Institute for Employment Research

The Research Institute of the
Federal Employment Agency

IAB-Discussion Paper 29/2015

Articles on labour market issues

Misreporting to looping questions in surveys

Recall, motivation and burden

Stephanie Eckman
Frauke Kreuter

ISSN 2195-2663

Misreporting to looping questions in surveys

Recall, motivation and burden

Stephanie Eckman
Institute for Employment Research (IAB)
Frauke Kreuter
University of Maryland & University of Mannheim & IAB

Contents

Abstract

Looping questions are used to collect data about several similar events, such as employment spells, retirement accounts, or marriages. The questions collect information about the number of events experienced as well as details about each one. The questions require respondents to think hard to recall each event and are often lengthy and repetitive. Looping questions can be asked in two formats, and which format a survey uses may affect the quality of the data collected. We develop theory-based hypotheses about the effects that the choice of format has on measurement error in looping questions and test the hypotheses using experimental data from a recent web survey with a link to administrative records. Results show that one format collects more accurate event reports, but the other format provides higher quality data to the follow up questions. We conclude with guidance for those who write survey questions as well as those who rely on survey data for substantive analyses.

Zusammenfassung

Fragebögen enthalten oftmals sogenannte „Schleifen“ um Daten über ähnliche Ereignisse zu sammeln. Ein Beispiel hierfür sind Frageschleifen über verschiedene Beschäftigungsperioden, Ruhestandskonten, Wohnorte oder Ehen. Mit Hilfe dieser Schleifen werden Informationen über eine Anzahl von erlebten Ereignissen gesammelt, sowie einzelne Details zu jedem Ereignis. In der Praxis kommen zwei verschiedene Formate von Schleifen zur Anwendung, welche unterschiedliche Auswirkungen auf die Qualität der gesammelten Daten haben. Ausgehend von Theorien der Kognitionspsychologie entwickeln und testen wir in diesem Beitrag Hypothesen über die Messfehler, die die verschiedenen Formate mit sich bringen können. Daraus ableitend geben wir Handlungsempfehlungen für alle die Survey Fragen erstellen, oder in ihren Analysen auf Befragungsdaten angewiesen sind.

JEL classification: B40

Keywords: looping questions, measurement error, misreporting

Acknowledgements: Funding for this project was provided by the Institute for Employment Research (IAB) and by Ludwig Maximilian University. We are grateful for assistance given by Barbara Felderer, Annette Jäckle, Antje Kirchner, Yuliya Kosyakova, and Ulrich Thomsen. Ruben Bach, Bob Belli, Georg-Christoph Haas, Arie Kapteyn, MaryBeth Ofstedal, Norbert Schwarz, and Roger Tourangeau commented on earlier versions of this work. Daniel Fehrle, Hannah Mautner, and Svenja Wippich provided research assistance.

1 Introduction

Many surveys use lengthy and repetitive sections of looping questions to collect data about several similar events, such as doctor visits or periods of unemployment. The questions collect information both about the number of such events experienced and details about each one. Looping questions can be asked in two formats. One format we call how many: “In how many different cities have you lived since you were 14?”. The response to this question then triggers loops of follow up questions about each city. The second format we call go-again: “Please tell me about the city where you lived when you were 14 ... How many people lived in that city or town? Did you live in a stand-alone home, an attached home, or an apartment? ... And what city did you live in after that?” This format collects the same information as the how many format, but in a different way. Both formats are widely used in surveys today, yet their effects on the quality of the collected data are not known. Because the formats' implications for data quality have not been thoroughly explored, surveys have no good basis for making a choice between the two. This paper develops and tests hypotheses about data quality in the two formats and provides evidence-based guidance to survey researchers who wish to ask looping questions or analyze data collected via such questions.

Table 1 gives examples of looping questions in seven surveys in the US and Europe. This list is not exhaustive of all looping questions in these surveys, but the questions collected here do span a range of topics: personal history, household composition, household finances, and health. In our search through questionnaires of major surveys, we find that the how many format is more common. However, we do not see any pattern to when a survey uses the how many format and when it uses the go-again format.

Two strains of research about the survey response process provide insight into how respondents answer looping questions in the two formats. The first is the literature on behavioral frequency questions, which ask respondents to report the number of times an event has occurred, for example: hours of television watched last week or the number of visits to the dentist in the last year. This literature focuses on the burden of the recall task and strategies respondents use to decrease this burden, such as satisficing. The other relevant body of research is that of motivated misreporting, which argues that respondents manipulate their answers to reduce the length of the interview. This theory is concerned not with the burden of the recall associated with any one question, but with the burden due to the length of the questionnaire as a whole. We discuss each of these viewpoints in more detail below and how each relates to the looping question response process.

Setting aside the follow up questions for a moment, the how many question is similar to a behavioral frequency question. This type of question has been the subject of much research (see in particular Blair/Burton, 1987; Burton/Blair, 1991). In answering a frequency question, the respondent generally follows one of four response strategies: she knows the true answer immediately (such as would often be the case with “How many biological children do you have?”); she remembers each of the relevant events individually and counts them (such as trips to the emergency room in the last three years); she applies a rate to the recall period (two one-week vacations per year over five years); she gives an

Table 1: Examples of Looping Questions from Current Surveys

Survey	Format	Question text	# loops possible	# follow up questions
PSID[a]	How many	During <year>, how many full-time or part-time jobs did you/he/she have (not counting work around the house)?	3	8
	Go-again	What is the street address and move-in date of your/head of household's current residence? Have/Has you/he/she lived anywhere else since January?	6	4
HRS[b]	How many	Altogether, how many times have you been married (including your current marriage)?	3	up to 4
PASS[c]	How many	How many friends or family members do you have a close relationship with, not counting members of your HH?	3	5
	Go-again	Was there a cut to the amount of income support you/someone in your household received at any point in time between <start>and <end>? Was there a further cut to the amount of income support during that time?	5	4
NSFG[d]	Go-again	I would like to get some additional information about the people in this household. Is there anyone else who usually lives here?	No limit	5
	How many	How many different females have you ever had intercourse with? This includes any female you had intercourse with, even if it was only once or if you did not know her well.	3	7
ESS[e]	How many	Including yourself, how many people – including children – live here regularly as members of this household?	No limit	3
SHARE[f]	How many	How many children do you have that are still alive? Please count all natural children, fostered, adopted and stepchildren.	20	4
NLSY97[g]	How many	Now I would like to ask you about any college or university experience you've had. How many different colleges or universities have you ever attended?	3	12
	How many	How many of your pregnancies were not live births, that is, they ended in a stillbirth, a miscarriage or an abortion?	No limit	2

