Nordic Seminar on Labour Market Statistics 2005 7 September 2005
Session 5: Absence due to sickness
Daisy Kyed, Statistics Denmark

New National Statistics on Absence in Denmark

Abstract
New national statistics on absence due to sickness have been in the process of development for the last year. During the next three years, the statistics will be gradually set up for completion at the end of 2007. The statistics will eventually cover the entire labour market, consisting of both the public and private sector. The main purpose of this paper is to demonstrate some of our considerations in the process of defining four of the central concepts. It is, however, necessary to begin the paper with a short presentation of the background, purpose, structure, concepts, and measures that form the foundation of the statistics. Due to its limited range, the presentation cannot be seen as exhaustive.

This paper will - if not give answers to - at any rate illustrate our considerations in solving the four questions of definition. To which year does a period of absence belong, when it begins in one year and ends in another? How does one define a full day of work, when full-time employment consists of a different number of working hours for different people? Which variables define the degree of employment of a person in an unambiguous way? Except for the ordinary days of vacation that an employee is entitled to, which other days should be excluded from the possible days of work?

1. Background

First initiatives

In the spring of 2000 a working group was formed at the request of the Ministry of Social Affairs. The purpose of this group was to explore the possibilities for establishing statistics on absence covering the entire Danish labour market. The considerations of the group resulted in a report that was finished in August 2001. This report not only describes the process of establishing statistics on absence, but also defines the populations and measures of absence. The development of the statistics then came to a standstill due to a reorganisation of the fields of responsibilities between the ministries. Now the initiative had to come from the Ministry of Employment.

Renewed foundation

The statistics on absence were reborn in December 2003, when the government put them on its plan of actions. The government proposed the establishment of new national statistics on absence. That is, statistics that covered both the private and the public sector, and included all causes of absence.

Final agreement

The Ministry of Employment and Statistics Denmark entered into a contract on the establishment and the future maintenance of the statistics in September 2004. The agreement was to follow the recommendations in the report from August 2001. Eventually, the development of the statistics could begin.

2. Purpose

Main purpose

The main purpose of the statistics on absence is to describe on a yearly basis the amount of work that is lost due to paid absence.

3. Time horizon

Three-stage establishment

From the year 2007, the statistics must cover the governmental, the municipal and the private sector entirely. The statistics are set up in three stages. The first data is published at the end of 2005, covering the year 2004 and only for the governmental sector. The year after in 2006, the municipal sector is included. Finally, the full-blown statistics are published for the very first time at the end of 2007, providing the full scale product for the year 2006.

4. The three sectors

Varying workload

The workload is varying in the establishment of the statistics in the three sectors. An overview of some of the challenges when collecting data is given in table 4.1.

The governmental sector

The data for the governmental sector already existed prior to the birth of these statistics. They had been collected by the 'Personalestyrelse' which is part of the Ministry of Finance. For several years, the 'Personalestyrelse' has published statistics on absence for this sector, with fewer indicators of

absence, however. Furthermore, comparison between their statistics and statistics on absence made in the private sector has not been possible.

Table 4.1

Sector	Cooperation partners	Scope	Data	Extra burden to respondents
Government	Personalestyrelsen	Population (full scale)	Existent	None
Municipal	FLD	Population (full scale)	Partly existent	Minimal
Private	DA FA	Sample	Existent to a limited degree	Some

The municipal sector

In the municipal sector, different statistics on absence have already been developed. These are, however, only compiled on a disaggregated level. The municipalities are using absence registration systems, which makes the extra burden to respondents of this data collection minimal. The market for data registration via a HR (human resource) system is primarily split between two big actors, KMD and SD. The contact to these system administrators has been made in association with FLD; an actor in the municipal sector. The aim of FLD is to develop statistics on absence that can satisfy the demands for municipality specific data made by the municipal sector.

