Measurement of Interviewer Workload within the Survey and an Exploration of Workload Effects on Interviewers' Field Efforts and Performance

Celine Wuyts¹ and Geert Loosveldt¹

Interviewer characteristics are usually assumed fixed over the fieldwork period. The number of sample units that require the interviewers' attention, however, can vary strongly over the fieldwork period. Different workload levels produce different constraints on the time interviewers have available to contact, recruit and interview each target respondent, and may also induce different motivational effects on interviewers' behavior as they perform their different tasks. In this article we show that fine-grained, time-varying operationalizations of project-specific workload can be useful to explain differences in interviewers' field efforts and achieved response outcomes over the fieldwork period. We derive project-specific workload for each interviewer on each day of fieldwork in two rounds of the European Social Survey in Belgium from contact history and assignment paradata. Project-specific workload is measured as (1) the number of sample units which have been and remain assigned on any day t (assigned case workload), and (2) the number of sample units for which interviewer activity has started and not yet ceased on any day t (active case workload). Capturing temporal variation in interviewers' workloads in a direct way, the time-varying operationalizations, are better predictors than are the interviewer-level operationalizations of typical (active or potential) workload that are derived from them, as well as the traditional total-count workload operationalization.

Key words: Nonresponse; interviewer effort; interviewer effects; time-varying interviewer characteristics; paradata.

1. Introduction

Interviewer workload is widely believed to be a constraining factor on the effort survey interviewers are able and willing to apply in the face-to-face recruitment of respondents, but the available empirical evidence in support of these beliefs is sparse. This article addresses an existing gap in the literature on interviewer workload as a source of survey interviewers' differential performance in the contact and recruitment task. The interviewers' performance in the contact and recruitment task, that is, the extent to which the demands of this task are adequately met, can be evaluated from a data quality perspective in terms of the interviewers' contribution to nonresponse error, which in turn depends on the interviewers' effort in applying the contact procedures (e.g., number and timing of contact attempts) (Loosveldt et al. 2004).

¹ Centre for Sociological Research, Catholic University of Leuven, Parkstraat 45 - bus 3601, 3000 Leuven, Belgium. Emails: celine.wuyts@kuleuven.be and geert.loosveldt@kuleuven.be

We focus specifically on interviewer workload in one particular survey project and will refer to this workload component as ‘project-specific workload’. We do not elaborate on work resulting from activities and responsibilities other than involvement in the survey under study (e.g., involvement in other survey projects and other jobs) even though all work components may combine to constrain time availability and alter interviewers’ incentives and motivations. Project-specific workloads can be controlled, if not fully determined, by the field supervisors, within the constraints imposed by the survey design and the available interviewer workforce.

The total amount of work involved in recruiting and interviewing respondents in any particular survey is driven by survey design features such as the mode, the contact procedures and the sampling design, and characteristics of the target population such as its geographical distribution and accessibility. The capacity of the interviewer workforce as a whole to deliver the desired results is determined by the number of interviewers available, the interviewers’ temporal and (in the case of face-to-face surveys) geographical distribution, and the interviewers’ experience and skill. The size and composition of the available interviewer workforce, relative to the size and composition of the (gross) sample size, places constraints on the way the work is allocated.

Project-specific interviewer workloads are conventionally quantified in terms of numbers of sample units, that is, the unit of measurement is the sample unit, and result from an intricate allocation process. Sample units may be (initially) allocated all at once at the start of the fieldwork, or in several batches over the fieldwork period. After the initial allocation, sample units may also be reallocated at some point. This may be done as a deliberate strategy to improve response rates as well as in response to interviewers dropping out of the interviewer workforce. Reallocation as a response enhancement strategy is premised on the recognition that even if (repeated) attempts to make contact and obtain cooperation by one interviewer are unsuccessful, additional attempts by a different interviewer may still yield a completed interview (e.g., [Calderwood et al. 2016](#)). Interviewers are usually assigned more sample units, both initially and in view of response enhancement, if they are expected to be more successful based on past performance, and if they reside in areas with high target population density relative to interviewer availability.

