

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

FDZ

FDZ-Methodenreport

Methodological aspects of labour market data

13/2017

EN

Identifying mothers in administrative data

Dana Müller,
Katharina Strauch



Bundesagentur für Arbeit

Identifying mothers in administrative data

Dana Müller (FDZ)

Katharina Strauch

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 the methodical aspects of FDZ data and thus help users in the analysis of data. In addition, through this series users can publicise their results in a manner which is citable thus presenting them for public discussion.

Content

Zusammenfassung	4
Abstract	4
1 Introduction	5
2 Used administrative information	6
3 Identification strategy	7
4 Conclusion	9
References	10
Appendix	11

Zusammenfassung

Die administrativen Daten der Bundesagentur für Arbeit bieten eine wichtige Datenbasis für die Arbeitsmarktforschung. Welche Informationen gesammelt werden, ist über die Aufgaben der Bundesagentur für Arbeit definiert. Daher sind nicht alle Informationen in den Daten enthalten, die für verschiedene Forschungsfragen relevant sind. Das betrifft zum Beispiel Informationen zu der Geburt von Kindern, die wichtig für die Analyse der Erwerbsbiografien von Frauen sind. Nach wie vor unterbrechen insbesondere Mütter ihre Erwerbstätigkeit, um sich der Kinderbetreuung zu widmen. Diese Erwerbsunterbrechungen können unterschiedliche Effekte auf die Erwerbsbiografien von Müttern haben, wie z.B. Lohneinbußen, Karrierenachteile oder vermehrte Teilzeitbeschäftigung. Mit diesem FDZ-Methodenreport zeigen wir eine Möglichkeit auf, familienbedingte Erwerbsunterbrechungen mit Hilfe indirekter Identifikatoren in den administrativen Daten zu ermitteln. Wir prüfen unsere Identifikationsstrategie mit Hilfe offizieller Geburtsstatistiken. Der Programmcode wird als Anhang zur Verfügung gestellt und kann nach Bedarf angepasst werden.

Abstract

Administrative data of the Federal Employment Agency provide an important data base for labour market research. However, the Federal Employment Agency's tasks define what kind of information is recorded. Therefore, some information is missing that is relevant for various research purposes. One example is information on childbirth, which is necessary to analyse female employment biographies. It is still mainly mothers who interrupt their employment to care for their children. These employment interruptions could have different effects on the career of mothers, e.g. wage reductions, reduced career opportunities or an increase in part-time employment. In this FDZ-Methodenreport we present one way to identify family-related breaks using indirect identifiers in the administrative data. We use official statistics on births to validate our identification strategy. We make our program code available as an appendix for the research community. Users can adapt when necessary.

Keywords: childbirth, mothers, administrative data

The authors would like to thank Ann-Christine Bächmann, Johanna Eberle, Corinna Frodermann, Katharina Schneider, Philipp vom Berge and Jonas Jessen for helpful comments and suggestions as well as all users and IAB-colleagues who gave feedback of a preliminary version of the program code. Furthermore we would like to thank Florian Zimmermann for editing the Excel sheet and Dagmar Theune for editing the layout.

1 Introduction

Administrative data of the German Federal Employment Agency provide an important data base for labour market research. They include information on employment, unemployment benefits, participation in labour market programs and registered job search as well as sociodemographic and establishment information. In comparison to survey data, those administrative data have various advantages. Large sample sizes, mandatory notifications and a longitudinal observation period on a daily basis that spans several decades make sure that panel mortality and selective responses do not influence empirical results. Furthermore, the importance of the notifications for social security contributions and other administrative purposes usually ensure a high data quality.¹

However, the Federal Employment Agency's tasks define what kind of information is recorded. Therefore, some information is missing that is relevant for various research purposes. One example is information on childbirth, which is necessary to analyse female employment biographies. It is still mainly mothers who interrupt their employment to take care of their children. These employment interruptions could have different effects on the career of mothers, e.g. wage reductions or an increase in part-time employment.

