limitations.
Data: BASiD 5109, Sample II

6 Discussion

The advantage of the BASiD dataset is that most of the variables necessary to calculate public pension entitlements are available, including the variables for earning points per year and the sum of pension entitlements. Therefore, the BASiD is a good basis for checking the functionality of a calculation of individual earning points. As we want to have the complete employment history and the final sum of earning points that constitute the pension entitlement of a retiree, we restrict ourselves to a particular subsample in the BASiD: persons who retired in 2007 at the latest. As a result, we end up with employees from three birth cohorts (1940–1942) who retire in 2000–2007 at 60–67 years of age. This sample restriction results in a relatively homogeneous age composition in each calendar year, but also reduces our sample to a relatively small number of individuals. A small sample size, coupled with limited information on many other variables that are potentially interesting in the study of old-age employment,

severely restricts the usability of BASiD for this particular research agenda. In this report, we propose a method to calculate two key variables of interest for the analysis of the labour market for older employees — the annual and total numbers of earning points obtained — in a large and widely used conventional administrative labour market history dataset: the SIAB. We combine information from individual employment histories with pension contribution limits, early and regular retirement rules, and the average gross earnings of the entire population per year to calculate the earning points obtained at the individual level.

It is clear that any administrative dataset that does not contain all of the necessary pension information implies errors in the pension entitlement calculation. We assess the magnitude of these errors and demonstrate the quality of our method under different sample restrictions. We show that there are only small errors in the recalculation for key socio-economic groups—for example, individuals with full-time work. We also show that the deviations decrease when we exclude groups with irregular working lives, adding restrictions step-by-step in a sensitivity analysis. The deviations we find using the BASiD Sample I that only includes employees without very large gaps in their career should be negligible for most research questions concerning the financial aspects of the pension decision. Depending on the research question, researchers may apply fewer or more sample restrictions (see Chapter 4).

The missing variables in the SIAB lead to a systematic underestimation of pension entitlements. The extent to which measurement error in the dataset significantly biases the estimates depends on which earning point variable (yearly or sum) is used. Tables 10 and 11 reveal several interesting patterns concerning the significance of the deviation. First, missing information is not necessarily bad. Missing birthdates or information on children does not lead to statistically significant deviations. Second, the largest and most systematically significant deviations are caused by gaps in the employment history, paired with low or no income. Thus, unemployment and marginal employment mainly cause deviations (see Chapter 5 and FigureD1 in Appendix D).

Based on these findings, we conclude that the systematic underestimation is, to a large extent, attributable to persons who earn money and pay contributions but whose income is not observable in the dataset or our information on income is limited because of periods with multiple jobs. In addition, persons who voluntarily contribute to the pension insurance generate calculation errors. Unfortunately, we have only indirect evidence for this hypothesis, based on the observation that entitlements are underestimated during periods of unemployment, parental leave and marginal employment. There unfortunately seems to be no information on the amount of these voluntary pension contributions. Pension deductions calculated for severely disabled persons, who in reality may not be devalued because of a lower statutory ERA, can lead to an underestimation of entitlements. In addition, the calculation of public supplements to pension entitlements during periods of unemployment, parental leave and marginal employment is complex (see our description in Chapter 2). We therefore cannot be certain about whether we were able to consider all of the legal rules adequately, especially for special socio-economic groups (e.g., craftsmen) with a variety of different specific regulations.

In sum, calculation errors seem to be, to a large extent, the consequence of rounding errors or minor accumulations of pension entitlements or deductions to these entitlements. This interpretation of the results is demonstrated by including important controls for employee groups in a multivariate regression. The explanatory power of observable characteristics such as gender, unemployment incidence, region and birth cohort for calculation errors is very small (R^2 much smaller than 1% under all restrictions and robustness checks in the BASiD Sample II regressions), and errors therefore seem to be largely unsystematic.

7 Conclusions

In this report, we show that it is possible to calculate individual pension entitlements using conventional administrative labour market data without information on earning points (*Entgeltpunkte*). Pension entitlements can be seen as an important driver of the retirement decision and of labour market attachment at older ages. The calculation of pension entitlements using labour market data therefore allows the exploration of a broad area of new research questions. We implement our method to calculate pension entitlements in the SIAB, a larger administrative dataset with information from a sample covering most employees in Germany.

The report shows which information is necessary for the approximation of pension entitlements, discusses how pension rules in Germany (and changes in these) can be taken into account and presents a step-by-step approach to the calculation of earning points.[95] We find that the small systematic deviations can be attributed to missing information for employees during periods of parental leave, unemployment, and marginal employment. For all other employment periods, the errors are unsystematic, small and mainly a result of rounding errors.

Overall, we believe our approach delivers sufficiently precise results for further research in the field of old-age employment and individual retirement decisions drawing on widely used labour market history data. Although we work with the SIAB, the following comparable administrative datasets could also be appropriate for implementing the procedure described here to calculate earning points: linked employer–employee data of the IAB (LIAB), the Linked Personnel Panel (LPP) and LPP survey data linked to administrative data of the IAB (LPP-ADIAB), the WeLL Panel—Employee Survey for the 'Further Training as a Part of Lifelong Learning' project and WeLL survey data linked to administrative data of the IAB (WeLL-ADIAB), the panel study 'Labour Market and Social Security' (PASS) and PASS survey data linked to administrative data of the IAB (PASS-ADIAB), Working and Learning in a Changing World (ALWA), ALWA

[95] In a companion report, we describe a strategy for the identification of the pensionable periods and determine whether an individual is eligible to obtain one of the old-age pension types. Eligibility for a pension types allows us to identify statutory retirement dates (ERA and NRA) and individual paths of older employees out of the labour market. Moreover, we describe sample restrictions necessary to get minimal errors in determining ERA, NRA, and pensionable periods. We also use the SIAB7514 for this report (cf. Lorenz et al. (2018)).

Literacy and Numeracy Data (ALWA-LiNu), and ALWA survey data linked to administrative data of the IAB (ALWA-ADIAB).

