

Visualizing Transitions between Multiple States – Illustrated by Analysis of Social Transfer Payments

Jacob Pedersen¹, Jakob Bue Bjorner^{1,2,3} and Karl Bang Christensen^{2*}

¹National Research Centre for the Working Environment, Copenhagen, Denmark

²Department of Public Health, University of Copenhagen, Denmark

³QualityMetric, Lincoln, RI, USA

Abstract

Background: Multi-state analyses are used increasingly in areas such as economic, medical, and social research. They provide a powerful analysis for situations where the research subjects move between several distinct states, but results are often complex. The purpose of the current paper is to present a simple descriptive analysis to visualize patterns in the transitions: the Top10 chart. Data on social transfer payments are used to illustrate the approach.

Methods: The Top10 chart summarizes the order in which the transitions between the states occur and the time spent in each state, and is constructed from individual level data. Persons with the same pattern of transitions between states are grouped together and average durations are calculated. We analyzed data from 4950 Danish employees aged 18-59 years who, during two years of follow-up, could at any time be in one of seven mutually exclusive states: work, unemployment, sick-listing, studying, parent leave, disability pension, and an absorbing state consisting of those who died, retired, or emigrated.

Results: The 10 most frequent transitional patterns described 84% of all women and 90% of all men in the sample. For women, the typical patterns involved working throughout the study (61.7%), patterns with sick-listing (12.0%), patterns with unemployment (5.3%), patterns with parent leave (3.6%), and studying (1.5%). For men, the typical patterns involved working throughout (68.8%), sick-listing (9.1%), unemployment (4.7%), parent leave (5.2%), and studying (0.9%).

Conclusion: The Top10 chart provides a simple descriptive visualization of complex transitional patterns.

Keywords: Longitudinal analysis; Multi-state models; Data visualization; Descriptive methods

Background

Today most studies of labor market outcomes like sickness absence use repeated observations over time, and the need to study dynamic pattern of the interplay of work environment, health, and consequences over time has been recognized [1]. One strategy for this is the use of multi-state models [2] that provide probabilities of transition among health states and explain differences among individuals during the course of a disease. Large longitudinal registers that contain information about social payments, medical diagnosis, payments, or employment status have gained increased importance in research areas like economics, medical, and labor market research. These registers can be local or national and are typically made available for research by central government or other major institutions with the authority to collect such data on an individual basis. Using registers like this, each individual can be in one of several distinct states. The purpose of the analysis will often be to predict the probability of transitioning from one state (e.g. work) to another state (e.g. sick-listing) and identify the variables that influence these transition probabilities.

The probability of observing a transition from one state to another often depends on previous transitions. For example, prior research has shown that earlier social transfer payments influence the risk of future payments [3,4]. This has increased the interest in understanding the order in which people receive different social transfer payments. One way of gaining more knowledge about the pathway to specific social payments is to identify typical patterns, e.g. patterns of repeated disease absence episodes. An individual pattern of social payments consists of combined information given by the type, order, and duration of each social payment period. These patterns are usually recorded over a

predefined study time. By understanding individual patterns of social payments the researchers gain valuable information that enables them to optimize the time of an intervention or to predict future social payments for specific groups of individuals.

Conventional methods for analyzing longitudinal data to gain information about patterns of social payments range from complex statistical methods to simple frequency tables. One of the many useable statistical methods is survival analysis, a method that is well suited for estimating probabilities of events like sickness absence or unemployment occurring over time [5,6]. Application of statistical models for survival analysis in the analysis of sickness absence is relatively new [5-7]. More complex methods such as multi-state models are useful for estimating transition probabilities between several different states, e.g. work, unemployment, and sickness absence, using larger models [7]. The use of multi-state models for the study of labor market outcomes as initiated by Lie et al. [7] have also been applied by other researchers [8-10] mainly to study return to work. Multi-state models have also recently been applied in studies of aging [11], cognitive impairment [12], and dementia [13].