[a] Panel Survey on Income Dynamics (Panel Survey on Income Dynamics, 2013)
[b] Health and Retirement Survey (Health and Retirement Survey, 2011)
[c] Panel für Arbeitsmarkt- und Sozialversicherung (Berg et al., 2014)
[d] National Survey of Family Growth (Lepkowski et al., 2010)
[e] European Social Survey (Central Co-ordinating Team, 2010)
[f] Survey of Health, Ageing and Retirement in Europe (Börsch-Supan/Jürges, 2005)
[g] National Longitudinal Survey of Youth, 1997 Cohort (Moore et al., 2000)

answer based on a general impression ("I don't go to the movies that often"). Retrieval of the true answer is rare and respondents usually use one of the other strategies, yet each of these strategies is prone to errors of overreporting and underreporting. When forming responses to behavior frequency questions, respondents often use contextual cues from the question wording, its placement in the questionnaire, and the response options (Tourangeau/Bradburn, 2010). Unfortunately, the how many question does not provide much help to the respondent as she formulates her response, and thus the answer may be of low quality. The go-again questions, on the other hand, do provide the respondent with some assistance with recall, by asking about related events one at a time and in a logical order. Additionally, the questions in the go-again format lead the respondent to make several attempts to remember relevant events, and repeated attempts have been shown to improve recall (Schwarz/Oysermann, 2001).

Consider the question from the National Survey of Family Growth in Table 1 about female sexual partners. As a how many question, it implicitly asks the respondent to use the recall-and-count method (or to retrieve an already-known true value). The question itself, though, provides no assistance with the recall task: the burden is entirely on the respondent to remember each partner. Because no help is provided, the respondent may use a suboptimal estimation strategy to arrive at the answer. The go-again version of this question would first ask about one's first female partner, then the second, and so on, which could elicit better reporting, by breaking the task into several smaller steps and structuring the memories in a logical order.

Based on the research into the responses process to behavioral frequency questions, we hypothesize that the go-again format leads to more accurate event reports. We call this the *better reporting hypothesis*.

However, we have so far not considered the role played by the follow up questions about each event. These questions add to the length and burden of the questionnaire, and thus the other relevant body of research is *motivated misreporting*. Previous studies have found that respondents misreport to filter questions and screener questions in an effort to reduce the burden of the interview (Tourangeau/Kreuter/Eckman, 2015) and a similar effect may occur with looping questions. The go-again format makes it clear that reporting an additional event will lead to another loop of follow up questions. Respondents who want to decrease the length of the questionnaire could underreport the number of events. The how many format, on the other hand, does not reveal the follow up questions until the respondent has reported the number of events, hiding the relationship between the number of events reported and the length of the interview, and thus removing the incentive to underreport.

This reasoning leads us to an alternative hypothesis, that the go-again format encourages underreporting and the how many format collects more event reports and more accurate event reports. We call this the *motivated misreporting hypothesis*.

The above two hypotheses point to two different aspects of the burdens that surveys place on respondents. One aspect is simply the time demanded of respondents: longer surveys are more burdensome (Bradburn, 1979; Sharp/Frankel, 1983). The motivated misreporting hypothesis is concerned with this type of burden and the ways that data quality can suffer when respondents try to shorten the interview. Another aspect of burden is the effort required to answer individual questions (Bradburn, 1979; Sharp/Frankel, 1983). When done well, responding involves comprehending what the question asks, retrieving relevant information, making judgments and estimates based on the retrieved data, and finally reporting the answer. This process requires careful thought and attention and places a burden on respondents. Some respondents may perform these steps half-heartedly or not at all, a response strategy called satisficing (Krosnick, 1991). The how many format of the looping questions is cognitively more demanding than the go-again format, and thus may be more susceptible to satisficing, such as via the rate- and impression-based estimation strategies discussed above. The go-again format, on the other hand, decomposes the frequency question into several questions, simplifying the recall task and reducing the recall burden (Cannell et al., 1989; Krosnick/Presser, 2010).

Thus, the two looping question formats each address a different aspect of respondent burden. The go-again format reduces the burden of the recall effort, but in doing so, opens the door to motivated misreporting to reduce the length of the questionnaire. The how many format hides the looping structure of the questions from the respondent, eliminating the incentive to reduce the length by misreporting, but at the same time increases the recall burden. The quality of the data collected by each format is likely related to which type of burden respondents find more troubling.

A recent conference presentation by Carley-Baxter/Peytchev/Black (2010) is the only study we are aware of that experimentally compared the how many and go-again formats of the looping questions (though the authors use the terms grouped and sequential). The topic of that telephone survey was sexual violence, and the looping questions asked about experiences as a victim of psychological aggression, physical violence, stalking and other such behaviors. The how many format collected more reports than the go-again format from male respondents (4.6 vs 3.5), but fewer from female respondents (5.9 vs 6.3) (test statistics and significance not reported in slides). The study also found more missing data in the go-again format but more breakoffs in the how many format.

Armed with the two competing hypotheses developed above and two aspects of respondent burden, we consider several measures of data quality in looping questions. We compare the performance of the two formats and conclude with recommendations for survey researchers who wish to use looping questions and for analysts who work with such data.

2 Data

To address the research questions given above and explore the quality of the data collected via looping questions, we conducted a web survey that experimentally varied the format of the loops. The survey also allowed us to merge responses to administrative records, and thus we can determine not only which format collects more events but also which is more accurate.

Sample

The sample of 11,836 named adults was selected from German federal databases (IAB Institut für Arbeitsmarkt- und Berufsforschung, 2011) in three strata representing diverse employment and unemployment histories. The first stratum contained persons who received income support in 2010 and held a social security contributing job in the last 10 years. The second consisted of persons who received unemployment insurance in the last 10 years, held a social security contributing job in the last 10 years, and never received income support. The third stratum consisted of persons who received neither income support nor unemployment insurance and held social security contributing jobs with two or more different employers in the last 10 years. Within each stratum, the sample was equal probability.

Each selected person was mailed a letter which asked them to go online to complete the web survey. Data collection was from February to April, 2012. The completed sample size was 1,068 cases with a response rate of 9.01% (AAPOR RR1). The median completion time, among those who completed the entire questionnaire, was 20.4 minutes. There were no differences in the mean or median time in the survey or in the looping question section between the two formats. Paradata indicate that twenty respondents completed the survey on a mobile device. Due to the small size, we do not analyze these respondents separately, but we do include them in our analyses.