The private sector

The data collection in the private sector is the greatest challenge. As previously mentioned, the employer unions, DA and FA, have published statistics on absence for some years. There will be some extra respondent burden in this sector because our sample exceeds the number of firms that participate in these statistics. In attempts to limit the extra burden, we have made contact to the HR-systems in the market to make this a viable way of collecting and sending data. In addition, it will be possible to register and report data via spreadsheet, a portal on the internet called 'Virk', and the systems that DA and FA already are offering. The firms in the sample that are members of either DA or FA are to report to these. DA and FA will then forward the data to Statistics Denmark. This is done to avoid that the firms are asked to report the same data twice.

5. Concepts and measures

Quick overview

In the following, the concepts and populations will be defined. Figure 5.1 illustrates how the concepts are connected and thereby gives a quick overview of the contents of this sequence of the paper.

Concepts

One statistical unit is a single period of absence. The variables that form the indicators of absence are partly from the reported absence data and partly from the earnings data that Statistics Denmark already possesses. We match the periods of absence with the job to which it belongs. The key used for this purpose is an unambiguous identification of the job. More information about the match is to be found in part 6 of this paper. Statistics are compiled on all periods of absence that have begun or ended in the tax year.[1]

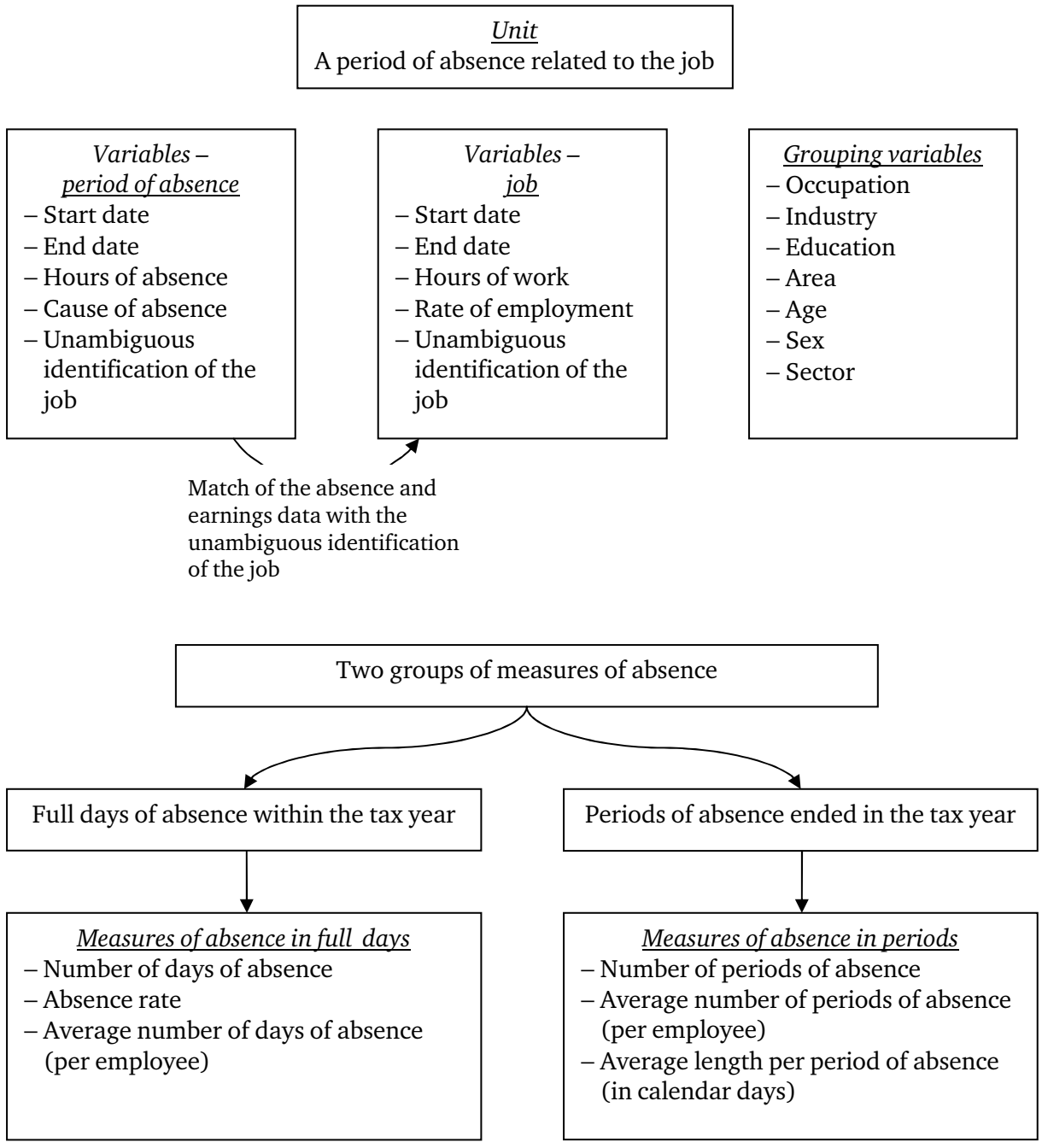
Measures of absence

We operate with two groups of measures of absence that are set apart by the statistical unit that we measure. In one group, we measure the lost days of

[1] The tax year is not always identical with the calendar year.

work that fall within the tax year. The lost days of work are measured in full day equivalents (referred to in the remainder of the paper as full days). In the other group, we examine the periods of absence that end in the tax year. All of the measures are grouped by variables like occupation, education, age, sex, and sector.

Figure 5.1



Absolute and relative measures

The measures can also be split into absolute and relative measures. The absolute measures are simply a count of days of absence and periods of absence, respectively. The relative measures include four measures, so far.

Absence rate

The rate of absence is the days of absence in percent of the possible days of work – both measured in full day equivalents. A discussion of the definition of the possible days of work is found in section 7.4.

Average number of days of absence

In order to get the average number of full days of absence, a measure of the number of employed in full-time equivalents is necessary. We find this measure by aggregating the actual number of possible days of work. This is then divided by the number of possible days of work for a full-time employed, employed for 12 months. By measuring per full-time employed, it is made possible to compare different groups with varying rates of employment and varying periods of the jobs over the year.