***Corresponding author:** Karl Bang Christensen, Department of Public Health, University of Copenhagen, Denmark; E-mail: kach@sund.ku.dk

Received August 19, 2013; **Accepted** September 27, 2013; **Published** September 29, 2013

Citation: Pedersen J, Bjorner JB, Christensen KB (2013) Visualizing Transitions between Multiple States—Illustrated by Analysis of Social Transfer Payments. J Biomet Biostat 4: 175. doi:10.4172/2155-6180.1000175

Copyright: © 2013 Pedersen J., et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The disadvantages of multi-state models, whether based on survival analysis or generalized linear models, are that results can be complex to interpret and that the methods rely on assumptions about the data that limit the range of possible outcomes. This obstacle has been addressed by the above cited studies in different ways. Lie et al. considered three states and reported the average number of days spent in each along with 95% confidence intervals for each of two groups. Furthermore, they plotted results from the Cox proportional hazards model and non-parametric effects with 95% point-wise confidence limits [7]. Pedersen et al. [8] who considered a multi-state model with four states, reported the number of persons per transition for each of the nine studied transitions and the hazard ratios describing the effect of covariates on each of these transitions. Oyeflaten et al. [9] considered a multi-state model with eight different states and calculated nonparametric transition intensities using the Aalen-Johansen estimator. They plotted state probabilities and reported the probability of being in each of the eight states at follow-up, conditional on the baseline state. Carlsen et al. [10] considered four states and nine transitions. They reported the average number of weeks from one state of employment to the next and the hazard ratios describing the effect of covariates on each of these transitions. Song et al. [11] considered four states and reported parameter estimates from multilevel generalized logit models. Abner et al. [12] considered six states and reported the one-step transition matrix that counts the number of times each of the possible transitions occur and parameter estimates from a multi-state model. The employed model was based on a multi-state Markov chain with two competing absorbing states [14]. Kryscio et al. [13] considered four states and reported the frequency of each of seven transitions along with parameter estimates from a semi-Markov model.

As the previous paragraph illustrates, there are many ways of analyzing multi-state data, and even more ways of reporting the results. There appears to be a lack of simple graphical descriptions of this kind of data.

This article introduces the Top 10 Chart as a new tool for visualizing patterns of transitions between multiple states. This chart shows the most commonly appearing patterns in an intuitive and easily comparable way. The Top10 chart is an alternative to frequency tables and a supplement to survival analysis. It relies on very few assumptions about the data. The concept borrows ideas related to Gantt charts [15], but is novel within multi-state analysis. The Top 10 chart is designed to show the most commonly occurring patterns and can be used to compare patterns across sub populations. The chart is shown to provide easily communicable results, and is proposed as a companion to the less easily interpretable results from multi-state models. We demonstrate its potential by an analysis of data on social transfer payments.

Methods

Data

The data set used is a representative sample of 4,950 Danish

employees between 18 and 59 years of age. We merged the DWECS 2005 (Danish Working Environment Cohort Study) [16] and the RSS register (The Danish Register of Sickness absence compensation and Social transfer payments) [8]. Each respondent was followed in the RSS register for a period of two years from the time the DWECS 2005 questionnaire was answered. Only employees with no registration of a social benefit at the start time (May 2005) were included in the study.

The Danish social security system contains many types of social transfer payments. For the purpose of multi-state analysis we grouped them into seven states. At each time during the two years of follow-up each respondent was classified as being in one of the following seven mutually exclusive categories: working (self-supporting), unemployed, sick-listed, studying, parent leave, receiving disability pension, or 'not at risk'. The latter state is *absorbing* in the sense that subjects do not transition out of the state. Study subjects entered this state if they reached 60 years of age (thereby qualifying them for retirement), emigrated, or died. These subjects were subsequently assigned to this state for the remainder of the follow-up period.

These seven categories cover the majority of all social payments existing in the Danish social security system for persons between 18 and 59. The working category was defined by time periods wherein a person did not receive any social transfer payments. Only sick-listing involving payments of sickness absence benefits were registered.

The top10 chart

The calculation of the Top10 chart requires data about the different states that subject visit during the follow-up period. Two things are recorded for each person: (i) the order in which the transition between the states occur, (ii) the duration of the stay in each state. Persons that move between states in the same order are said to share the same pattern and are gathered together. The pattern is depicted in the Top10 chart using the average duration. This is illustrated in Figure 1 for three subjects who share the pattern 'Work', Sickness absence', 'Work'.