Different workload levels produce different constraints on the time interviewers have available to contact, recruit and interview each target respondent, and induce interviewers to adjust their behavior accordingly. Interviewers have limited available time and energy to allocate over the sample units assigned to them. Larger workloads imply that each case in an interviewer’s workload on average receives a smaller share of the interviewer’s total available time, and in particular time available during the most productive days (weekends) and on the most productive hours (evenings) ([Botman and Thornberry 1992](#); [Groves and Couper 1998](#), 274). The attempts that are made to contact and recruit each target respondent are likely fewer in number and less optimally timed if workloads are large ([Japac 2008](#)). Given that (face-to-face) interviewers typically have full discretion in dividing their time and energy among the sample units assigned to them, and especially under piece-rate payment schemes, whereby interviewers are paid per completed interview, larger workloads are likely to induce interviewers to focus their attention to easily reached and cooperative target respondents. Interviewers may make fewer

round-trips to relatively sparsely populated areas, disregard when target respondents are likely to be at home, and push for a quick decision at the doorstep, rather than carefully planning visits, tailoring and maintaining the interaction in a single or over multiple visits. Geographical clustering of sample units and scale efficiencies in travel time may encourage interviewers to make more round-trips to distant geographical areas when workloads for those areas are large, and induce them to give up on these areas entirely when workloads are small.

The idea of interviewers taking shortcuts when they experience tasks as burdensome, for example because of a heavy workload, was advanced by Japac (2005; 2008). These shortcuts are generally intended to reduce the amount of time and effort invested for each individual case below the level corresponding to the researchers' data quality targets. Japac (2005; 2008) therefore uses the term 'interviewer satisficing'. Interviewer satisficing behaviors in contact and recruitment put additional pressure on the trade-off between achieving response rates that are acceptable from a quality perspective and the cost of extending and/or intensifying fieldwork operations.

Workload-related design choices and allocation procedures may therefore affect the survey data collection process and the quality of the collected data. Investigations into the nature and impacts of interviewer workload should be highly relevant to survey practitioners but are surprisingly sparsely documented. Most of the available literature on explaining differential interviewer performance, whether in the contact and recruitment task or in the interview administration task, focuses on sociodemographics and interviewer characteristics related to experience, skill and confidence (West and Blom 2016), rather than motivation and effort or their determinants. The limited available evidence on the workload-performance relationship in the contact and recruitment task of survey interviewers is not only sparse but the usual approach to measuring workload has also not been altogether convincing and invites further research.

In the current article, we examine the association between interviewers' project-specific workload on the one hand and expended efforts and achieved response rates on the other. Contact history and assignment paradata for two rounds of the European Social Survey in Belgium allow the construction of more fine-grained measurements of project-specific workload over the fieldwork period than traditionally have been used. Project-specific workload and other interviewer characteristics (see e.g., West and Blom 2016 for a recent overview) are usually considered as fixed over the fieldwork period. As will be discussed in the following section, valuable insights into the underlying mechanisms may be gained by observing project-specific workload as a time-varying interviewer characteristic.

The following section presents some important considerations in the measurement of project-specific workload, summarizes the available literature on the link between interviewers' project-specific workloads and response rates and elaborates on the limitations of the traditional approach to workload measurement, before introducing an alternative perspective and associated alternative operationalizations. In particular, we advance two approaches to measure interviewer workload on each day in the fieldwork: (1) the number of sample units which have been and remain assigned on any day t (assigned case workload), and (2) the number of sample units for which interviewer activity has started and not yet ceased on any day t (active case workload).

2. Measurement of Project-Specific Workload

A common approach is to measure project-specific workload as the total number of sample units worked on by an interviewer for the survey project (e.g., [Blom 2012](#); [Beullens et al. 2016](#)). [Beullens et al. \(2016\)](#), for example, observe that interviewers in Round 7 of the European Social Survey worked on average between 12 (Czech Republic) and 53 (Switzerland) sample units over the course of the fieldwork. This is, however, only one of multiple possible approaches.

There are two key measurement issues to consider: (1) which sample units to include, and (2) whether and how temporal variation over the fieldwork period is taken into account. With regard to the first question, two approaches appear reasonable. The first approach would be to include sample units if they are assigned to the interviewer and thus could potentially be pursued by the interviewer. The second approach would be to include sample units only if they actually are actively being pursued by the interviewer.

Both the number of sample units that have been and remain assigned and the number of sample units that are actually being pursued may vary over time. Many sample units naturally cease to entail any additional work at some point, because an interview has been completed, ineligibility has been established, or the sample unit is abandoned as unproductive. The remaining number of sample units that require any attention therefore (more or less gradually) shrinks over time, at least until additional sample units are assigned. More erratic fluctuations may occur especially when sample units are assigned in batches or reallocated over the fieldwork period. Whether and how these changes are taken into account in measuring project-specific workload constitutes an important measurement decision.