Average number of periods of absence

The average number of periods of absence is found by dividing the number of periods that end in the year by the number of full-year employed. The rate of employment is not taken into account in the measurement of the periods. We only measure the number of periods and the lengths of these. In order to make the numerator and the denominator compatible, the periods of absence are set relative to the number of full-year employed. This makes sense as we do not expect a part-time employed to have a lesser number of sickness cases than a full-time employed, given that the jobs are of equal length.

Average length of the period of absence

The length of the period of absence is measured either in calendar days or in working days. One could argue that most people and firms refer to the number of calendar days when we talk about the length of a sickness period. That is, when people talk about a fourteen-day period of sickness, they include weekends even though these are not working days.

6. Match with the data from the earnings statistics

Matching procedure

One of the critical points in the establishment of the statistics is the match between the absence and earnings data. The match is done with a key consisting of variables that can unambiguously identify the job and that is existing in both datasets. The variables forming the key are not the same in all three sectors. Though, the two variables in the key that all three sectors have in common are the civil registration number and an employee number of some sort. Choosing this method for arriving at the final dataset gives rise to both problems and advantages.

6.1 Problems

No-match of the individual

In the matching process some problems can arise if the periods of absence do not have a matching job. This means that the period cannot be attached to a person in the earnings data because no such person exists.

No-match of the dates

Problems with the dates can also arise in the process of matching. Some periods of absence do not lie within any job for the person. Even when the individual with the period of absence can be identified in the earnings data, we cannot be sure that the match is correct until we ensure that the period of absence is within the start and end date of the job.

Extent of no-match

It is hard to know the extent of these non-matches in advance. Therefore, we were relieved when we found that less than 5% of the data in the governmental sector suffered from these problems.

6.2 Advantages

Minimization of the burden of respondents

With the potential problems mentioned in section 6.1, the advantages must be eminent to justify choosing this method for arriving at the final dataset. The most important advantage is the minimization of the burden of respondents. Two thirds of the variables that we require are found in data that we already possess. Furthermore, we do not have any reason to believe that the amount of useable data will be greater if we asked respondents to report all of the variables. The extent of mistakes and sloppiness due to frustration with the heavy registration could very well exceed the extent of non-match problems.