Figure 1 illustrates how three people with the same order of social payments result in a single line. Each shade of grey in Figure 1 represents a state and the length of each box of social payments represents the average duration of state occupancy. It should be noted that the pattern arising from grouping persons that have the same sequence of social transfer differs from the individual trajectories.

The Top10 chart is read from left to right. Each line represents a pattern, i.e., a series of social transfer payments that occur in a specific order. Each type of social transfer payment is represented by its own color pattern and the length of each event represents the expected duration in days. Time is read on the horizontal axis starting at zero (0)

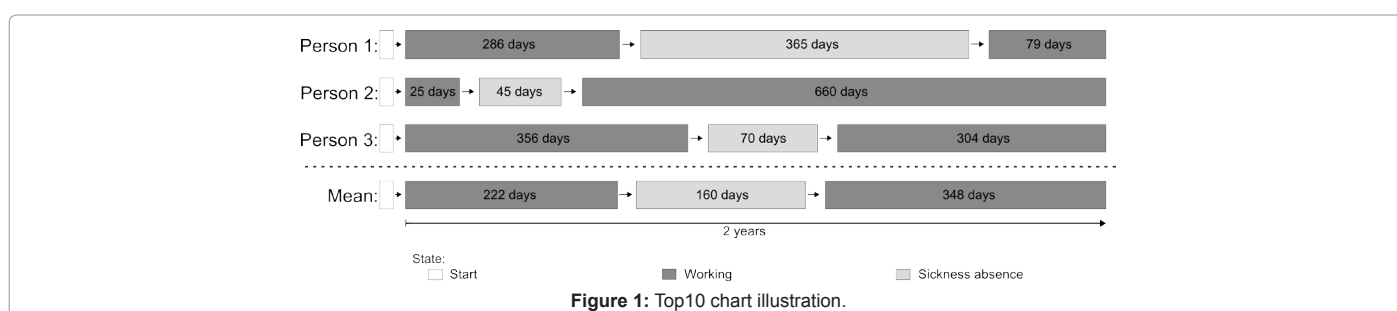


Figure 1: Top10 chart illustration.

indicating the date at which the monitoring started, and ends after two years of follow-up (730 days).

For each line shown in the Top 10 chart the proportion of people with that particular type of pattern is presented. This proportion is called the contribution margin (CM) of the pattern. The sum of the contribution margins of a number of selected patterns, e.g. those containing periods of sickness absence, is called a combined contribution margin (CCM). The CCM for a chart with 10 patterns is the percentage of the population that can be characterized by one of these 10 most frequently occurring patterns. The CCM is shown on the far right of the chart.

We suggest study of the combined contribution margin (CCM) over the 100 most frequent transitional patterns to evaluate whether the top 10 patterns represent the best compromise between brevity and detail. In the current analysis, charts were made for men and women separately to compare patterns between men and women. Stratification can be made over any discrete variable to provide a visual description of the impact of exogenous variables on transitional patterns.

SAS version 9.3 was used for all analyses; SAS code for making the Top10 chart is available at the website www.nrcwe.dk/riskstud (click on "Publications"). This website also contains an introduction to how the data should be processed and arranged to comply with the SAS code. It is possible to edit the SAS code to show any number of paths within the limitations of the GCHART procedure in SAS.

Results

Table 1 shows the representative sample of 4,950 Danish employees stratified by gender and age group. The population consists of a slight majority of women, and people in the age group 40-49 years. The table also shows the percentage and number of subjects who have been in each of the six categories during the follow-up period (recurrent visits are not shown): Sickness absence is the most frequently occurring category, and is more frequently occurring among women, in particular those aged 20-29 years. Similar results are seen for unemployment. Parent leave is the fourth most frequent category for both genders; it is more frequently seen among women aged 20-29, whereas men visiting this state are typically older (aged 30-39). Education is most frequent among the women aged 18-29 years. The 'not at risk category' occurs most frequently for men aged 50-59 years. Disability pension is the least frequent category and occurs more often among women than among men. The work category is not shown because all individuals are employed at baseline.