While the administrative data of the Federal Employment Agency do not have any direct information on childbirth, other information can be used to identify them indirectly. One possible strategy is to identify employment interruptions due to family formation as it is described by Schönberg (2008). She uses the meanwhile outdated IAB Employment Sample with an observation period from 1975 until 2001. This data contains a variable ("btyp") that indicates employment interruptions. Furthermore, Schönberg (2008) tests her results with linked pension register data that include children's date of births. Her results show that it is possible to identify nearly 90% of childbirths in West Germany using the administrative data if the age of mothers falls into a specific range. However, the identification of first childbirth during un/employment episodes is easier because further pregnancies often take place during parental leave or non-employment periods which are not observable in the data. Nevertheless, the "btyp" variable does no longer exist in the follow up data versions of the "Sample of Integrated Labour Market Biographies" (see Antoni et al. 2016).

In this FDZ-Methodenreport we present another way to identify family-related breaks. While Schönberg (2008) uses employment interruptions to identify mothers, we use an indirect identifier specifying the reason of cancellation/notification/termination of an employment or unemployment spell in the variable 'grund'. We use official statistics on births in Germany to validate our identification strategy.

The paper is structured into three parts. The following chapter describes the information we use from the administrative data for our procedure. Afterwards, we describe the identification

¹ This is not always the case for information that is not directly relevant for the calculation of pensions or benefits, for example information on schooling.

of mothers and expected childbirths in general (details are included in the available program code). We sum up our identification strategy in the last chapter.

2 Used administrative information

There are two main ways to identify maternity in the administrative data, one using employment notifications and one using process data of the Federal Employment Agency about unemployment and benefits. The relevant information can be found in the variable „grund“, the reason of cancellation/notification/termination of a spell.

Identification from employer's notifications uses employment interruptions due to entitlement to other compensation by the statutory health insurance provider (value 51 of 'grund'). Unfortunately, there are two reasons to file this kind of notification that cannot be easily separated. First, maternity allowance is paid by the statutory health insurance provider during paid maternal leave. Second, sickness allowance is paid by the provider when a person is absent due to illness for more than six weeks. Although we cannot clearly differentiate between long-term sickness and maternity leave in the administrative data, it is more likely to become a mother at a certain age. At that age it is also less likely to suffer long periods of disease. According to Meyer et al. (2015), these are more likely for older workers. Our identification strategy therefore rests on the assumption that women in a specific age bracket do not face employment interruptions due to long-term sickness.

The time period of paid maternity leave is determined by the estimated birth date from the mother's booklet. The employer informs the employee of the reason for the employment interruption once maternity leave has started. The employee receives 13 euros per day from the statutory health insurance provider as maternity allowance. The employer makes an additional payment so the total compensation equals the previous net wage. Social insurance contributions, both by the employer and the employee, are not due during paid maternity leave. Pregnant women who are privately insured can apply for a one-time maternity benefit of 210 euros.

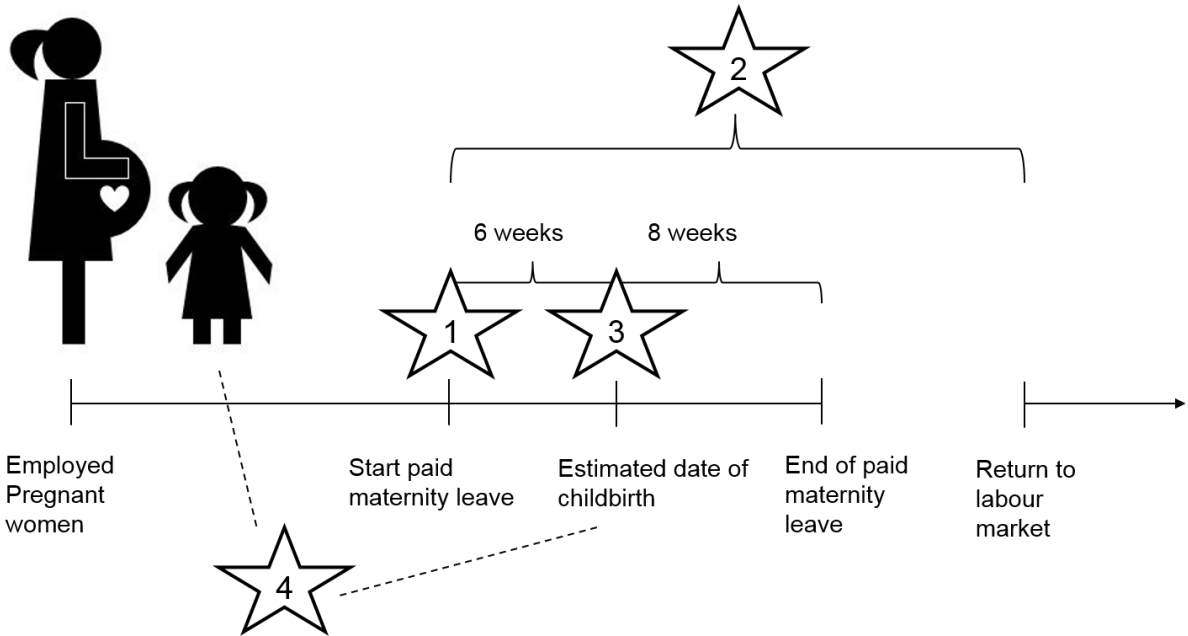
Identification from non-employment information in the administrative data uses the withdrawal into maternity allowance as part of the internal procedures of the Federal Employment Agency. This deregistration takes place when women are unemployed and receive unemployment benefits according to Social Code Books II or III. Unemployed women who received unemployment benefits I before maternity leave receive maternity allowance in the amount of the previous benefits. The allowance is paid by the statutory health insurance provider instead of the Federal Employment Agency because a pregnant mother is no longer available to the labour market due to the employment ban during maternity leave. Women who work and are additionally dependent on unemployment benefits II also receive maternity allowance. Women who depend solely on unemployment benefits II (i.e. welfare) do not receive maternity allowance.