References

AltersTZG (n. y.): Altersteilzeitgesetz vom 23. Juli 1996 (BGBl. I S. 1078), das zuletzt durch Artikel 151 des Gesetzes vom 29. März 2017 (BGBl. I S. 626) geändert worden ist.

Antoni, M./Ganzer, A./vom Berge, P. (2016): Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975 - 2014, FDZ Datenreport 04/2016, Nürnberg.

Arent, S./Nagl, W. (2010): A Fragile Pillar: Statutory Pensions and the Risk of Old-Age Poverty in Germany. In: FinanzArchiv, 66 (4), pp. 419-441.

Berkel, B./Börsch-Supan, A. (2004): Pension Reform in Germany: The Impact on Retirement Decisions. In: FinanzArchiv: Public Finance Analysis, 60 (3), pp. 393-421.

Beye, C.-J. (2009): Gesetzliche Rentenversicherung als Teil der Altersvorsorge, Heidelberg.

Bieber, U./Stegmann, M. (2000): Sozialversicherungspflichtige Teilzeitbeschäftigung in den Erwerbsbiografien der zukünftigen Rentnerinnen. Eine Kohortenbetrachtung auf Basis der Untersuchung Altersvorsorge in Deutschland 1996 (AVID '96). In: DRV-Schriften, 55 (6), pp. 364-383.

BMAS (2016): Ergänzender Bericht der Bundesregierung zum Rentenversicherungsbericht 2016 gemäß § 154 Abs.2 SGB VI (Altersicherungsreport 2016), Bundesministerium für Arbeit und Soziales, Berlin.

Boockmann, B./Steiner, V. (2006): Cohort effects and the returns to education in West Germany. In: Applied Economics, 38 (10), pp. 1135-1152.

Börsch-Supan, A. (1992): Population Aging, Social Security Design, and Early Retirement. In: Journal of Institutional and Theoretical Economics, 148, pp. 533-557.

Börsch-Supan, A./Bucher-Koenen, T./Coppola, M./Lamla, B. (2015): Savings in Times of Demographic Change: Lessons from the German Experience. In: Journal of Economic Surveys, 29 (4), pp. 807-829.

Börsch-Supan, A./Ferrari, I. (2017): Old-age Labor Force Participation in Germany: What Explains the Trend Reversal among NBER Working PaperOlder Men? And What the Steady Increase among Women?, NBER Working Paper Series, 24044, Cambridge.

Börsch-Supan, A./Schnabel, R. (1999): Social Security and Retirement in Germany. In: Gruber, J./Wise, D.A. (Eds.): Social Security and Retirement around the World, Chicago, pp. 135-180.

Börsch-Supan, A./Schnabel, R./Kohn, S./Mastrobuoni, G. (2004): Micro-Modeling of Retirement Decisions in Germany. In: Gruber, J./Wise, D.A. (Eds.): Social Security Programs and Retirement around the World: Micro-Estimation, Chicago, pp. 285-343.

Börsch-Supan, A./Wilke, C.B. (2004): The German Public Pension System: How it Was, How it Will Be, NBER Working Paper Series, 10525, Cambridge.

Brussig, M. (2015): Alter beim Austritt aus sozialversicherungspflichtiger Beschäftigung ist gestiegen, Altersübergangs-Report, 2015-01, Institut Arbeit und Qualifikation (IAQ), Universität Duisburg-Essen.

Burkert, C./Hochfellner, D. (2017): Employment Trajectories Beyond Retirement. In: Journal of Aging & Social Policy, 29 (2), pp. 143-167.

Card, D./Kluve, J./Weber, A. (2010): Active Labor Market Policy Evaluations: A Meta-Analysis. In: Economic Journal, 120 (548), pp. 452-477.

De la Croix, D./Pierrard, O./Sneessens, H.R. (2013): Aging and pensions in general equilibrium: Labor market imperfections matter. In: Journal of Economic Dynamics and Control, 37 (1), pp. 104-124.

Deutsche Rentenversicherung Bund (2014): 125 Jahre gesetzliche Rentenversicherung, Deutsche Rentenversicherung Bund, Berlin.

Deutsche Rentenversicherung Bund (2015a): Arbeitslos - was Sie beachten sollten, Deutsche Rentenversicherung Bund, 503, Berlin.

Deutsche Rentenversicherung Bund (2015b): Gesetzesänderungen Rentenversicherung von 1989 bis 2014, Deutsche Rentenversicherung Bund, Berlin.

Deutsche Rentenversicherung Bund (2015c): Rentenversicherung in Zeitreihen, DRV-Schriften, 22, Berlin.