Figure 2 shows the top 10 most frequent patterns of social payments for a total population of 2,641 employed women, and Figure 3 shows

the corresponding ten most frequent patterns for a total population of 2,309 employed men.

Work

The top pattern in both of these figures is a two year period with no payment of any social benefit. This patterns accounts for 61.7% of the women and for 68.8% of the men.

Sick-listing

Three of the ten most frequently occurring patterns involve sick-listing (CCM is 12.0% for women and 9.1% for men). The second most frequently occurring pattern (line number 2) for both women (CM=8.3%) and men (CM=6.3%) is a pattern of work followed by sick-listing followed by return to work. The subjects worked for an average of 354 days followed by sick-listing with an average duration of approximately 54 days (a little longer for the women), and then returned to work for the remainder of the follow-up period. Pattern number four for the women (CM=2.0%) is comparable with the sixth most frequently occurring pattern for the men (CM=1.6%). This pattern shows two periods of sick-listing for subjects who are otherwise working; the average length of the sick-listing periods being shorter for the men. Finally, the seventh most frequently occurring pattern for women (CM=1.7%) and eighth most frequently occurring pattern for men (CM=1.2%) consists of a single period of sick-listing that extends beyond end of follow-up. The average time of working before being sick-listed started was 583 days.

Unemployment

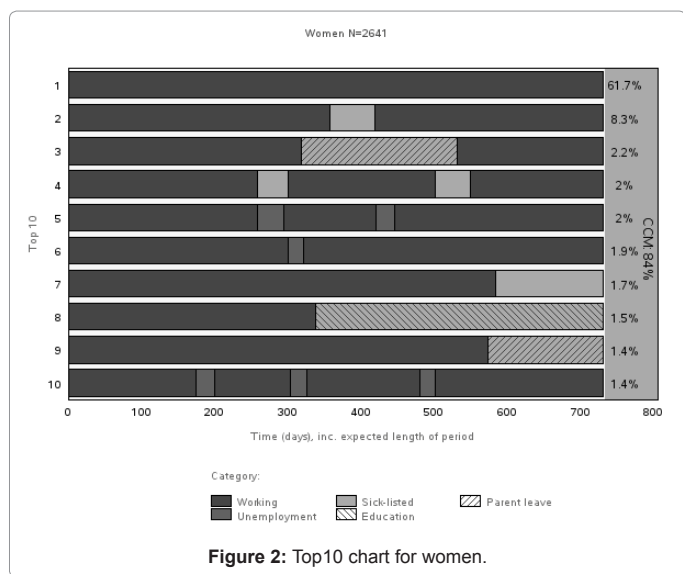
Three of the patterns for women (CCM=5.3%) and two of the patterns for men (CCM=4.7%) involved unemployment. For women, the fifth (CM=2.0%), sixth (CM=1.9%) and tenth (CM=1.4%) most frequent patterns involved two, one, and three periods of unemployment, respectively. The expected length of each unemployment episode was 20-36 days. For men, the fourth (CM=3.1%) and fifth (CM=1.6%) most frequent patterns involved one and two periods of unemployment, respectively. The expected length of each unemployment episode was 31-41 days—thus slightly longer than the expected length for women.

Parent leave

For both women and men, two of the ten patterns involved parent leave (CCM=3.6% for women and CCM 5.2% for men). For women, the two patterns involved a single episode of parent leave followed either by return to working (on average after 212 days) or end of follow-up. For men, the third most frequent pattern was a parent leave of, on average,