The availability of the data and the logic of the notification procedures lead to additional caveats that researchers need to keep in mind when using the proposed algorithm: (i) we can identify (expected) childbirths only if women have a record in the administrative data sources (so that we do not capture maternity leave of self-employed, public servants or before women enter the labour market), (ii) we cannot differ between live birth and stillbirth and (iii) the number of children per woman can be underestimated since twin births are not identifiable and successive births are often hard to detect (for example when a mother has no employment episode between having two children).

3 Identification strategy

There are four steps in the identification of mothers and (expected) childbirths which are pictured in graph 1 and explained step by step below. More detailed information is included in the program code. The program code can be used for all administrative datasets at the Research Data Centre. However, the identification strategy is only applicable with the weakly anonymous versions which can be used via on-site use or remote execution. The Scientific Use File of the SIAB for instance includes mixed aggregated information of the variable 'grund' which might lead to a misspecification of mothers. We tested the program code with the Sample of Integrated Employment Biographies (SIAB 7514 v1) using Stata 14.

Graph 1: Steps of mother's respectively (expected) childbirth's identification



The first step marked with a star and the number 1 in the graph 1 identifies the beginning of paid maternity leave. The entry in paid maternity leave is recorded in the variable 'grund' as mentioned above. We mark all deregistrations with the value 51, which come from the notification procedure of the social security system and are filed 'due to other entitlements from

statutory health insurance'. Value 51 is included in the source 'Employment History'. Furthermore, we mark all deregistrations into maternity leave and maternity allowance with the values 2015, 3002 and 6010, which come from process data of the Federal Employment Agency and are included in the source 'Benefit Recipient History', 'Unemployment benefit II Recipient History' and 'Jobseeker History'. We do not taken into account the value 52 of 'grund' which includes the notification 'childcare leave'. This notification is filed by employers when mothers take the last year of their parental leave at a later date, which is possible until the child reaches its 8th birthday. At this stage, we consulted childbirth statistics of the Federal Statistical Office in order to restrict the plausible childbearing age. The number of identified age-specific childbirths in our data needs to be lower than the number of age-specific childbirths of the official statistics because not all women are recorded in the administrative data of the Federal Employment Agency. Using an iterative process we determined the maximum childbearing age to be 40. After comparing our outcomes with the childbirth statistics of the Federal Statistic Office, we noted that it was necessary to add another restriction: we ignore cases where the first childbirth occurs after the 38th birthday of an employed women.

The second step marked with a star and the number two in graph 1 involves the computation of the employment interruptions' duration to avoid misspecification due to long-term sickness. Therefore, each interruption needs to be longer than 97 days (duration of maternity leave). We also take into account women who do not re-enter the labor market. In this case, deregistration is on the last observed spell. Here, we generate the employment interruption using the end of the observation period in the data. Furthermore, we create a separate variable to make these spells visible. All information with a shorter interruption than 98 days are replaced.