Only a handful of empirical studies have reported estimates of correlations between interviewers' project-specific workloads and response rates, usually with workload as one of many interviewer characteristics under consideration. None have taken into account temporal variation or investigated the effect of interviewer workload on nonresponse error in a comprehensive way.

An early mention of workload as an interviewer characteristic was made by [Singer et al. \(1983\)](#). [Singer et al. \(1983\)](#) observed a negative correlation between interviewer workload (operationalized as the total number of cases assigned to each interviewer) and response rates in a random digit dialing, RDD telephone survey. The observed screening rate dropped from 91% for interviewers with fewer than 30 cases to 84% for interviewers who were assigned a total of 88 or more cases. The cooperation rate dropped from 78% to 60%. They offer workload fatigue or burden as one possible explanation, but also acknowledge that the result may be an artefact of the case assignment procedure. Some interviewers were assigned large numbers of cases at the end of fieldwork and these cases may have been more difficult and/or may not have been pursued as intensively due to lack of time.

[Nicoletti and Buck \(2004\)](#) reported that interviewer workload (operationalized as the total number of households assigned to each interviewer) was negatively associated with the probability of contact in two of four household panel surveys studied and with the probability of cooperation, conditional on contact, in three of the surveys. For the German Socio-Economic Panel survey, they observe the expected negative association between

interviewer workload and the probability of cooperation but a *positive* association for the probability of contact. A positive association between interviewer workload (operationalized as the total number of previous-wave respondents assigned at the start of the fieldwork) and contact probability in a panel survey (at least for reasonably sized workloads of fewer than 124 cases) has also been reported by [Watson and Wooden \(2009\)](#). While claiming that large workloads may in fact be beneficial for response, they do admit that the observed pattern may be an artefact of the case assignment procedure. Larger numbers of cases had commonly been assigned to the best interviewers.

[Blom et al. \(2011\)](#), on the other hand, found no evidence in support of the hypothesized negative effect of interviewer workload (operationalized as the total number of cases worked on over the course of the fieldwork) on contact and cooperation rates in seven countries in the first round of the European Social Survey. [Blom \(2012\)](#), examining a partially different set of seven countries, likewise did not observe the expected association, except for one country. She also suggests that in most countries larger numbers of cases may have been assigned to the best interviewers, counterbalancing a potential negative effect of workload on response rates.

In all of these studies, project-specific workload has been operationalized by total-count measures at the interviewer level, that is, the total number of sample units assigned to each interviewer or the total number of sample units worked on by each interviewer. Total-count measures at the interviewer level are usually readily available or easy to derive but have their limitations. The main limitation of using a total-count measure to explain interviewers' differential task performance, expressed amongst others by [Blom \(2012\)](#), is the risk of bias due to reverse causality. Concerns about interviewers' performance in the survey project driving their total case workload are warranted. In addition, by adopting a total-count measure, one makes abstraction of any temporal variation in the number of sample units that require the interviewers' attention over the fieldwork period. Note that because interviewers rarely leave sample units that have been assigned to them completely untouched over the entire fieldwork period, the question of which sample units to include should not be of great importance for this type of operationalization. Any discrepancies between the number of sample units assigned and the number of sample units actively worked on will be minor, and negligible in practice.

As previously discussed, temporal variation results from particularities of the allocation process and the natural progression of fieldwork. At any moment in time, we would expect interviewers to adjust their behavior in response to the number of sample units that require their attention at that time. We therefore propose that operationalizations of project-specific workload that take temporal variation over the fieldwork into account are more appropriate to explain interviewers' fieldwork effort and task performance.

A useful strategy would thus be to measure project-specific workload at different points in time over the fieldwork period. Time-varying operationalizations of project-specific workload capture temporal variation in a direct way. We would also argue that by observing variability in project-specific workload over time in addition to variability in project-specific workload across interviewers, the risk of bias due to reverse causality between workload and task performance may be partially mitigated. At a particular point in time, it seems much more likely that the interviewers' workload affects their expended effort and achieved response rates at that time than the other way around.