Gender	Age (years)	N	Sickness absence	Unemployment	Disability pension	'not at risk'	Education	Parent leave
Women	18-19	17	17.6% (3)	11.8% (2)	0.0% (.)	0.0% (.)	58.8% (10)	5.9% (1)
	20-29	268	23.9% (64)	26.9% (72)	0.0% (.)	0.0% (.)	17.2% (46)	26.9% (72)
	30-39	661	21.0% (139)	19.4% (128)	0.3% (2)	0.0% (.)	3.8% (25)	16.3% (108)
	40-49	869	19.3% (168)	12.0% (104)	0.7% (6)	0.1% (1)	2.5% (22)	1.5% (13)
	50-59	826	21.5% (178)	7.3% (60)	0.5% (4)	3.3% (27)	1.1% (9)	0.0% (.)
sum:		2641	20.9% (552)	13.9% (366)	0.5% (12)	1.1% (28)	4.2% (112)	7.3% (194)
Men	18-19	12	16.7% (2)	8.3% (1)	0.0% (.)	0.0% (.)	41.7% (5)	0.0% (.)
	20-29	231	15.2% (35)	14.3% (33)	0.4% (1)	0.4% (1)	9.5% (22)	12.1% (28)
	30-39	623	14.0% (87)	13.2% (82)	0.0% (.)	0.5% (3)	0.8% (5)	18.6% (116)
	40-49	735	13.5% (99)	7.9% (58)	0.1% (1)	0.3% (2)	0.8% (6)	2.7% (20)
	50-59	708	17.5% (124)	10.9% (77)	0.6% (4)	4.5% (32)	0.7% (5)	0.1% (1)
Total Sample:		2309	15.0% (347)	10.9% (251)	0.3% (6)	1.6% (38)	1.9% (43)	7.1% (165)

Table 1: Sample by gender and age at baseline, pct. subject ever in one of the six states: sickness absence, unemployment, disability pension, 'not at risk', education, and parent leave.



25 days followed by return to work (CM=3.9%). The seventh most frequent pattern for men is two periods of parent leave (CM=1.3%). On average, the first period is approximately 25 days, followed by a period of working for approximately 200 days, followed by a second parent leave for approximately 51 days.

Other patterns

Approximately 1.5% of women and 0.9% of men started receiving educational grants. Finally, 1% of men entered the absorbing state. While entering an absorbing state was approximately as frequent among women (0.8%), this pattern was not among the top 10, due to a larger diversity of patterns among women (the 12th most frequently occurring pattern contained this state).

Figure 4 shows the combined/cumulated contribution margin (CCM) calculated up to a Top 100 chart for the women. From this Figure it can be seen that the CCM would only increase a small fraction by expanding the Top 10 chart to a Top 11 chart or higher (the horizontal solid reference line shows the CCM of the Top10).

Discussion

This article focuses on longitudinal data where each subject can shift between multiple non-overlapping states. We describe a visual presentation of the most typical patterns of transitions between states. By presenting the ten most typical patterns in a single chart, the researcher gains a tool for comparing and understanding the typical patterns in the investigated population or cohort. The Top10 chart is suitable for direct reporting, presentation of patterns, and for direct comparison of different subgroups.

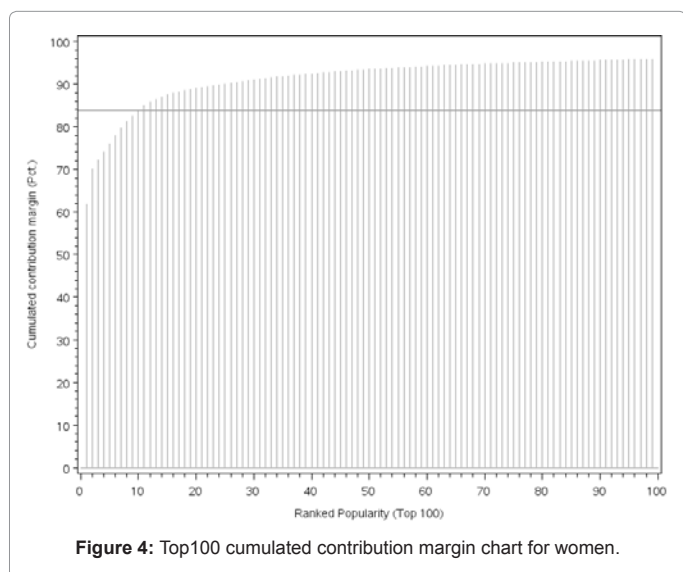
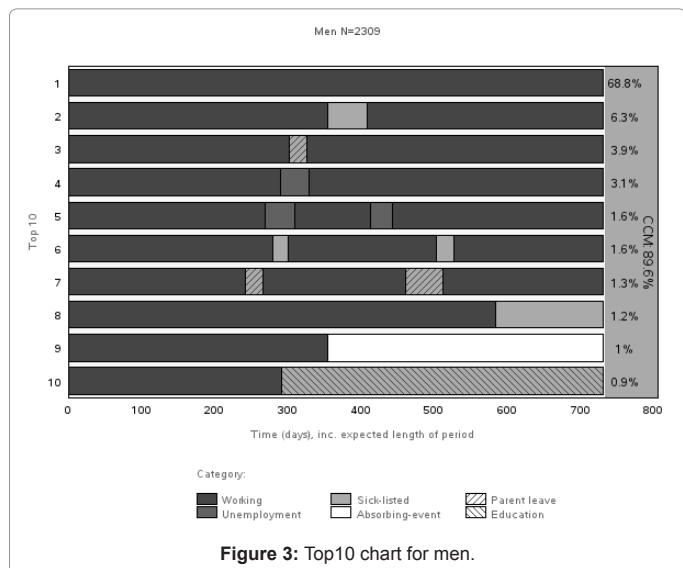
The cohort used in the present article has previously been analyzed – in particular with regards to sick-listing and early retirement. However, the present simple analysis highlights a number of observations that have not received much attention before.