An additional 143 cases started the survey and answered some of the looping questions, but did not finish the entire interview. Those that finished both looping sections ($n = 34$) are included in all of our analyses, for a total of 1,102 cases. Those cases that broke off during the looping sections ($n = 109$) are analyzed only when addressing the fourth research question on breakoffs. There were additional cases that broke off before answering any of the looping questions, but we do not consider these cases here at all.

Questionnaire

The survey contained two sections of looping questions. One asked about employers and the other about places the respondent had lived. These questions are similar to those in the PSID (see Table 1), though we asked about lifetime employer and residences rather than those in the last twelve months. We chose these topics due to the expected availability of administrative data with which to evaluate response accuracy (see below).

Each respondent was randomly allocated to receive both looping sections in the how many format or both in the go-again format. For each mentioned employer or location, the respondent was asked four follow up questions. The order of the two sections was randomized, with half the respondents asked first about employers and half asked first about locations. The order of the loops within each section was fixed: Respondents were asked to report about their employers/locations in chronological order, starting with the earliest. The full text of all the looping and follow up questions, translated from German by the authors, is given in the Appendix.

To keep the length of the survey reasonable, and reduce the risk of breakoffs that would harm later experiments, we limited respondents to seven loops through the follow up questions in the first looping section they saw, and five in the second. Thus, for respondents in the how many format, we have the full number of employers/locations, but in the go-again format we have at most seven (or five, in the second loop). In order to compare the responses between the two formats, we re-coded the number of events reported in the how many format to match what was possible in the go-again format.

After respondents in the how many condition had reported about all the employers or locations that they had indicated, they were asked if they had forgotten anything and wanted to report another.[1] If they said yes, they were sent through the follow up questions again, and then again asked if they had yet another to report.

The web survey we analyze here also contained other experimental manipulations: a consent experiment and an incentive experiment. Both of these experiments were fully crossed with the assignment to looping question format and the order of the looping questions. The consent experiment varied the placement and wording of the question about consent to link survey responses to administrative data (see Sakshaug/Kreuter, 2014: for details on the consent questions): we do not expect these manipulations to affect our results. The incentive experiment occurred after the looping questions and thus has no effect on our results (Felderer/Kreuter/Winter, 2013).

Administrative Data

Because the sample was selected from administrative databases, we are able, with respondent consent, to link the respondents' answers with administrative records. Overall, 61.6% of the respondents ($n = 679$) consented to the link,[2] and the consent rate did not differ by the two looping question formats. The linked administrative records come from the database of social security contributions made each year by employers in Germany (IAB Institut für Arbeitsmarkt- und Berufsforschung, 2013). The database contains reports of all spells of social security contributing employment in Germany since 1975. Each spell contains an identifier for the firm making the contribution and information about the employee as well. From these records we can count the number of different employers each linked respondent has had.

The records captured in the administrative data do not entirely match the common understanding of employment, and thus some mismatch between the survey responses and the administrative data is expected. Positions such as civil servant, police officer, professor, and self-employed, which are not covered by social security, are not captured in the administrative data (Jacobebbinghaus/Seth, 2007). Respondents, however, may include these positions when they report the employers they had. However, the mismatch between reports and the administrative records should be similar across the two looping question

[1] This question was asked only if the number of events initially reported was less than the maximum number of trips through the follow-up questions (seven in the first set, five in the second).

[2] The consent rate given here differs slightly from that reported by Sakshaug/Kreuter (2014) for the same survey, due to different decisions about which partially completed cases to include in the analysis data set.

formats, due to the random assignment of respondents to formats, and thus it should not bias our findings.

As a check on the randomization, we tested the difference between the number of employers in the administrative data for those assigned to the how many format and those assigned to the go-again format, and it was not significant ($F(1, 668) = 2.37, p = 0.12$). The number of employers reported in the administrative data was also re-coded to match the number of possible reports in the go-again condition.

In designing the survey, we had also planned to verify the number of residences, but the concept used in the administrative data (the German *Gemeinde*, an administrative unit that can refer to a city or a collection of neighboring villages) was too difficult to ask respondents about. Furthermore, the administrative records only go back to 1995.

3 Methods

We designed this survey specifically to run experiments related to looping questions and test our two hypotheses about data quality in looping questions. We explore four aspects of data quality: the number of events reported in the two formats, accuracy in the number of the events reported, missing data in the follow up items, and breakoffs.

To explore how the format effect interacts with other question and respondent characteristics to influence the number of events reported, we run two regression models. The data set for the two models is at the section level and thus each respondent appears twice. The models differ only in terms of their case base: the first model includes all respondents and loops, and the second includes only those respondents who consented to the linkage of administrative data. The dependent variable in the models is the number of events reported by the respondent in that loop, re-coded as described above. Because the dependent variable is a count variable, we used Poisson regression models. The mean number of events reported is 3.02 with a variance of 3.11. The nearly equal mean and variance support our use of the Poisson model. The independent variables are the format (how many versus go-again); the two looping sections (employers verusus locations); and an indicator for the first or second loop. Additionally, the first model includes an indicator of whether the respondent consented to linkage or not, and the second model includes the respondent gender, merged in from administrative data.

To understand which looping question format leads to more accurate reports of the number of reported events, we use only the response to the number of employers and only those cases that consented to linkage of their responses to administrative records. For each respondent, we flag whether she overreported, underreported, or responded correctly, relative to the administration data. We then compare the extent of each type of reporting by looping question format, using F tests.

Because the administrative data cannot be used to verify the quality of the reports to the follow up questions in either section, we operationalize data quality in the follow-up items by the fraction of don’t knows or refusals. For each loop through the follow up items, we calculate the number of don't know and refusal responses, out of four, and compare the results between the two formats.

We then bring the 109 cases that broke off during the looping sections back into the analysis data set and look at the factors that influenced the breakoff decision, using Pearson χ^2 statistics to test for significance.

All analyses are unweighted, as our goal is not to make inference to the population in the three strata used in selection, but to compare the performance of the two formats of looping questions for the collection of biographical data.

4 Results

With this survey and its link to administrative data, we can address four elements of data quality in looping questions: the number of events reported; the accuracy of the event reports; missing data in the follow up items and survey breakoffs. The literature review above led us two competing hypotheses: the motivated misreporting hypothesis and the better recall hypothesis. The first states that the how many format collects reports of more events and is more accurate, because it hides the repetitive structure of the loops from the respondents and minimizes underreporting. The second holds that the go-gain format leads to better recall, because it reduces the burden of the recall task and takes respondents through their life histories in a logical way. In comparing the quality of the data collected in the two formats, we also test these hypotheses.