The question of which sample units to include at each time point remains to be answered. The inclusion of sample units hinges on the identification of the moments in time at which sample units can be presumed to enter and leave the interviewers' workloads. We previously presented two reasonable approaches. The first approach would consider a sample unit as being in an interviewer's workload as long as the interviewer could potentially pursue the unit in question. Each sample unit is assumed to enter the interviewer's workload when the interviewer is assigned the sample unit and leaves the interviewer's workload when the sample unit is returned to the field office. Applying this approach, 'assigned' case workload at time t is measured as the number of sample units that have been and remain assigned at that time. The second approach would consider a sample unit as being in an interviewer's workload as long as the interviewer is actively pursuing the unit in question. Each sample unit is assumed to enter the interviewer's workload when the interviewer first attempts to contact the sample unit and leaves the interviewer's workload after the last attempt. Applying this second approach, 'active' case workload at time t is measured as the number of sample units for which interviewer activity has started and not yet ceased at that time. Figure 1 illustrates the application of the two approaches for a hypothetical interviewer with a total case workload of four sample units. Sample unit 1 leaves both active and assigned case workload after the completed interview. Sample units 2, 3 and 4 leave assigned case workload after being returned to the field office (at day 9), but leave active case workload after the last unsuccessful attempt (at day 7 for sample units 2 and 3, at day 8 for sample unit 4). Note that for these time-varying operationalizations, the choice of which sample units to include is actually highly relevant. Discrepancies between the number of sample units that have been and remain assigned and the number of sample units actively worked on may be considerable.

The first operationalization of *assigned* case workload is conceptually superior to the operationalization of *active* case workload in that it better captures the amount of work interviewers are expected to do, rather than just the amount of work interviewers are currently engaged with in the field. Cases that are assigned but not (yet) actively engaged with may carry comparable weight because such cases still involve some planning and

Time	1	2	3	4	5	6	7	8	9	10
SU1	A		NR	I						
SU2	A	NR	NR				NR		R	
SU3						A	NR		R	
SU4						A	NR	NR	R	
Assigned case workload	2	2	2	2	1	3	3	3	3	0
Active case workload	0	1	2	2	1	1	3	1	0	0

Note: A = sample unit is assigned; R = sample unit is returned to the field office; NR = unsuccessful contact attempt (no interview); I = interview.

Fig. 1. Illustration of two approaches to measuring project-specific workload over time.

administrative work. In addition, the *active* case workload operationalization is endogenous to interviewers' field efforts. Active case workload will be lower if assigned sample units are being shelved or abandoned, and higher when more sample units are being attempted. As this operationalization itself depends heavily on actual interviewer activity, it may not be particularly well suited to assess workload effects on field efforts and outcomes. Estimated "effects", especially with respect to interviewers' field efforts, should be interpreted with caution. Because *assigned* case workload depends primarily on the survey agency's case assignment practices, and is less sensitive to actual interviewer activity, causal interpretations of estimated workload effects should be less questionable for this operationalization.

From the two time-varying operationalizations of project-specific workload, we can also derive operationalizations at the interviewer level. A (temporally weighted) average (active/assigned) case workload better approximates workload on any one typical fieldwork day than the traditional total-count measures at the interviewer level.

3. Study Objectives

The main objective of this study is to develop and assess alternative project-specific workload operationalizations at two levels of aggregation. We adopt new operationalizations expressing *time-varying* interviewer characteristics (at the level of interviewer days) and operationalizations derived thereof and expressing *fixed* interviewer characteristics (at the interviewer level).

Our conception of project-specific workload measured at different points in time over the fieldwork period is inspired by, and analogous to, the conception of within-survey experience measured at each interview over the fieldwork period. The introduction of *time-varying* interviewer characteristics to explain changes in interviewer behavior and performance can be attributed to [Olson and Peytchev \(2007\)](#), who proposed using interview order to capture interviewers' increasing familiarity with the survey instrument over the fieldwork period. Whereas interview order has become the standard operationalization of interviewers' within-survey experience (e.g., [Loosveldt and Beullens 2013](#); [Kirchner and Olson 2017](#)), most other interviewer characteristics studied in the survey research literature remain defined and operationalized at the interviewer level. Many interviewer characteristics can indeed be assumed sufficiently stable over the fieldwork period. This assumption is plausible for interviewer characteristics such as age, gender, race and personality traits, and likely for many attitudes as well, but not for the amount of work interviewers are expected to do.

The secondary objective is to apply these different operationalizations in an exploration of whether the commonly held beliefs about large interviewer workloads negatively affecting fieldwork outcomes in face-to-face surveys are legitimate. Since face-to-face contact and recruitment demand a large share of the total time and effort expended by the interviewers, we may expect a negative effect of workload on the number of personal visits made to each individual case in their workload because of limited time availability. Both contact and cooperation rates may to some extent be affected by large workloads implying that relatively fewer visits are made during the more productive hours, and cooperation rates may be further reduced by less effortful doorstep interactions. We therefore will

evaluate the validity of the following two assertions, using the alternative workload operationalizations.