1. A rather large percentage of men go on parent leave. This pattern may be specific to countries like Denmark with a well-developed reimbursement system for men's parent leave. However, the percentage of men taking parent leave, and in particular the percentage taking parent leave twice, was higher than expected.

2. Each of the Top10 patterns involved at most, only two states. Since previous sick-listing is a predictor of unemployment and early retirement, we had anticipated patterns where sick-listing was followed by unemployment or early retirement. Although such patterns occurred, they were rare, and not among the ten most frequently occurring.

3. Among the Top10, patterns with frequent unemployment episodes seemed more common among women than among men. Since many men work in sectors (e.g. construction) with high seasonal unemployment, we had expected that more men would experience temporary unemployment than women.

Since the Top10 chart does not take possible confounding factors into account and does not provide measures of robustness (in the form of confidence intervals, p-values etc.) not all of these results may be confirmed by further analysis. However, these examples illustrate the potential of this model for pointing out the unexpected. The Top10 chart show the unfiltered patterns and this makes it easy to clarify the most relevant types of events, find the most common transitions, and the most likely competing risks. All of this is important information when making a multi-state model. Making only few assumptions about the data, the chart shows the general picture with multiple outcomes.



The present study considered a population with no loss to follow-up. In future studies the methodology could be extended to accommodate differing lengths of follow-up by including a censoring period due to time truncation, however this was beyond the scope of the present study.

The Top10 chart has precedents in tables enumerating the transition path frequencies, but expands on these by providing extra information about the mean duration of the stay in each state. A further expansion of the Top10 chart would be to include information about the variability. As illustrated in Figure 1, a high variability of state durations may be summarized in the mean duration plotted. Future work is needed to evaluate whether measures of variability can be included in the graphs without disturbing the clear visual representation.

While the Top10 chart conveys a very intuitive picture, a few precautions should be taken when interpreting the chart. Each relevant state is represented by the mean length in days. Using this mean length as the expected time period to the next event only is not recommended. Because the mean length is always conditional on the future, the mean length can only be used if the future is exactly the order of events that makes the pattern. Also, the combined contribution margin of the chart will depend on the duration of the study and on the definition of states. Two choices have been made in our data analysis before making the charts: the length of the follow-up period, and the number of studied states. If the follow-up period is too wide, the chart would become less informative. A similar limitation would result if a large number of different states were to be considered. When some states are both common and highly transient, as is the case in some applications [12], our methodology is not useful. In situations like this the number of distinct patterns increases and we would not be able to capture a large percentage of the study population with patterns consisting of only five or six states. This would show as low CM and CCM values, which indicates that the ordering of the pattern is less robust, as no particular pattern is more frequent than the other. For this task, the Top100 cumulated contribution margin chart is helpful, as it shows the number of paths needed to increase the CCM value. The task at hand is to find a balance between the information monitored, and the simplification needed for visualizing general patterns. The number 10 was chosen to represent a reasonable number of patterns that can be grasped by the researcher while still providing a description of a large percentage of the population. In working with register data on social transfer payments, we found that the 10 most frequent patterns capture more than 80% of the population.