Deutsche Rentenversicherung Bund (2016a): Altersrenten im Zeitablauf, Deutsche Rentenversicherung Bund, Berlin.
Deutsche Rentenversicherung Bund (2016b): http://www.deutsche-rentenversicherung.de/Allgemein/de/Inhalt/6_Wir_ueber_uns/downloads_drv/orgplan_sv_pdf.html (12 August 2017).
Deutsche Rentenversicherung Bund (2016c): Rentenversicherung in Zeitreihen, DRV-Schriften, 22, Berlin.
Deutsche Rentenversicherung Bund (2017): Rentenversicherung in Zahlen 2017, Deutsche Rentenversicherung Bund, Berlin.
Engels, B./Geyer, J./Haan, P. (2017): Pension Incentives and Early Retirement. In: Labour Economics, 47, pp. 216-231.
Fisher, W.H./Keuschnigg, C. (2010): Pension reform and labor market incentive. In: Journal of Population Economics, 23 (2), pp. 769-803.
Fitzenberger, B./Osikominu, A./Völter, R. (2005): Imputation rules to improve the education variable in the IAB employment subsample, FDZ-Methodenreport, 3, Nürnberg.
Fitzenberger, B./Schnabel, R./Wunderlich, G. (2001): The gender gap in labor market participation and employment: a cohort analysis for West Germany, ZEW Discussion Paper, 47 (1), Mannheim.
Geyer, J./Steiner, V. (2014): Future public pensions and changing employment patterns across birth cohorts. In: Journal of Pension Economics and Finance, 13 (2), pp. 172-209.
Geyer, J./Welteke, C. (2017): Closing Routes to Retirement: How Do People Respond?, DIW Discussion Papers, 1653, Berlin.
Grogger, J./Wunsch, C. (2012): Unemployment Insurance and Departures from Employment: Evidence from a German Reform, Working Paper, Department of Economics, VU University Amsterdam.
Gruber, J./Wise, D.A. (1999): Social Security and Retirement around the World, Chicago.
Gürtzgen, N./Nolte, A. (2017): Imputation rules for the implementation of the pre-unification education variable in the BASiD Data Set. In: Journal for Labour Market Research, 50 (1), pp. 45-65.
Hanappi, T.P. (2012): Retirement Behaviour in Austria: Incentive Effects on Old-Age Labor Supply, NRN Working Paper, 1213, Linz.
Hanel, B. (2010): Financial Incentives to Postpone Retirement and Further Effects on Employment – Evidence from a Natural Experiment. In: Labour Economics, 17 (3), pp. 474-486.
Himmelreicher, R./Frommert, D. (2006): Gibt es Hinweise auf zunehmende Ungleichheit der Alterseinkünfte und zunehmende Altersarmut? In: Vierteljahrshefte zur Wirtschaftsforschung, 75 (1), pp. 108–130.
Himmelreicher, R./Stegmann, M. (2008): New Possibilities for Socio-Economic Research through Longitudinal Data from the Research Data Centre of the German Federal Pension Insurance (FDZ-RV). In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 128 (4), pp. 647-660.
Hochfellner, D. (2014): Older Workers: Employment beyond Retirement and Old Age Poverty, Dissertation, Bamberg University.
Hochfellner, D./Müller, D./Wurdack, A. (2011): BASiD - Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland, FDZ Datenreport 09/2011, Nürnberg.
Hochfellner, D./Voigt, A./Budzak, U./Steppich, B. (2009): Das Projekt BASiD - Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland: Projektinhalte, aktueller Stand der Arbeiten und Analysemöglichkeiten. In: DRV-Schriften, 55, pp. 74-86.
Honekamp, I./Schwarze, J. (2010): Pension reforms in Germany: Have they changed savings behaviour? In: Pensions: An International Journal, 15 (3), pp. 214-225.
Kluth, S./Gasche, M. (2013): Ersatzraten in der Gesetzlichen Rentenversicherung, MEA Discussion Papers, 11, München.
Kortmann, K./Schatz, C. (1999): Altersvorsorge in Deutschland 1996 - (AVID'96) Zusammenfassung wichtiger Ergebnisse der Untersuchung "Strukturen und Trends

der Altersvorssorge von 40- bis 60-jährigen Rentenversicherten und ihrer Ehepartner". In: DRV-Schriften, 10-11, pp. 573-597.
Larsen, M./Pedersen, P.J. (2017): Labour force activity after 65: what explain recent trends in Denmark, Germany and Sweden? In: Journal for Labour Market Research, 50 (1), pp. 15-27.
Lewicki, M.P. (2014): Sustainability of the German Pension Scheme, Dissertation, Karlsruher Institut für Technologie (KIT).
Lorenz, S./Pfister, M./Zwick, T. (2018): Identification of the statutory retirement dates in the Sample of Integrated Labour Market Biographies (SIAB), mimeo.
Manoli, D.S./Weber, A. (2016): Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions. In: American Economic Journal: Economic Policy, 8 (4), pp. 160-182.
Müller, D./Strauch, K. (2017): Identifying mothers in administrative data, FDZ-Methodenreport 13/2017, Nürnberg.
Neugschwender, J. (2014): Pension Income Inequality: a Cohort Study in Six European Countries, LIS Working Paper Series, 618, Luxemburg.
OECD (2017): Pensions at a Glance 2017: OECD and G20 indicators, OECD Publishing, Paris.
Potrafke, N. (2011): Unemployment, human capital depreciation and pension benefits: an empirical evaluation of German data. In: Journal of Pension Economics and Finance, 11 (02), pp. 223-241.
Rasner, A. (2014): Gender Pension Gap in Eastern and Western Germany. In: DIW Economic Bulletin, 4 (11), pp. 42-50.
Riphahn, R.T./Schmidt, P. (1995): Determinanten des Rentenzugangs: Lockt der Ruhestand oder drängt der Arbeitsmarkt?, MEA Discussion Papers, 10, München.
Rüb, F.W./Lamping, W. (2010): German Pension Policies: The Transformation of a Defined Benefit System into … What? In: German Policy Studies, 6 (1), pp. 143-183.
Schatz, C./Merz, J./Kortmann, K. (2002): Künftige Alterseinkommen. Eine Mikrosimulationsstudie zur Entwicklung der Renten und Altersvorsorge in Deutschland (AVID '96). In: Schmollers Jahrbuch, 122, pp. 227-260.
Schönberg, U. (2009): Does the IAB employment sample reliably identify maternity leave taking? A data report. In: Zeitschrift für Arbeitsmarkt Forschung, 42 (1), pp. 49-70.
Seibold, A. (2017): Statutory Ages and Retirement: Evidence from Germany, Working paper, School of Economics.
SGB (n. d.): Das Sechste Buch Sozialgesetzbuch – Gesetzliche Rentenversicherung – in der Fassung der Bekanntmachung vom 19. Februar 2002 (BGBl. I S. 754, 1404, 3384), das zuletzt durch Artikel 7 des Gesetzes vom 21. Dezember 2015 (BGBl. I S. 2517) geändert worden ist.
Social Security Administration (2014): Social Security Programs Throughout the World: Europe, 2014, SSA Publication, 11801 (13), Washington.
Stegmann, M. (2006): Aufbereitung der Sondererhebung „Versicherungskontenstichprobe (VSKT)“ als Scientific Use File für das FDZ-RV. In: DRV-Schriften, 79, pp. 17-33.
Steiner, V. (2017): The labor market for older workers in Germany. In: Journal for Labour Market Research, pp. 1-14.
Stock, J./Wise, D.A. (1990): Pension, the Option Value of Work, and Retirement. In: Econometrica, 58 (5), pp. 1151-1180.