7. Definition of central concepts

Differing definitions

When new statistics are developed, some questions about methods and definitions will arise. It is also hard to avoid differing definitions when cooperating with interested parties from all three sectors. In the following four sections, some of the considerations we have made are illustrated.

7.1 Inclusion of periods

Statistical period

The periods of absence that can be included in the statistics on absence are limited by the reporting period of the earnings statistics. It has been necessary to limit the statistical period to the tax year as not all reports on earnings data are completed at the end of the calendar year. Some of the earnings data run from the middle of December in one year to the middle of December in the next. This is the case for some employees paid on an hourly basis. The absence data for the individual naturally has to follow the earnings data in order to make matching feasible.

Periods starting or ending in another year

We operate with both days of absence and periods of absence as statistical units of absence. This was seen in chapter 5. Cases where periods of absence begin in one year and end in the following will occur in the data. For this reason, we need a definition of which periods to include.

Four criteria

Theoretically, one can choose between four criteria for which periods to include in a given tax year. The first three are based on inclusion of periods as a whole. The first possible criteria for inclusion could be to include all periods beginning in the year. The second, on the other hand, could be to include periods ending in the year. The third possibility is to use all complete periods within the year. Finally, one could include all days of absence within the year. This means that some periods must be divided between two years, however.

Two unrealistic criteria

Some of the mentioned solutions are only theoretical, in that they either cannot be implemented in reality or will result in underestimation of the absence. Statistics built on periods of absence beginning in the year are not a realistic choice because some time could and would pass before all of the periods are ended. Publication of such statistics would have to be years after the reference year. This criterion is therefore only theoretical. If we only include the periods that both start and end in the year, we underestimate the indicators of absence. The reason is that a period beginning in one year and ending in the next would not be counted.

Days of absence within the year

We have a problem in measuring the length and number of periods if we only include days of absence that fall within the year. Some periods will be split between two years. As a result, these periods would count as two and the length of them would be shortened. The indicators that are founded on days would on the other hand be very precisely related to the specific year. This is the reason why we have chosen to count the days that fall within the year. Such a definition of days gives rise to some challenges in programming due to the days in periods that have to be split between two years. Further complications arise because the statistical period is the tax year which differs from individual to individual. Thus, the separation point of periods that are split between two years is individual specific.

Periods of absence that end in the year

Alternatively, one could include all periods that end in the year. This is not as precise as the previous criterion when one wants to measure the days but it is a good solution when one wishes to measure the periods. We have selected this criterion for inclusion of periods. Another consideration is needed when looking at the periods. Some periods are very long because they began more than a year earlier than the beginning of the year in question. The credibility of these periods is small. It must be decided whether such periods should be discarded as outliers.

7.2 Definition of a full day

Standard scenario

A full day can be defined as one whole day of work or a whole day of absence for a person who is employed full time. An unambiguous definition of a full-timer is, however, hard to arrive at. The standard scenario in Denmark is an employee with 37 hours as the normal working hours. Such a person would be considered as a full-time employee. The person has 37/5 = 7.4 hours of work on a normal working day. If the same individual has a 20-hour period of absence he has 20/7.4 = 2.7 full days of absence.

Alternative definition

Persons who have 39 hours of work also exist. Is a full day of work for these people 39/5 = 7.8 hours? According to this line of thought, this person will have 20/7.8 = 2.56 full days of absence if he also has a 20-hour period of absence. Alternatively, one could define a full employment as always being 7.4 hours. In this case a 20-hour period of absence would always be handled as 2.7 full days of absence even if the person normally works 10 hours per day and therefore only has been absent 2 days.

Varying purposes of the concept

The choice of definition depends partly on how the hours are registered and partly on the purpose of the measure 'number of days of absence'. The definition of a full day of work as being 7.4 hours is correct if the users should be able to calculate costs of lost days of work in monetary measures. By multiplying the number of full days of absence with 7.4 they would get the number of hours of absence. It is then possible to multiply the hours with some measure of average labour cost per hour. This calculation does not make sense if the full days of absence are calculated with a varying definition of a full day of work. If, on the other hand, we want the measure of full days of absence to make sense when we look at individuals, the individual definition of a full day of work is necessary. If we use the standard definition of 7.4 hours then the measure for one full day of absence would not always be equal to one full day of work.