Number of Events Reported by Format

Figure 1 shows the distribution of the number of events reported, by question and by format. The graphs in the left column refer to the employers loop and those in the right column refer to the locations loop. The first row shows the number of events reported in the how many condition, before any re-coding. Those in the second row show the number of events in the how many format after re-coding. (Re-coding was to five or seven employers and locations, depending on which loop a respondent was asked first: see Section 2 for more information.) The last row of graphs are for the go-again format, where no re-coding was done (because this format was constrained by the web instrument to seven or five loops).

Comparing the first and last rows of graphs in each column, there is clearly a different pattern in the reports between the two formats. Respondents reported more employers and locations in the how many format than in the go-again format. Even when we re-code the how many reports to match the maximum number of possible events in the go-again questions, as in the second row of graphs, the reports in the how many format are still higher.

We also see that some respondents reported zero employers, which should not happen given the way we selected the sample. This occurred in both formats, though more often in the go-again format. The effect is likely due to underreporting to the filter question

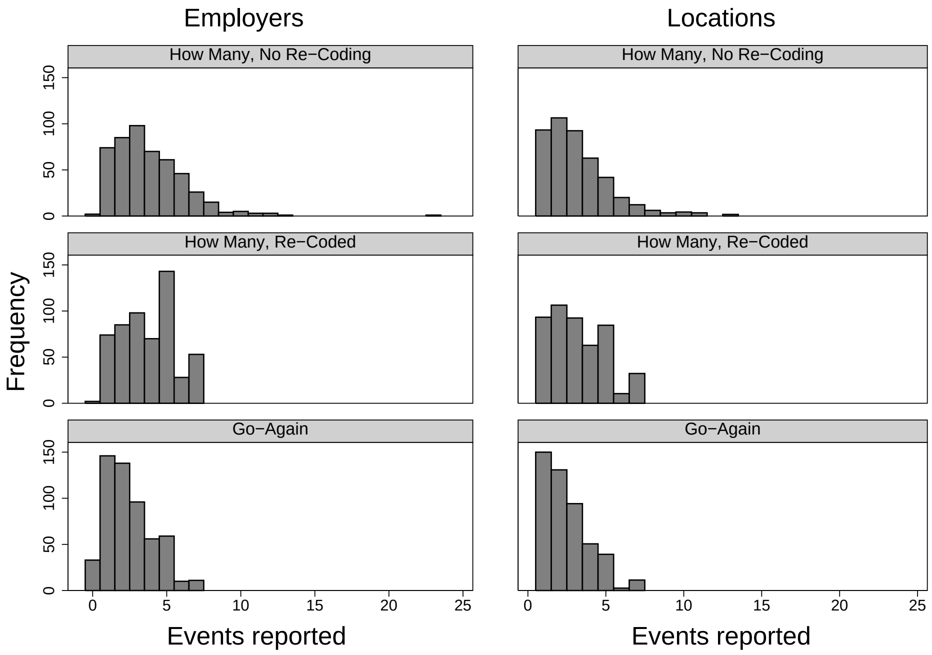

Figure 1: Number of Events Reported, by Format

that began the employer loop in the go-again format: “Have you ever been employed?” (There was no such filter in the locations loop, because we presumed everyone has lived somewhere.)

The how many loops in our questionnaire included a question at the end asking whether respondents had any additional events that they had failed to include when responding to the first question. (See the Appendix for details on how this question was worded.) 17% of respondents in the how many condition did in fact report one or more extra employers or locations, for a total of 57 additional employers and 68 additional living locations. One respondent used this technique to report six additional employers and another to report six additional locations – essentially turning the how many loop into a go-again loop. These additional events are not included in Figure 1, but including them would only strengthen our conclusion that the how many format collects more events.

We do not often see this question about additional events asked in surveys which use the how many format, but clearly it can capture some events that would otherwise go unreported. The reporting of additional events speaks against the motivated misreporting hypothesis and for the aided recall hypothesis: It seems that answering the follow up questions triggers recall of additional relevant events. If respondents were interested only in decreasing the length of the survey, they would say “no” to the question about additional events. Because most surveys do not include this question, we exclude these additional events from all further analyses but return to this issue in the discussion.

Two Poisson regression models, described in Section 3, help us understand the factors influencing the number of events reported. The dependent variable in the models is the number of events reported in a given loop. Table 2 reports estimated coefficients and marginal effects for both models.

Table 2: Factors Affecting Number of Events Reported

	Model 1		Model 2	
Variable	$\widehat{\beta}$ (Std. Error)	Marg. Effect (Std. Error)	$\widehat{\beta}$ (Std. Error)	Marg. Effect (Std. Error)
Format *reference category: How Many*				
Go-Again	-0.478*	-0.996*	-0.302*	-0.831*
	(0.0506)	(0.0766)	(0.0541)	(0.0975)
Section *reference category: Employers*				
Locations	-0.0916*	-0.273*	-0.0997*	-0.303*
	(0.0214)	(0.0639)	(0.0277)	(0.0841)
Section Order *reference category: First*				
Second	-0.142*	-0.339*	-0.139*	-0.328*
	(0.0269)	(0.0636)	(0.0356)	(0.0839)
Consent to Link *reference category: No*				
Yes	-0.0145	0.163*		
	(0.0348)	(0.0795)		
Gender *reference category: Female*				
Male			-0.000634	-0.0183
			(0.0428)	(0.0972)
Interactions				
Go-Again * Consent	0.170*			
	(0.0555)			
Go-Again * Second Section	0.0683		0.0702	
	(0.0434)		(0.0555)	
Go-Again * Male			-0.0122	
			(0.0645)	
n Loops	2204		1338	
n Respondents	1102		669[a]	
F test of model	$F(6, 1096) = 35.91*$		$F(6, 663) = 16.47*$	

Dependent variables is number of events reported in looping section
Standard errors adjusted for the fact that each respondent appears twice
Constant not shown
* $p < 0.05$
[a] Although 679 cases consented to linkage, ten were missing data on gender

Looking at the first model (columns 1 and 2), in the first row of the table, we see the strong format effect revealed in Figure 1: The go-again format collects one fewer event than the how many format (marginal effect -0.996). In the second row, we see that respondents report fewer locations than they do employers, which is simply a topic effect. The order of the sections also matters: Respondents report on average one-third fewer events in the second section of looping questions than in the first. There are two possible explanations for this finding. Respondents were not able to report more than seven in the first set, and five in the second set, and for this reason, we see reduced reporting in the second section. It is also possible that respondents learn and remember across sections and report fewer events in the second section. Unfortunately, our design does not let us distinguish between these two possible explanations. Those who provide consent to merge in administrative data report slightly more events (marginal effect 0.163) and the format effect is a slightly weaker for consenters than nonconsenters. The format effect is not stronger or weaker in the second section.