Hypothesis 1: When interviewers carry larger workloads, their contact effort per case (in quantitative terms) is reduced.

Hypothesis 2: When interviewers carry larger workloads, their contact and cooperation rates are reduced.

4. Data and Methods

We use data from Round 6 and Round 7 of the European Social Survey (ESS) in Belgium (European Social Survey 2014, 2016). The units in the Belgian gross samples, individual persons drawn from the National Register, were assigned to individual interviewers in several batches over the fieldwork period, and relatively large numbers of non-responding sample units (both initial non-contacts and initial refusals) were reassigned to other interviewers. Such a fieldwork strategy complicates the operationalization of project-specific workload over the fieldwork period, compared to a fieldwork strategy with all units allocated only once at the start of fieldwork. The ESS in Belgium therefore represents an attractive case for the current study.

In Round 6 (ESS6-BE) 3,267 sample units were issued (of which 622 initial nonrespondents were reassigned and worked on by another interviewer), 155 interviewers were employed and 1,869 interviews were administered (response rate 58.7%) between September 2012 and the end of December 2012. In Round 7 (ESS7-BE 2014-2015) 3,204 sample units were issued (of which 1,041 initial nonrespondents were reassigned and worked on by another interviewer), 151 interviewers were employed and 1,769 interviews were administered (response rate 57.0%) between September 2014 and the end of January 2015. The same survey agency was contracted and the two groups of interviewers partially overlap. The interviewers were paid per completed interview, with the piece rate itself adjusted if an interviewer exceeded a threshold response rate, administered interviews and collected paradata of sufficiently high quality, and completed his or her assignments on time.

We construct several measurements of the main interviewer characteristic of interest, project-specific workload, from the publicly available ESS contact history data (“contact forms data”) and additional interviewer assignment data. The ESS contact history data set contains a detailed record of when sample units were attempted, by which interviewer, by which mode and with which outcome. The interviewer assignment data set on which we can draw for these two survey rounds of the ESS in Belgium contains a record of when sample units were assigned to which interviewer.

The following two subsections offer a detailed description of the operationalizations at the level of interviewer-days (workload as a time-varying interviewer characteristic) and at the interviewer level (workload as a fixed interviewer characteristic), respectively.

4.1. Operationalizing Project-Specific Workload at the Level of Interviewer-Days

We derive daily project-specific workload measures along the lines of the two approaches described above (active case workload and assigned case workload), and by taking into account reassigned sample units as well as sample units initially assigned at the start or over

the course of the fieldwork ($N = 3,889$ in Round 6 and $N = 4,245$ in Round 7). The number of sample units for which interviewer activity has started and not yet ceased (active case workload) can be derived for each interviewer and on each fieldwork day from the contact history data. Each sample unit enters an interviewer's *active* case workload on the date of the first recorded attempt and leaves the interviewer's active case workload on the date of the last recorded attempt by that interviewer. The number of sample units that have been and remain assigned (assigned case workload) could similarly be derived for each interviewer and each fieldwork day from detailed assignment data. Each sample unit enters an interviewer's *assigned* case workload on the recorded date of assignment, unless the date of assignment is missing ($N = 2$ in Round 6, $N = 6$ in Round 7) or follows rather than precedes the first contact attempt ($N = 83$ in Round 6, $N = 97$ in Round 7), in which case the date of the first recorded attempt is taken. Ideally, the interviewer assignment data set would include, for each sample unit that is assigned to a particular interviewer, not only the date of assignment but also the date of return to the field office. Because data on case returns to the field office are not available in this study, each sample unit is assumed to leave the interviewer's assigned case workload on the date of the last recorded attempt in the assignment set (i.e., the group of sample units assigned on the same date and in the same geographical area). Note that assignments of sample units to interviewers for which no attempt is recorded in the contact history data are therefore necessarily excluded here (4% and 7% of assignments in Round 6 and Round 7, respectively). Respondents leave the interviewer's workload on the date of the completed interview. Sample units identified as deceased also leave the interviewer's workload. Because under the Belgian ESS fieldwork strategy each sample unit is assigned to a single interviewer at any one point in time, we should not observe any sample units simultaneously counted in multiple interviewers' workloads. We observe only 0.1% and 2.1% of sample units simultaneously counted in multiple interviewers' assigned case workloads in Rounds 6 and 7, respectively. Aggregating the number of sample units that have entered, and have not yet left, the (active or assigned) workload of the interviewer at time t yields the interviewer's (active or assigned) project-specific workload at that time t .