Appendix

A Legal rules on the pension entitlement calculation: Thresholds for marginal employment

Table A1. Threshold for marginal employment by year in €, 1957–2014

Year	West Germany	East Germany	Year	West Germany	East Germany
1957	575.20		**1987**	2638.27	
1958	575.20		**1988**	2699.62	
1959	613.55		**1989**	2760.98	
1960	651.90		**1990**	2883.69	
1961	690.24		**1991 until 30 June**	2945.04	1349.81
1962	728.59		**1991 from 1 July**	2945.04	1533.88
1963	766.94		**1992**	3067.75	1840.65
1964	843.63		**1993**	3251.82	2392.85
1965	920.33		**1994**	3435.88	2699.62
1966	997.02		**1995**	3558.59	2883.69
1967	1073.71		**1996**	3619.95	3067.75
1968	1227.10		**1997**	3742.66	3190.46
1969	1303.79		**1998**	3804.01	3190.46
1970	1380.49		**1999**	3865.37	3865.37
1971	1457.18		**2000**	3865.37	3865.37
1972	1610.57		**2001**	3865.37	3865.37
1973	1763.96		**2002**	3900.00	3900.00
1974	1917.34		**2003 until 31 March**	3900.00	3900.00
1975	2147.43		**2003 from 1 April**	4800.00	4800.00
1976	2377.51		**2004**	4800.00	4800.00
1977	2607.59		**2005**	4800.00	4800.00
1978	2392.85		**2006**	4800.00	4800.00
1979	2392.85		**2007**	4800.00	4800.00
1980	2392.85		**2008**	4800.00	4800.00
1981	2392.85		**2009**	4800.00	4800.00
1982	2392.85		**2010**	4800.00	4800.00
1983	2392.85		**2011**	4800.00	4800.00
1984	2392.85		**2012**	4800.00	4800.00
1985	2454.20		**2013**	5400.00	5400.00
1986	2515.56		**2014**	5400.00	5400.00

(Source: Deutsche Rentenversicherung Bund (2016c), p. 264)

B Sample selection and sample characteristics

For our analyses, we use the BASiD 5109 (Biographical Data of Selected Social Insurance Agencies in Germany) to generate three datasets (see Table 2). First, we make basic restrictions that apply to all three BASiD datasets. We remove people who do not retire before 2007, self-employed people, civil servants, and miners. We additionally delete persons without an old-age pension but with a pension for surviving dependants or a reduced earnings capacity pension, because we do not have the information needed for the specific construction of these pension entitlements in conventional datasets. Furthermore, we concentrate our report on the 1940–1942 birth cohorts. After these adjustments, we get the BASiD dataset in the left-hand column of Table 2. In our BASiD Samples I and II, we introduce further restrictions in addition to those in the BASiD dataset described in Section 3.

For the first part of the report, the evaluation of the functionality and quality of the recalculation of the earning points, we create the BASiD Sample I and make the following additional restrictions: We drop people who are unemployed or out of labour force for more than 50% of the observation period, those who insure themselves completely voluntarily without other pensionable income, and those with less than five contribution years during the observation period. This means that people with extremely large gaps in their data records and people who have only brief periods on the labour market are deleted. These individuals are often not entitled to a pension under the German pension system or they do not have the choice whether to work or to retire at old age, LORENZ ET AL. (2018). These employees therefore are not of particular interest in most research questions concerning the retirement decision and the retirement transition.

BASiD Sample II is used in order to check whether the errors found for the BASiD Sample I can be transferred into the SIAB environment. BASiD Sample II thus corresponds to the information situation in the SIAB, which is presented in detail in LORENZ ET AL. (2018). Specifically, for BASiD Sample II, we further restrict the BASiD Sample I as follows: We delete all persons who do not have at least one employment period subject to social security contributions before their 55th birthday, those whose last labour market activity is not completed prior to age 59 and those whose first labour market activity is not observable before or at the age of 41. In addition, all individuals whose reason for the end of their last labour market observation is death are removed. The last step is taken in order to delete people whose compulsory contribution status does not appear in the last 10 years before leaving the labour market. The most important limitation, however, is probably restricting the observations to the years of 1975–2007. The SIAB can only be observed from 1975 onwards, and the BASiD is only available until 2007. Individual labour market biographies can therefore be considered over a maximum of 32 years in the BASiD Sample II.

Table B1. Overview of sample characteristics in the BASiD, BASiD Sample I and BASiD Sample II

	BASiD (1957–2007)	BASiD Sample I (1957–2007)	BASiD Sample II (1975–2007)
	mean (sd)	mean (sd)	mean (sd)
Share of full-time employment[a]	35.68 (0.2626)	41.69 (0.2546)	70.97 (0.2847)
Share of part-time employment	10.94 (0.1729)	11.06 (0.1794)	17.53 (0.2097)
Share of marginal employment	3.34 (0.0772)	2.39 (0.0571)	1.21 (0.0225)
Share of unemployment	5.86 (0.1013)	5.63 (0.0847)	6.20 (0.1390)
Share of illness	0.76 (0.0171)	0.82 (0.0166)	0.74 (0.0164)
Share of maternity and care-giving periods	18.84 (0.2506)	15.03 (0.2137)	5.53 (0.1441)
Share of people with children	49.87	45.56	42.58
Average number of children (of those with children)	2.29 (1.2164)	2.17 (1.1490)	2.13 (1.0895)
Average number of children (all)	1.14 (1.4300)	0.99 (1.3298)	0.89 (1.2628)
Share of the severally disabled	9.19	8.60	7.16
Average retirement age	62.51 (2.1729)	62.27 (2.1172)	62.22 (1.906)
Average EP per year	0.8151 (0.5384)	0.8661 (0.5214)	1.0502 (0.5265)
Average sum of EP	30.5190 (19.2912)	33.1371 (20.3406)	30.0861 (14.0696)
Share of people with early retirement	40.26	39.07	61.16
Share of people with normal retirement	34.72	32.77	24.88
Share of people with late retirement	25.02	28.16	13.96
Share of people with bridge retirement[b]	27.83	22.27	32.64
N	10,883	7269	4204