Conclusions

We conclude that the Top10 chart provides a simple descriptive visualization of complex transitional patterns. As illustrated in the analysis of social transfer payments, this simple visualization may provide perspective to patterns overlooked in previous analyses.

References

1. Burdorf A (2012) The need for novel strategies to analyze the dynamic pattern of worker's health over time and the consequences for sustained employability. *Scand J Work Environ Health* 38: 485-488.
2. Andersen PK, Keiding N (2002) Multi-state models for event history analysis. *Stat Methods Med Res* 11: 91-115.

3. Duijts SF, Kant IJ, Landeweerd JA, Swaen GM (2006) Prediction of sickness absence: development of a screening instrument. *Occup Environ Med* 63: 564-569.
4. Reis RJ, Utzet M, La Rocca PF, Nedel FB, Martin M, et al. (2011) Previous sick leaves as predictor of subsequent ones. *Int Arch Occup Environ Health* 84: 491-499.
5. Christensen KB, Andersen PK, Smith-Hansen L, Nielsen ML, Kristensen TS (2007) Analyzing sickness absence with statistical models for survival data. *Scand J Work Environ Health* 33: 233-239.
6. Gjesdal S, Ringdal PR, Haug K, Maeland JG (2008) Long-term sickness absence and disability pension with psychiatric diagnoses: a population-based cohort study. *Nord J Psychiatry* 62: 294-301.
7. Lie SA, Eriksen HR, Ursin H, Hagen EM (2008) A multi-state model for sick-leave data applied to a randomized control trial study of low back pain. *Scand J Public Health* 36: 279-283.
8. Pedersen J, Bjorner JB, Burr H, Christensen KB (2012) Transitions between sickness absence, work, unemployment, and disability in Denmark 2004-2008. *Scand J Work Environ Health* 38: 516-526.
9. Oyeflaten I, Lie SA, Ihlebæk CM, Eriksen HR (2012) Multiple transitions in sick leave, disability benefits, and return to work. A 4-year follow-up of patients participating in a work-related rehabilitation program. *BMC Public Health* 12: 748.
10. Carlsen K, Harling H, Pedersen J, Christensen KB, Osler M (2013) The transition between work, sickness absence and pension in a cohort of Danish colorectal cancer survivors. *BMJ Open* 3.
11. Song C, Kuo L, Derby CA, Lipton RB, Hall CB (2011) Multi-stage transitional models with random effects and their application to the Einstein aging study. *Biom J* 53: 938-955.
12. Abner EL, Kryscio RJ, Cooper GE, Fardo DW, Jicha GA, et al. (2012) Mild cognitive impairment: statistical models of transition using longitudinal clinical data. *Int J Alzheimers Dis* 2012: 291920.
13. Kryscio RJ, Abner EL, Lin Y, Cooper GE, Fardo DW, et al. (2013) Adjusting for mortality when identifying risk factors for transitions to mild cognitive impairment and dementia. *J Alzheimers Dis* 35: 823-832.
14. Salazar JC, Schmitt FA, Yu L, Mendiondo MM, Kryscio RJ (2007) Shared random effects analysis of multi-state Markov models: application to a longitudinal study of transitions to dementia. *Stat Med* 26: 568-580.
15. Gantt HL (1910) *Work, Wages and Profit*, published by The Engineering Magazine, New York; republished as *Work, Wages and Profits*, Hive Publishing Company, Easton, Pennsylvania.
16. Feveile H, Olsen O, Burr H, Bach E (2007) Danish work environment cohort study 2005: from idea to sampling design. *Statistics in Transition-new series* 8: 441-458.

Citation: Pedersen J, Bjorner JB, Christensen KB (2013) Visualizing Transitions between Multiple States—Illustrated by Analysis of Social Transfer Payments. *J Biomet Biostat* 4: 175. doi:10.4172/2155-6180.1000175