The third and fourth columns show the results from a similar model run on a subset of cases, those that consented to administrative record linkage. The results are substantively the same, which supports our use of the cases that consented to linkage between survey and administrative data to address the issue of response accuracy below. We do not see a difference in reporting behavior for men versus women, nor a different format effect for men, in contrast to Carley-Baxter/Peytchev/Black (2010): Their finding that the two formats worked differently for men and women may have been due to the sensitive nature of their topic, violence in sexual relationships.

As predicted by the motivated misreporting hypothesis, the how many format collects more reports of employers and locations than the go-again format. We cannot tell from Figure 1 or from the regression models in Table 2, however, which format's reports are more accurate.

Data Quality in Event Reports

The linked administrative data lets us examine measurement error in the number of employers reported. As discussed in the Section 2, we do not expect the administrative data to match the survey reports in all cases, due to mismatches between respondents' understanding of employers and what is captured in the administrative data, which are derived from records of social security contributions. However, because respondents were randomly allocated to the two looping question formats, we should not see differences in the reporting patterns between the formats unless they perform differently.

Overall, 29% of the respondents' reports on the number of employers matched the count of employers in the administrative data (see the top row of Table 3). Most (62%) of the respondents underreported and about 9% overreported. When we break the report accuracy down by the two formats, however, we see significant and substantial differences between the two. Among respondents in the how many format, 41% reported a number of employers that matches the administrative data, and 48% reported fewer employers than we see in the data. Respondents in the go-again format were less likely to report accurately (18%

of reports) and much more likely to underreport (77%). All of the differences between the formats in Table 3 are significant at the five percent level.[3]

Table 3: Measurement Error in Number of Employers Reported, by Format (Row Percents)

	% Match		% Reporting Fewer		% Reporting More	
Overall	28.7	(1.7)	62.4	(1.9)	8.8	(1.1)
How Many	40.5	(2.7)	48.3	(2.7)	11.1	(1.7)
Go-Again	17.9	(2.1)	76.9	(2.3)	5.2	(1.2)
n^a	195		424		60	
$F(1,678)^b$	44.4*		64.2*		7.93*	

[a] Cases consenting to linkage of responses to administrative data, $n = 679$

[b] Significance test of difference between formats in each column

$*$ Column difference significant at 5% level

Some readers may be surprised at the low match rate between the administrative data and the survey responses even in the how many condition. As discussed in Section 2, there are valid reasons to expect some disagreement between survey responses and the administrative records. Some governmental positions as well as self-employment are not captured in the records (though respondents were instructed not to count periods of self-employment). Firms may change their reporting identification number for reasons that go unnoticed by the employee, and thus what looks like a new employer in the administrative records may not be reported as such. At the same time, the how many format is itself not without error: Respondents may satisfice and give an estimated answer rather than counting each employer. Despite these issues, we believe that the higher match rate, the the lower underreport rate, in the how many format in Table 3 are signals that the how many format is more accurate.

The results in Table 3 support the motivated misreporting hypothesis and are in line with our findings with filter questions (Eckman et al., 2014). When collecting the correct number of events is important in a survey, the how many format performs better than the go-again format.

Data Quality in Follow Ups

The first two research questions addressed the number of events reported in the two formats, but the quality of responses to the follow up questions are also of interest to data analysts. Because we cannot validate responses to the follow up questions with administrative data, we instead compare the missing data rates between the formats.

Figure 2 shows the average number of times a respondent answered “don’t know” to a follow up question or refused to answer a follow up item, by each event reported. The vertical axis is the event number: respondents could report up to seven events in the first

[3] When we include the additional 57 employers collected by the extra question in the how many condition, the share of matching reports falls to 39.9%. All differences are still significant.

loop and up to five in the second loop, though as we saw in Figure 1, many reported fewer. The horizontal axis is the average number of missing responses to the four follow up questions. This analysis combines the employer and location sections, though the two perform similarly.

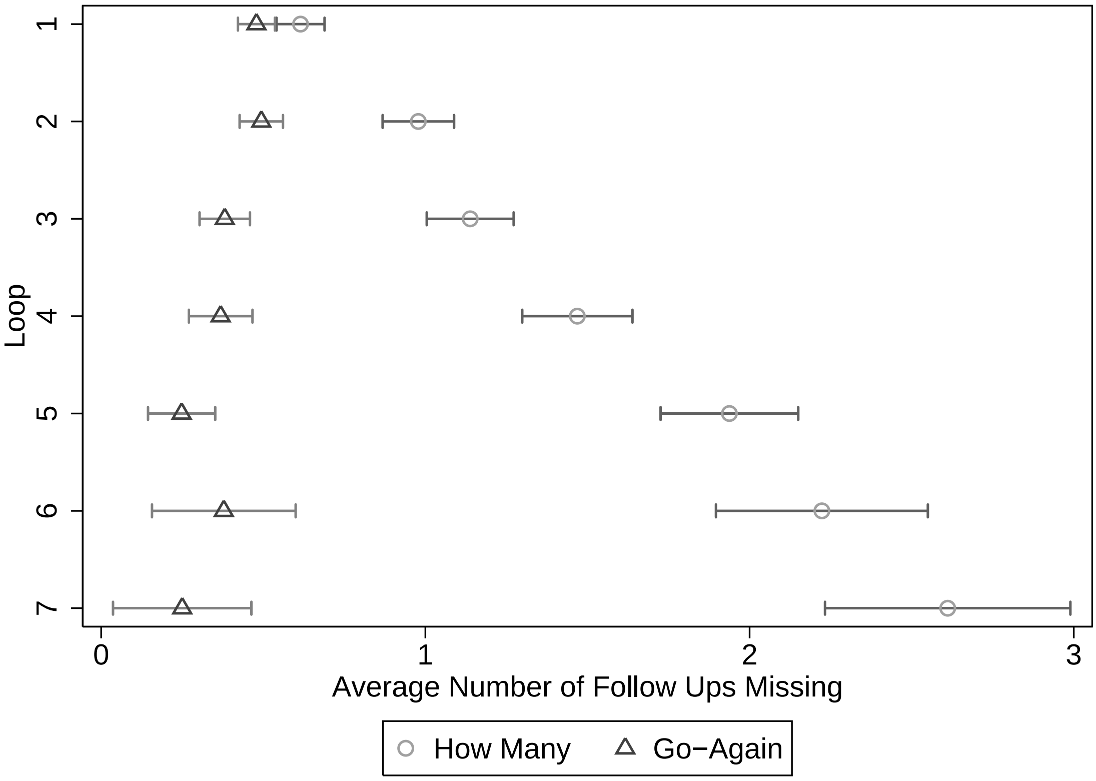

Figure 2: Average Number of Missing Follow Ups, by Event Reported & Format (Standard errors adjusted for the fact that each respondent can appear more than once)