Notes: [a]Shares over the total period of labour market observation, with standard errors in parentheses. [b]'Bridge retirement' means postponing retirement and bridging employment with alternatives, such as unemployment, until the early retirement age (ERA) or the normal retirement age (NRA). Individuals in this category are included either in the number of people with an NRA or in the number of people with an ERA.
EP means earning points. N is the number of individuals per group.
Data: BASiD 5109, Samples I and II

We see remarkably age-specific retirement patterns when the sample is split by retirement age. There are obvious spikes for women and men at ages 60, 63 and 65. The spike at age 65 corresponds with regular old-age pension. The old-age pension for unemployed or persons

under a progressive retirement offers the earliest possible option for men to leave the labour market at age 60 for the cohorts considered in this report.[96] For women, the old-age pension for women is the most advantageous pension type, with an ERA of 60 years.[97] Age 63 is the earliest option to retire for those who are eligible for the old age pension for long-term insured.

If we look at the other variables besides labour market exit type, we can see few striking differences between the three samples. The proportions of the various labour market states in the BASiD Sample II differ in comparison to the BASiD and BASiD Sample I because more observations at the beginning of the career are missing. The shares of labour market states do not add up to 100%, because labour market activity also includes school education and vocational training, active labour market measures and unemployment without registration with the Federal Employment Agency. These periods have a share of approximately 20 percentage points in the BASiD and BASiD Sample I and only a few percentage points in Sample II. The reason for this difference is again that these labour market states typically occur in the first years of the career. Also the proportion of maternity and caregiving periods is therefore lower in BASiD Sample II than in the other two samples, but the proportion of women with children and the average number of children is very similar in all three samples. The proportions of full-time and part-time employment are correspondingly higher in the BASiD and BASiD Sample II.

[96]See Lorenz et al. (2018) for a detailed list of the ERA and NRA per cohort and the presentation of retirement patterns by date of birth.
[97]This conclusion is also reached by Börsch-Supan and Ferrari (2017), p. 5; Seibold (2017), p. 69 and Grogger and Wunsch (2012), p. 4.

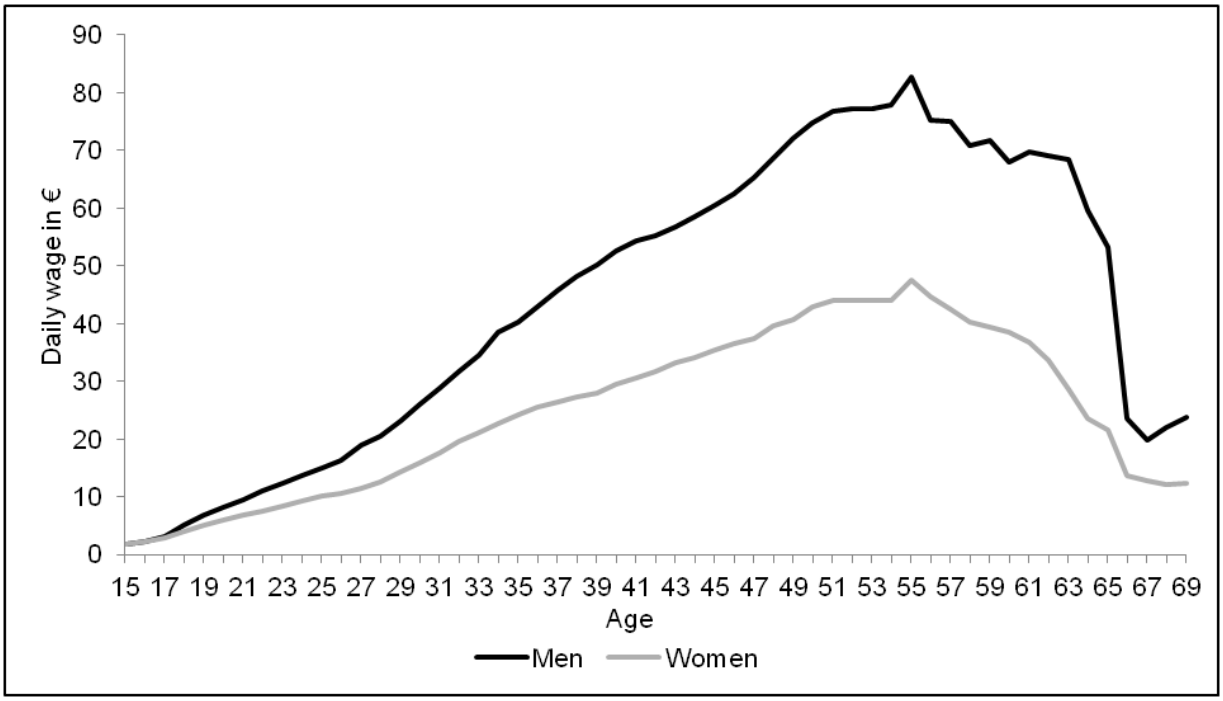
C Further sample characteristics: BASiD (Biographical Data of Selected Social Insurance Agencies in Germany)

Table C1. Overview of the shares of individuals by retirement entry path, BASiD Samples I and II

	Sample I	Without East Germany	Without large labour market gaps	Without women	Sample I	Without East Germany	Without large labour market gaps	Without women
	1	2	3	4	5	6	7	8
Average retirement age	62.27	62.27	61.53	61.97	62.22	62.21	62.16	62.55
Share of full-time employment[a]	41.69	41.81	47.00	57.10	70.97	71.40	72.68	79.23
Share of part-time employment	11.06	11.09	11.39	2.54	17.53	17.67	17.06	3.16
Share of unemployment	5.63	5.60	5.58	5.10	6.20	6.10	5.88	5.65
Share of marginal employment	2.39	2.39	1.79	1.39	1.21	1.21	1.13	0.83
Number of people with NRA[b]	2382 (32.77)	2346 (32.80)	887 (17.75)	383 (15.31)	1046 (24.88)	1027 (24.85)	950 (24.04)	399 (18.91)
Number of people with ERA	2840 (39.07)	2788 (38.98)	2354 (47.10)	1084 (43.34)	2571 (61.16)	2531 (61.25)	2471 (62.54)	1636 (77.54)
Number of people with retirement after NRA	2047 (28.16)	2018 (28.22)	1757 (35.15)	1034 (41.34)	587 (13.96)	574 (13.89)	530 (13.41)	75 (3.55)
Number of people with bridge retirement[c]	1619 (22.27)	1897 (26.52)	955 (19.11)	708 (28.31)	1372 (32.64)	1346 (32.58)	1285 (32.52)	754 (35.73)
N	7269	7152	4998	2501	4204	4132	3951	2110