The third step includes the computation of the expected date of delivery. Only 4% of all childbirths are exactly at the expected date of delivery. Researchers need to be aware that an inexact date of childbirth may affect their research results. We use the original end date (endorig) and add 42 days (maternity leave starts six weeks before estimated childbirth) to compute childbirth.²

Last but not least, we check the gap between siblings in the fourth step. We do not use a hard limitation of 40 weeks (281 days) of pregnancy because the expected delivery is only correct for 4% of all childbirths. Furthermore, the preterm rate was 9.2% in 2010 for Germany (March of Dimes et al. 2012). Therefore, we use a more realistic measure and use a gap between siblings of at least 32 weeks (224 days)³. With our identification strategy, we generate 277.012 (expected) childbirths and 205.102 mothers, respectively in the SIAB 7514.

Finally, we prepare some variables that might be useful for research purposes like the number of children per episode, the number of children and a dummy variable for being mother or

² We cannot use the end date of the episode (endepi) because overlapping periods of employment or non-employment have different durations and time specific start and end dates after episode splitting.

³ In the program we produce both measures of 281 days and 224 days to show how many deliveries are affected.

childless. Researchers need to be aware that we underestimate the number of children for the reasons mentioned in chapter 2.

We validate our identified (expected) childbirth with the official statistics of the Statistical Office of Germany using annual statistics on childbirth by age of mothers. We extrapolate the numbers of (expected) childbirth with a factor of 50 because the SIAB 7514 is a 2% subsample of the Integrated Employment Biographies. The results are presented in an Excel sheet which is available online (see online appendix). The results show that the numbers of generated childbirths are smaller than the number of childbirths in the official statistics. This is not surprising because we do not have any information on civil servants and self-employed. Furthermore, we can only identify a childbirth if the mother has a record in the administrative data sources.

4 Conclusion

In this FDZ-Methodenreport we describe a strategy to identify mothers respectively (expected) childbirths in the administrative data of the Federal Employment Agency by using indirect identifiers. We validate our results with official statistics on childbirths. Our identification strategy has some limitations. We cannot differ between life birth and stillbirth. We only identify (expected) childbirths if the mother has a record in the data sources we use (we do not have information on civil servants and self-employed). Furthermore, we have to limit our algorithm to a certain range of the possible child-bearing age of mothers to avoid misspecification. Nonetheless, our identification strategy is helpful for researchers who want to use the administrative data of the Research Data Centre for their research purposes in the context of female employment biographies.

References

Antoni M.; Ganzer A.; vom Berge P. (2016): Sample of integrated labour market biographies (SIAB) 1975-2014. FDZ-Datenreport, 04/2016 (en).

March of Dimes, PMNCH, Save the Children, WHO (2012): Born Too Soon: The Global Action Report on Preterm Birth. Eds CP Howson, MV Kinney, JE Lawn. World Health Organization. Geneva. http://www.who.int/pmnch/media/news/2012/201204_borntoosoon-report.pdf (accessed on October 31, 2017).

Meyer M.; Böttcher M.; Glushanok I. (2015): Krankheitsbedingte Fehlzeiten in der deutschen Wirtschaft im Jahr 2014. In: Badura B., Ducki A., Schröder H., Klose J., Meyer M. (eds) Fehlzeiten-Report 2015. Fehlzeiten-Report, vol 2015. Springer, Berlin, Heidelberg. DOI https://doi.org/10.1007/978-3-662-47264-4_30.

Schönberg, U. (2008): Does the IABS reliably identify maternity leave taking? FDZ-Methodenreport, 03/2008 (en).

Appendix

Download of the Stata do-file:

- http://doku.iab.de/fdz/reporte/2017/MR_13-2017_EN_programs.zip

Download of the Excel sheet with the comparison of official statistics on childbirth by age of mother and childbirth by age of mother generated with the SIAB 7514:

- http://doku.iab.de/fdz/reporte/2017/MR_13-2017_EN_comparison_of_official_statistics.zip

Imprint

FDZ-Methodenreport 13/2017 (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

All rights reserved

Reproduction and distribution in any form, also in parts,
requires the permission of FDZ

Download

http://doku.iab.de/fdz/reporte/2017/MR_13-17_EN.pdf

Internet

<http://fdz.iab.de/>

Corresponding author:

Dana Müller
Institute for Employment Research (IAB)
Research Data Centre (FDZ)
Regensburger Str. 104
D-90478 Nürnberg
Phone: +49-911-179-2409
Email: Dana.Mueller@iab.de