There are strong differences in the quality of follow up information collected in the formats. Respondents in the how many format were much more likely to say “don’t know” or to refuse to answer the follow up questions, than those in the go-again format. Even when reporting about the first event, respondents in the how many format do not answer 0.6 of the four follow up questions, on average. For those in the go-again format, slightly fewer than 0.5 of the four questions have missing values. At the fifth event, the quality of the responses to the how many format has decreased: Respondents do not provide answers to 1.9 of the four follow up items, on average. In contrast, the reporting of those in the go-again format has improved and respondents fail to answer fewer than 0.3 follow up questions. The difference between the two formats is statistically significant, as shown in Figure 2 by the non-overlapping confidence intervals.

In the employment section, the start and end dates of each spell were most often missing. In the locations section, the item on the number of residents of the city or town most often had missing data. While the rates of missing data varied by format and by loop, these items always had the highest rates of missing data.

Both sections asked about the earliest event first, and the follow up questions about this

Table 4: Breakoff Cases, By Format and Section

	n	Breakoffs (%)	Test Statistic[a]
Overall	109	9.0	
Format			
How Many	64	10.4	2.89
Go-Again	45	7.6	
First Loop			
Employers	67	11.0	6.08*
Locations	42	7.0	

[a] $\chi^2(1)$ test of difference in breakoff rates

$*$ Difference significant at 5% level

event might be the hardest for respondents to answer, due to memory decay (Tourangeau/Rips/Rasinski, 2000: Section 3.3.1). As the loop continues, respondents are answering about relatively recent employers or locations, which makes the recall task easier, and thus we might expect fewer missing values to later events. We do see improving data quality with later events in the go-again format: The point estimates in Figure 2 shift to the left as we move down the vertical axis and forward in time. In the how many format, however, the quality of the responses to the follow up questions deteriorates (shifts right) as we ask about later events. By the seventh job or location, these respondents are answering about a relatively recent event, but are failing to answer 2.6 out of the four follow up questions. These results suggest that it is in fact lack of motivation to provide high quality data, rather than the difficulty of recalling the details from long-ago events, that drives the missing data rates in the follow up questions in the how many format.

Breakoffs

An additional aspect of data quality that we might worry about with looping questions, especially when administered via a web survey, is breakoffs. Without an interviewer to encourage continued reporting, web surveys are known to suffer from high breakoff rates (Peytchev, 2009). A respondent who wanted to reduce the burden or length of the questionnaire may choose simply to close her browser window, rather than use the more elaborate underreporting techniques discussed above.

Nine percent (n=109) of the cases that started the first looping section exited the web survey before finishing both looping sections (Table 4). Breakoffs were slightly more common in the how-many format (10.4%) than in the go-again format (7.6%), but the difference is not significant. Breakoffs were also more likely when the employer section was asked first (11.0%) than when the locations section was first (7.0%).

5 Discussion & Conclusion

Looping questions are vulnerable to motivated misreporting, just as filter and screener questions are. The how many and go-again formats collect different numbers of events, and the higher reports of the how many format appear to be more accurate. However, there is a trade-off between the number of events reported and the quality of the data collected by the follow up items: Respondents in the how many format are more likely not to answer the follow up items. Although the breakoff rate is high, about ten percent, there is no significant difference in the rates between the two formats. In other modes, the breakoff rates would likely be smaller and the difference between the formats smaller as well. Taken together, these results are consistent with the interpretation that respondents in the how many format provide higher quality answers to the number of events question, but lower quality in the follow up questions, than respondents in the go-again format.

These results support the motivated misreporting hypothesis more so than the better recall hypothesis. When given an opportunity to shorten the questionnaire, as in the go-again format, respondents do so. Any positive effects of the assistance with recall available in the go-again format are overshadowed by the motivation to answer fewer loops of repetitive follow up questions. The only support we found for the better recall hypothesis was in the responses to the extra follow up question in the how many format, which asked respondents if they had any additional events to report. We were surprised to find that 17% of the responses in the how many format reported one or more additional events to this question. This results suggests that the how many loops stimulated recall and that the respondents did not want to skip additional loops of follow up questions. However, on balance the evidence is more strongly in favor of motivated misreporting in looping questions.

This study asked looping questions about two topics, employers and living locations. Both types of events are at the top level of the memory hierarchy (Belli, 1998), and thus recall is unlikely to be a problem for most respondents. With other topics, for which respondents may not as easily know the true answer, the how many format question may elicit more satisficing, that is more, rate- or impression-based estimation. In such cases, the better recall hypothesis may hold.

The finding of motivated misreporting in looping questions, together with previous evidence of similar misreporting in filter and screener questions (Tourangeau/Kreuter/Eckman, 2012; Eckman et al., 2014), underscores the importance of incentivizing respondents to provide high quality data. It is not enough to write clearly worded questions and to persuade respondents to participate in the survey – evidence is mounting that data quality can suffer if respondents find the questions repetitive or burdensome, and survey designers need to acknowledge and adapt to this phenomenon.

Other question types are likely to be similarly affected by motivated underreporting. A related question type, not shown in Table 1 but present in several of the surveys listed there, involves handing the respondent a show card and asking which of the listed choices apply. For example, in Wave 1 of the SHARE survey, the respondent was shown a list of different pension types and asked which ones she had received income from in 2003; for each type indicated, she was then asked six follow up items. On the web, such questions might be asked in a check-all-that-apply grid. Such questions are similar to the how many loop format in that respondents commit to the different types before they realize that there are follow up questions to come. An alternative formulation of the same questions that is more like the go-again loops would ask a series of filters about each pension type and ask follow up items about the indicated ones. Such loops do not fit into either the how many or go again category, but they are undeniably similar and probably also suffer from motivated misreporting once respondents catch on to how they work.

An alternative to looping questions for collecting life history data is the event history calendar, a form of less standardized interviewing which involves asking respondents about big life events to provide structure and aid in recall. For example, to collect an employment history, a survey may start by having the respondent fill in a calendar of moves, job changes, and relationships. Such events are at the top level in the memory hierarchy and thus help give structure to the recall task (Belli, 1998). Only afterward are respondents asked follow up questions about each job, such as occupation, pay, etc. Belli/Shay/Stafford (2001) compare the event history calendar to what they call the Q-List format, which in our terminology is a mixture of filter questions and go-again and how many loops. They find that the calendar collects more events and is in most cases more accurate, though it at times leads to overreporting. The calendar format also facilitates the cleaning up of spells to ensure that they are complete and consistent (Reimer/Matthes, 2007; Drasch/Matthes, 2013) and seems to provide more cues and context that aid the recall task (Belli et al., 2004).