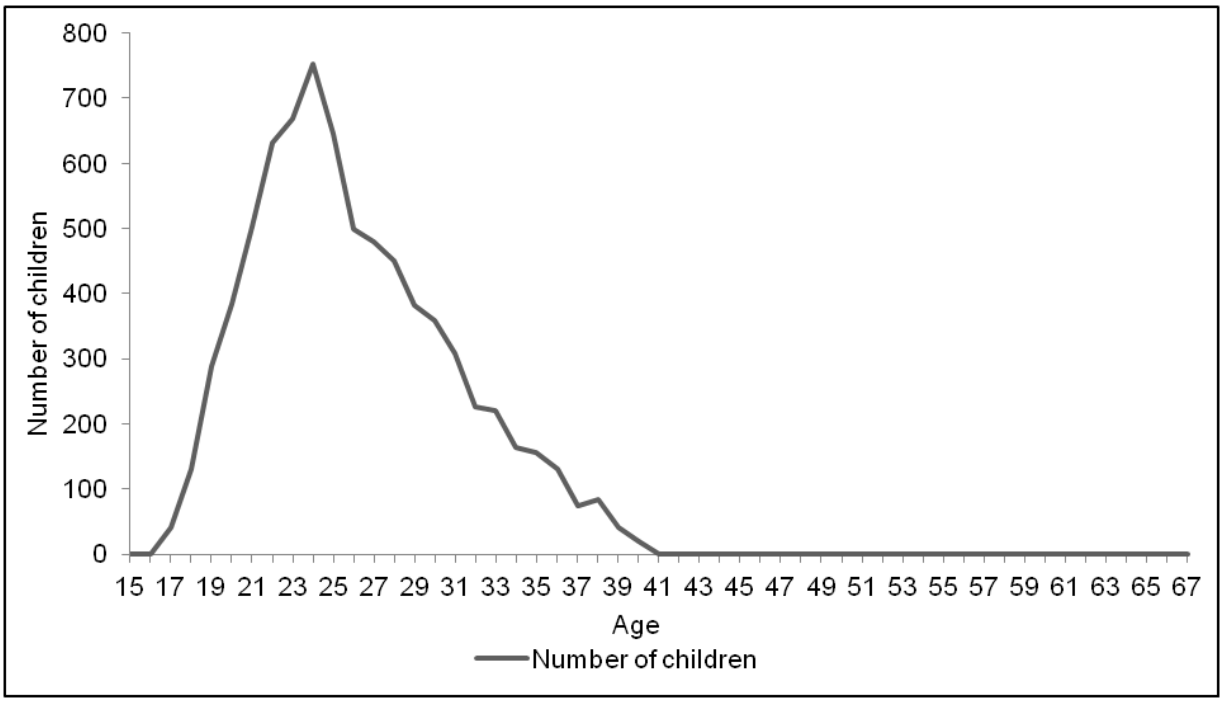
Notes: [a]Shares over the total period of labour market observation.
[b]Shares of persons in total number of persons per group in parentheses.
[c]'Bridge retirement' means postponing retirement and bridging employment with alternatives, such as unemployment, until the early retirement age (ERA) or the normal retirement age (NRA). Individuals in this category are included either in the number of people with an NRA or in the number of people with an ERA.
N is the number of individuals per group.
Data: BASiD 5109, Samples I and II

Figure C1. Development of average daily wage by age



(Source: own illustration based on BASiD 5109, Sample I)

Figure C2. Total number of children born by age of parent



(Source: own illustration based on BASiD 5109, Sample I)

Table C2. Number of people by birth cohort and year of retirement (BASiD Sample I)

Birth cohort	Men	Women	All	Year of retirement	Men	Women	All
1940	1087	1240	2327	**2000**	387	579	966
1941	1117	1326	2443	**2001**	418	644	1062
1942	1142	1357	2499	**2002**	462	658	1120
				2003	360	365	725
				2004	371	261	632
				2005	579	525	1104
				2006	365	453	818
				2007	404	438	842
N = 7269							