Registration of absence

Another aspect when defining a full day of absence is whether the person who registers the absence takes the normal hours of work of the individual into account or not. If he does take the normal hours of work into account, he will register 7.8 hours per day of absence for a person with 39 hours a week and 7.4 hours per day of absence for a person with 37 hours a week. Otherwise, he almost certainly registers 7.4 hours per day of absence.

7.3 Definition of jobs

Level of detail

Jobs can be defined more or less detailed, depending on the possibilities and requirements of the statistics. The more detailed the definition is, the more variables are needed. When more variables are used in the definition, the final number of jobs becomes higher even though we have the same number of individuals. The reason is, that more variables make it more likely that one of them changes which leads to at new job. If you, for instance, use occupation in the definition of the job then every shift in occupation will lead to a new job.

Problems with a very detailed definition

In the statistics on absence we match the period of absence to the job as explained in chapter 6. A large number of jobs for each person can result in problems when matching. The periods of absence can easily have a start date that is within one job period and an end date that is within another. This problem would be nearly impossible to solve - at least within a reasonable time frame.

Contents of the job

A simple definition of the job is obviously preferable but we also need to ensure that we can get an unambiguous match between the job period and the period of absence. Furthermore, each period of absence must have variables describing the extent of the employment. These variables are to be used as an important element in creating the measures of absence. To comply with this demand for information, the length of each job is indicated by a start date and an end date. The rate of employment and hours of work is also compiled for each job.

7.4 Definition of possible days of work

Differing views

A possibility of confusion is already present in the formulation 'possible days of work'. When the year begins, an employee anticipates a number of working days. This number could be defined as the number of calendar days minus Saturdays, Sundays and holidays. Some would consider this number as the number of possible days of work. Others would argue that vacation days and other days, which the employee has a possibility of taking, also should be excluded. The issue here is whether only days that are not optional days off should be excluded, or if optional days off should be excluded too.

Definition

The name of the concept, 'possible work days' is actually misleading in relation to the use of the concept. When calculating the rate of absence, the possible days of work make up the denominator. This is the reason why the *correct definition of possible days of work is the actual days worked plus the days of absence*. This is the number of days the individual actually would have worked had he not been absent due to sickness etc. In order to reach the correct definition, vacation days and other days off should be excluded.

Problems in collecting data

The definition is, however, not clear because a dataset that includes this information can be difficult to get. In many firms such days are not registered because the employee receives pay continuously.

8. Status

The governmental sector

The data for the governmental sector has been tested and adapted for its purpose and the periods of absence and the jobs have been matched. Thus, the first attempt has been made in creating the measures of absence. The documentation of the statistics is under progress and the time of publication is nearby.

The municipal sector

In the municipal sector, cooperation with FLD is initiated. The purpose of the cooperation is to ensure that the data is reported in a coordinated way. The contents of the data have been communicated to the two large HR-systems in this sector. A separate agreement has also been made with the biggest municipalities which have, to a greater or lesser degree, developed their own systems. The coverage of the population looks promising. The first test data from SD has been received.

The private sector

In the private sector, cooperation with DA has been extensive. The work has primarily been focused on three fronts. Firstly, meetings have been held with many different HR-system suppliers during the spring and early summer. The participants at the meetings have shown a positive attitude towards the development of the statistics on absence and towards developing or adjusting the existing systems for our requirements.

On another front, the sample has been determined in cooperation whit the Methods Unit of Statistics Denmark.

Finally, alternative media for registration and reporting of data are being developed. The purpose of this is to give the firms, and to some degree municipalities, an alternative way of complying with our demand for data if they do not want to use an existing systems.

Process toward the full range statistics

Given the current situation, the Statistics Denmark envisages a promising process towards the end of 2007 where we expect to have full range statistics on absence for the year 2006. However, we anticipate more questions to arise before we are home free.