Event history calendars are a promising approach to the collection of life history data, and Schwarz/Oysermann (2001) recommend their use in other contexts as well, such as short-period recall. However, some researchers worry that event history calendars are too difficult to complete in the telephone or web mode and that non-standardized interviewing can introduce additional measurement error (Drasch/Matthes, 2013). It is clear from Table 1 that surveys have not discontinued the use of looping questions, and thus additional research into the best way to ask such questions is important. The event history calendar research, in turn, should consider the burdens placed on respondents by their questionnaires, and design calendars that motivate respondents to think hard and provide complete data and that do not reveal any repetitive loops of follow up questions before respondents fill in all events.

Practical Advice

Because looping questions are widely used in surveys, and repetition in survey questions cannot be entirely avoided, we close with practical advice for those who wish to use them. In choosing how to ask looping questions, survey designers need to consider which data are of highest importance to later analyses. When the number of events is more important, the how many format should be used. This choice may result in a high rate of missing data in the follow ups. In other research contexts, however, the goal may be to collect detailed information only about a few events: In Table 1, we saw that many surveys cap the number of loops of follow up questions that are asked. In these cases, the go-again format may be a better choice. However, researchers analyzing data collected via the go-again format should keep in mind that the number of events reported is likely too low.

The how many format does have one important advantage over the go-again format. In the how many format, the missing data in the follow ups is obvious, and imputation could be used to fill in the data, making a complete data set. In the go-again format, however, it is not obvious that entire events are missing and analysts can easily overlook this fact. Moreover, because it is not clear how many events are missing for each respondent, imputation is less useful. Although the data set provided by the go-again format appears more complete, because there are fewer cells with missing value codes, the number of events collected by the how many format is more accurate and thus the responses may contain more useful information for analysts. Researchers may favor the how many format for this reason.

We strongly recommend that researchers using the how many format include the additional question at the end asking if there are any other evens the respondents wants to mention. This question give the respondents a chance to report any events that she recalled while answering the follow up loops, and in this way takes advantage of all of the available recall cues.

We also suggest further study with a hybrid type of looping question that asks about the different events one at a time ("Where did you live when you were 14?... And where did you live after that?... And where did you live after that?") but then asks all of the follow up questions at the end. We have found a few instances of such looping questions in the questionnaires we searched through. This approach may provide the benefits of assisted recall, by asking several times about related events in a logical order, but it is also not susceptible to to motivated misreporting, because the follow up questions are not revealed until the respondents has committed to a number of events.

References

Belli, Robert F. (1998): The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys. In: Memory, Vol. 6, No. 4, p. 383–406.

Belli, Robert F.; Lee, Eun Ha; Stafford, Frank P.; Chou, Chia-Hung (2004): Calendar and QuestionList Survey Methods: Association Between Interviewer Behaviors and Data Quality. In: Journal of Official Statistics, Vol. 20, p. 185–218.

Belli, Robert F.; Shay, William L.; Stafford, Frank P. (2001): Event History Calendars and Question List Surveys: A Direct Comparison of Interviewing Methods. In: Public Opinion Quarterly, Vol. 65, No. 1, p. 45–74, URL http://poq.oxfordjournals.org/content/65/1/45.abstract.

Berg, Marco; Cramer, Ralph; Dickmann, Christian; Gilberg, Reiner; Jesske, Birgit; Kleudgen, Martin; Bethmann, Arne; Fuchs, Benjamin; Huber, Martina; Trappmann, Mark (2014): Codebuch und Dokumentation des “Panel Arbeitsmarkt und soziale Sicherung” (PASS) Band III: Personendatensatz (PENDAT). FDZ Datenreport, URL http://doku.iab.de/fdz/reporte/2014/DR_02-14_III.pdf.

Blair, Edward; Burton, Scot (1987): Cognitive Processes Used by Survey Respondent to Answer Behavioral Frequency Questions. In: Journal of Consumer Research, Vol. 14, p. 280–288.

Börsch-Supan, Axel; Jürges, H. (2005): The Survey of Health, Ageing and Retirement in Europe – Methodology. Tech. Rep..

Bradburn, Norman M. (1979): Respondent Burden. In: Health Survey Research Methods: Second Biennial Conference, Williamsburg, VA.

Burton, Scot; Blair, Edward (1991): Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys. In: Public Opinion Quarterly, Vol. 55, p. 50–79.

Cannell, C.F.; Oksenberg, L.; Kalton, G.; Bischoping, K.; Fowler, F. J. (1989): New Techniques for Pretesting Survey Questions. Tech. Rep., Survey Research Center University of Michigan.

Carley-Baxter, Lisa R.; Peytchev, Andy; Black, Michele Lynberg (2010): Effect of Questionnaire Structure on Nonresponse and Measurement Error: Sequential vs. Grouped Placement of Filter Questions. Presentation at American Association for Public Opinion Research Conference, phoenix, AZ.

Central Co-ordinating Team (2010): European Social Survey Round 3 2008/2009. Final Activity Report ESS4e03.0, City University London.

Drasch, K.; Matthes, B. (2013): Improving Retrospective Life Course Data by Combining Modularized Self-Reports and Event History Calendars: Experiences From a Large Scale Survey. In: Quality & Quantity, Vol. 47, No. 2, p. 817–838.

Eckman, Stephanie; Kreuter, Frauke; Kirchner, Antje; Jäckle, Annette; Presser, Stanley; Tourangeau, Roger (2014): Assessing the Mechanisms of Misreporting to Filter Questions. In: Public Opinion Quarterly, Vol. 78, No. 3, p. 721–733.

Felderer, Barbara; Kreuter, Frauke; Winter, Joachim (2013): Can We Buy Good Answers? The Inuence of Respondent Incentives on Item Nonresponse and Measurement Error in a Web Survey. Presented at the American Association for Public Opinion Research Annual Conference, Boston, MA.

Health and Retirement Survey (2011): Sample Sizes and Response Rates. Tech. Rep., http://hrsonline.isr.umich.edu/sitedocs/sampleresponse.pdf.

IAB Institut für Arbeitsmarkt- und Berufsforschung (2013): Nuremberg: Integrierte Erwerbsbiographien (IEB) V09.03. Tech. Rep..