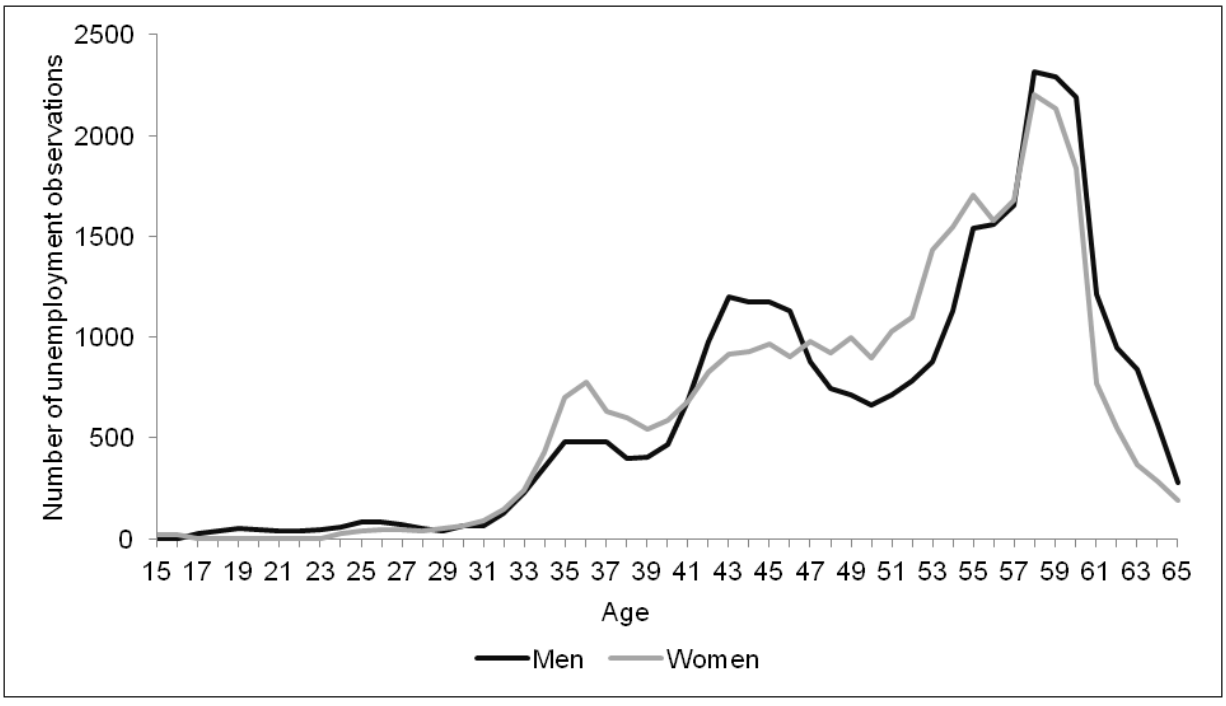
(Data: BASiD 5109, Sample I)

Table C3. Number of people by birth cohort and year of retirement (BASiD Sample II)

Birth Cohort	Men	Women	All	Year of retirement	Men	Women	All
1940	715	635	1350	**2000**	149	257	406
1941	753	708	1461	**2001**	257	340	597
1942	733	660	1393	**2002**	340	403	743
				2003	291	290	581
				2004	312	226	538
				2005	367	217	584
				2006	234	150	384
				2007	251	120	371
N = 4204							

(Data: BASiD 5109, Sample II)

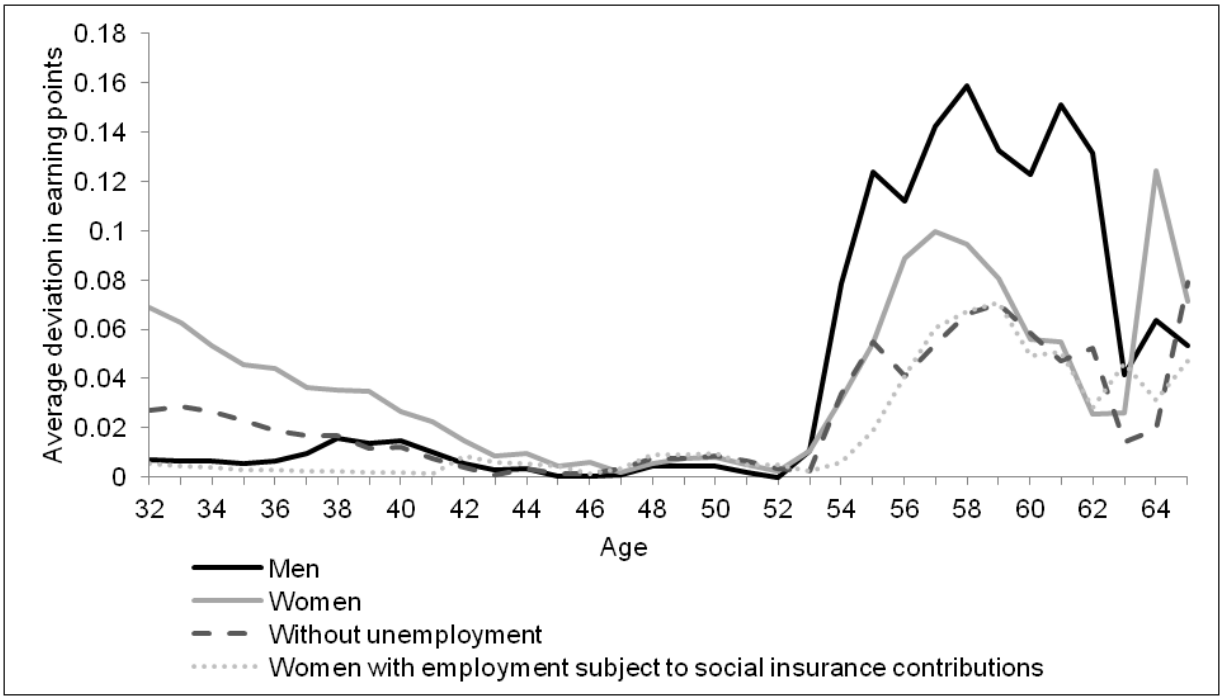
Figure C3. Total number of unemployment observations by age



(Source: own illustration based on the BASiD 5109, Sample I)

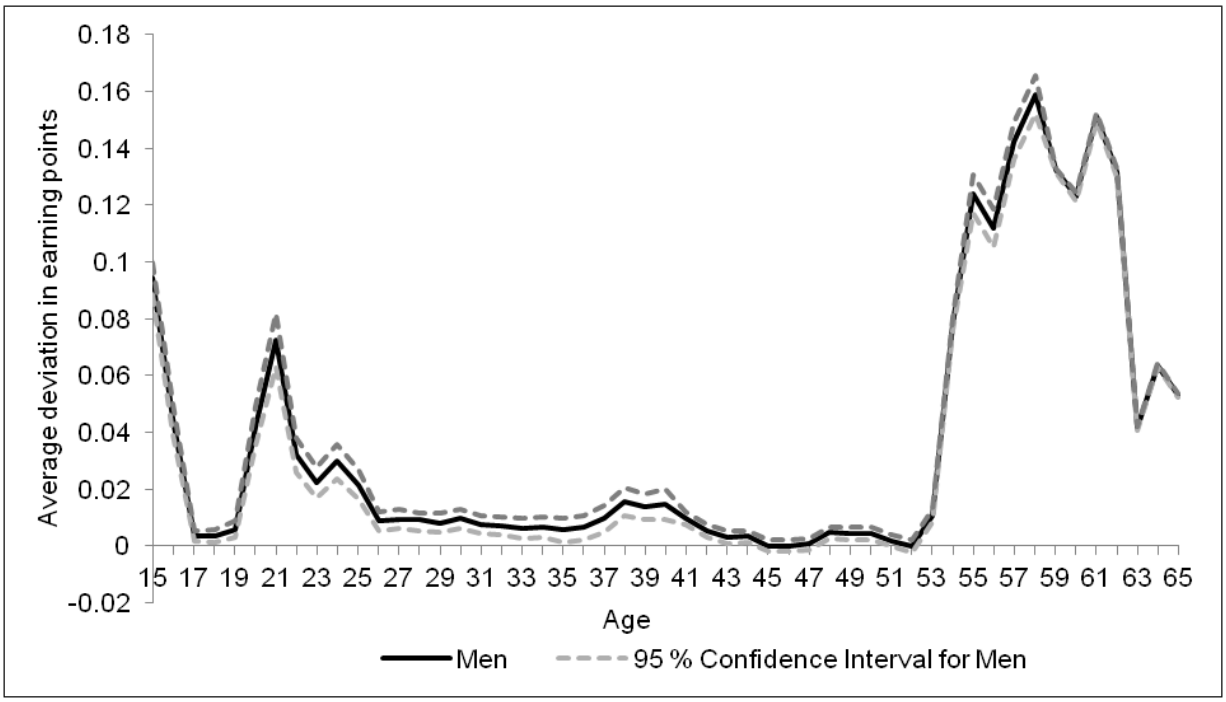
D Description of the average deviation in earning points in the BASiD (Biographical Data of Selected Social Insurance Agencies in Germany)

Figure D1. Development of the average deviation in earning points for selected groups, BASiD Sample II



(Source: own illustration based on the BASiD 5109, Sample II)

Figure D2. Development of the average deviation in earning points by age for men, with 95% confidence intervals, BASiD Sample I



(Source: own illustration based on the BASiD 5109, Sample I)

Table D1. Robustness check of the multivariate analysis of the annual deviations in BASiD Sample II

	Sample II	
Variable	**Coefficient**	**P > |t|**
Female	0.1014050	0.647
East	0.0391721	0.668
Childcare	0.4539108	0.000
No children	0.0003197	0.994
1 child	0.0039742	0.926
2 children	0.0168010	0.690
3 children	0.0210145	0.650
5 children	0.0407861	0.633
6 children	0.3017173	0.106
> 6 children	−0.2108689	0.146
Unemployment	0.1608934	0.000
Marginal employment	−0.4489412	0.000
Cohort dummies	Yes	
Year dummies	Yes	
Age dummies	Yes	
Number of observations	111,397	
R^2	0.0025	

Notes: Sample II without further limitations. All variables are dummy variables. Negative signs mean overestimation of earning points, and positive signs mean underestimation of earning points.
The dummy variable ‘4 children’ serves as the reference group.
Data: BASiD 5109, Sample II

Table D2. Robustness check of the multivariate analysis of the sum of the deviations in BASiD Sample II

	Sample II	
Variable	**Coefficient**	**P > |t|**
Female	−2.3818720	0.000
East	−0.5894954	0.120
No children	−2.4648380	0.000
1 child	0.2324170	0.499
2 children	0.2284843	0.490
3 children	0.2286557	0.532
5 children	0.4511234	0.443
6 children	−0.6397832	0.627
> 6 children	0.8500750	0.353
Unemployment	−0.1082482	0.376
Marginal employment	3.0590520	0.000
Severe disability	0.4187533	0.008
Cohort dummies	Yes	
Number of observations	242,908	
R^2	0.0017	

Notes: Sample II without further limitations. All variables are dummy variables. Negative signs mean overestimation of earning points, and positive signs mean underestimation of earning points.
The dummy variable ‘4 children’ serves as the reference group.
Data: BASiD 5109, Sample II

E Further sample characteristics: SIAB (Sample of Integrated Labour Market Biographies)

Table E1. Number of people by birth cohort and year of retirement (SIAB)

Birth cohort	Men	Women	All	Year of retirement	Men	Women	All
1936	2706	1123	3829	**1996**	240	208	448
1937	3146	1359	4505	**1997**	664	383	1047
1938	3597	1672	5269	**1998**	962	565	1527
1939	3990	1860	5850	**1999**	1877	823	2700
1940	4234	2282	6516	**2000**	2480	1085	3565
1941	4104	2625	6729	**2001**	2897	1387	4284
1942	3479	2396	5875	**2002**	3107	1622	4729
1943	3558	2588	6146	**2003**	2975	1679	4654
1944	3475	2604	6079	**2004**	2748	1726	4474
1945	2678	1843	4521	**2005**	2852	1888	4740
1946	3142	2349	5491	**2006**	2612	1781	4393
1947	3632	2627	6259	**2007**	2992	1989	4981
1948	3995	2894	6889	**2008**	3155	2130	5285
				2009	2705	1901	4606
				2010	2370	1743	4113
				2011	2326	1496	3822
				2012	1955	1266	3221
				2013	1784	1197	2981
				2014	5035	3353	8388
N = 73,958							

(Data: SIAB 7514)

Imprint

FDZ-Methodenreport 01/2018 (EN)

Publisher
The Research Data Centre (FDZ)
of the Federal Employment Agency
in the Institute for Employment Research
Regensburger Str. 104
D-90478 Nuremberg

Editorial staff
Dana Müller, Dagmar Theune

Technical production
Dagmar Theune

Download
http://doku.iab.de/fdz/reporte/2018/MR_01-18_EN.pdf

Internet
http://fdz.iab.de/

Corresponding author:
Mona Pfister
Chair for Human Resource Management
and Organisation
Sanderring 2
97070 Würzburg
Phone: +49 931 31-89274
Email: mona.pfister@uni-wuerzburg.de