IAB Institut für Arbeitsmarkt- und Berufsforschung (2011): Nuremberg: Integrierte Erwerbsbiographien (IEB) V09.00. Tech. Rep..

Jacobebbinghaus, Peter; Seth, Stefan (2007): The German Integrated Employment Biographies Sample IEBS. In: Schmollers Jahrbuch Zeitschrift für Wirtschafts- und Sozialwissenschaften, Vol. 127, p. 335–342.

Krosnick, Jon A. (1991): Response Strategies for Coping With the Cognitive Demands of Attitude Measures in Surveys. In: Applied Cognitive Psychology, , No. 5, p. 213–236.

Krosnick, Jon A.; Presser, Stanley (2010): Question and Questionnaire Design. In: Marsden, Peter V.; Wright, James D. (Eds.) Handbook of Survey Research, Emerald Group Publishing Limited, p. 263–313.

Lepkowski, James M.; Mosher, William D.; Davis, K.E.; Groves, Robert M.; Hoewyk, John Van (2010): The 2006–2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey. In: Vital Health Statistics, Vol. 2, No. 150.

Moore, Whitney; Pedlow, Steven; Krishnamurty, Parvati; Wolter, Kirk (2000): National Longitudinal Survey of Youth 1997 (NLSY97) Technical Sampling Report. Tech. Rep..

Panel Survey on Income Dynamics (2013): PSID Main Interview User Manual. Tech. Rep., http://psidonline.isr.umich.edu/data/Documentation/UserGuide2011.pdf.

Peytchev, Andy (2009): Survey Breakoff. In: Public Opinion Quarterly, Vol. 73, No. 1, p. 74–97, URL http://poq.oxfordjournals.org/content/73/1/74.abstract.

Reimer, M.; Matthes, B. (2007): Collecting Event Histories With TrueTales: Techniques to Improve Autobiographical Recall Problems in Standardized Interviews. In: Quality & Quantity, Vol. 41, No. 5, p. 711–735.

Sakshaug, Joseph W.; Kreuter, Frauke (2014): The Effect of Benefit Wording on Consent to Link Survey and Administrative Records in a Web Survey. In: Public Opinion Quarterly, Vol. 78, No. 1, p. 166–176, URL http://poq.oxfordjournals.org/content/78/1/166.abstract.

Schwarz, N.; Oysermann, D. (2001): Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction. In: American Journal of Evaluation, Vol. 22, No. 2, p. 127–160.

Sharp, Laure M.; Frankel, Joanne (1983): Respondent Burden: A Test of Some Common Assumptions. In: Public Opinion Quarterly, Vol. 47, No. 1, p. 36–53.

Tourangeau, R.; Kreuter, F.; Eckman, S. (2012): Motivated Underreporting in Screening Interviews. In: Public Opinion Quarterly, Vol. 76, No. 3, p. 453–469, URL http://dx.doi.org/10.1093/poq/nfs033.

Tourangeau, Roger; Bradburn, Norman M. (2010): The Psychology of Survey Response. In: Marsden, Peter V.; Wright, James D. (Eds.) Handbook of Survey Research, Emerald Group Publishing Limited, p. 315–346.

Tourangeau, Roger; Kreuter, Frauke; Eckman, Stephanie (2015): Motivated Misreporting: Shaping Answers to Reduce Survey Burden. In: Engel, Uwe (Ed.) Survey Measurements. Techniques, Data Quality and Sources of Error, Frankfurt/New York: Campus, p. 24–41.

Tourangeau, Roger; Rips, Lance J.; Rasinski, Kenneth (2000): The Psychology of Survey Response. Cambridge University Press, cambridge.

Recently published

No.	Author(s)	Title	Date
14/2015	Reichelt, M. Abraham, M.	Occupational and regional mobility as substitutes: A new approach to understanding job changes and wage inequality	4/15
15/2015	Zapf, I.	Individual and workplace-specific determinants of paid and unpaid overtime work in Germany	4/15
16/2015	Horbach, J. Janser, M.	The role of innovation and agglomeration for employment growth in the environmental sector	6/15
17/2015	Dorner, M. Fryges, H. Schopen, K.	Wages in high-tech start-ups – do academic spin-offs pay a wage premium?	6/15
18/2015	Möller, J.	Verheißung oder Bedrohung? Die Arbeitsmarktwirkungen einer vierten industriellen Revolution	6/15
19/2015	Hecht, V.	Location choice of German multinationals in the Czech Republic: The importance of agglomeration economies	7/15
20/2015	Wiemers, J.	Endogenizing take-up of social assistance in a microsimulation model: A case study for Germany	7/15
21/2015	Wanger, S. Weigand, R. Zapf, I.	Measuring hours worked in Germany: Contents, data and methodological essentials of the IAB working time measurement concept	8/15
22/2015	Weigand, R. Wanger, S. Zapf, I.	Factor structural time series models for official statistics with an application to hours worked in Germany	8/15
23/2015	Zapf, I.	Who profits from working-time accounts? Empirical evidence on the determinants of working-time accounts on the employers‘ and employees‘ side	8/15
24/2015	Dietrich, H.	Jugendarbeitslosigkeit aus einer europäischen Perspektive: Theoretische Ansätze, empirische Konzepte und ausgewählte Befunde	9/15
25/2015	Christoph, B.	Empirische Maße zur Erfassung von Armut und materiellen Lebensbedingungen: Ansätze und Konzepte im Überblick	9/15
26/2015	Bauer, A.	Reallocation patterns across occupations	9/15
27/2015	Dauth, W. Fuchs, M. Otto, A.	Long-run processes of geographical concentration and dispersion: Evidence from Germany	9/15
28/2015	Klinger, S. Weber, E.	Detecting unemployment hysteresis: A simultaneous unobserved components model with Markov switching	10/15

As per: 2015-10-26

For a full list, consult the IAB website
http://www.iab.de/de/publikationen/discussionpaper.aspx

Imprint

IAB-Discussion Paper 29/2015

Editorial address
Institute for Employment Research
of the Federal Employment Agency
Regensburger Str. 104
D-90478 Nuremberg

Editorial staff
Regina Stoll, Jutta Palm-Nowak

Technical completion
Gertrud Steele

All rights reserved
Reproduction and distribution in any form, also in parts,
requires the permission of IAB Nuremberg

Website
http://www.iab.de

Download of this Discussion Paper
http://doku.iab.de/discussionpapers/2015/dp2915.pdf

ISSN 2195-2663

For further inquiries contact the author:

Frauke Kreuter
Phone +49.911.179 1358
E-mail frauke.kreuter@iab.de