Using these two time-varying operationalizations of project-specific workload, we first descriptively examine the degree to which temporal variation in project-specific workload over the fieldwork period actually occurs. Figure 2 provides an illustration of an interviewer's project-specific workload trajectory over the fieldwork period according to the two approaches. The interviewer presented was assigned a total of 64 sample units. The first panel tracks the number of sample units the interviewer started, and has not yet stopped, pursuing on each day over the fieldwork (active case workload). The second panel tracks the number of sample units the interviewer was assigned and remain to be worked (assigned case workload). This visualization shows strong fluctuations in project-specific workload over the fieldwork period, and supports the proposed relevance of workload operationalizations that take temporal variation into account. The visualization also demonstrates that the two time-varying operationalizations produce measurements that are strongly related but far from identical. As would be expected, assigned case workload is not only higher on average (as indicated by the dotted line) but also changes in broader steps over time.

We observe project-specific workload (for interviewer j on day t) for 155 interviewers over 106 fieldwork days in Round 6 and for 151 interviewers over 145 fieldwork days in

Fig. 2. Illustration of an interviewer's project-specific workload over the fieldwork period.

Round 7. [Table 1](#) presents descriptive statistics for the two project-specific workload operationalizations at the interviewer-day level. The upper panel presents the mean and standard deviation across all interviewer-days. The lower panel presents the mean and standard deviation across the interviewer-days that will be included in the analysis. Given

Table 1. Descriptive statistics of project-specific workload at the interviewer-day level.

	ESS6-BE		ESS7-BE	
	Mean	SD	Mean	SD
<i>All interviewer-days</i>				
Active case workload	2.51	3.63	2.22	3.60
Assigned case workload	6.55	7.28	6.08	7.84
<i>N</i>	16,430		21,895	
<i>Interviewer-days with strictly positive active case workload</i>				
Active case workload	5.35	3.58	5.26	3.82
Assigned case workload	11.9	5.54	12.31	6.96
<i>N</i>	7,702		9,259	

that no interviewer activity can be expected when interviewers have no cases to work on, the analysis only makes sense for interviewer-days on which workload is strictly positive. The boundaries of the relevant analytical sample thus also depend on the chosen operationalization. In order to make valid comparisons across operationalizations, we drop interviewer-days with zero workload according to either operationalization, which in practice corresponds to dropping interviewer-days with zero *active* case workload because assigned case workload is always at least as large as active case workload.

Including all interviewer-days, the two measures exhibit disproportionally high variability and are strongly correlated ($r = .69$ in Round 6, $r = .71$ in Round 7). The variances and correlation are driven up, and the averages are driven down, by the large numbers of interviewer-days at which workload is zero. Active case workload is zero on 53% and 58% of interviewer days in Round 6 and Round 7, respectively. Assigned case workload is zero on 46% and 51% of interviewer days in Round 6 and Round 7, respectively. The number of days with zero workload varies considerably across interviewers. The interviewers had strictly positive active case workloads on between 1 and 131 fieldwork days in Round 6 ($M = 61.32$, $SD = 30.57$) and between 4 and 100 fieldwork days in Round 7 ($M = 49.69$, $SD = 21.61$). This is to a large extent due to the selective engagement of the geographically better located and the more effective interviewers among those available for the second phase of the fieldwork, when many non-responding sample units were reallocated. Within the relevant analytical samples ($N = 7,702$ in Round 6 and $N = 9,259$ in Round 7), the bivariate correlation between the two workload measures is much reduced ($r = .49$ in Round 6, $r = .53$ in Round 7), but remains positive and significant.

Fig. 3. Average active case workload (and 95% confidence interval) for active interviewers over the fieldwork period.

Note: The pattern for assigned case workload is highly similar.

The intensity of project-specific workload variation over the fieldwork period is clearly demonstrated by the workload pattern observed among active interviewers (Figure 3). The cyclic pattern in the average project-specific workload over the fieldwork period is very similar for the two rounds. The average workload remains fairly stable over the first month, then gradually decreases over the remainder of the first phase of the fieldwork. At the start of the second fieldwork phase, characterized by intensive reallocations, the average workload rebounds to a level similar to that at the start of the fieldwork, before gradually decreasing again.

4.2. Aggregating Project-Specific Workload to the Interviewer Level

In order to compare the predictive power of project-specific workload as a time-varying interviewer characteristic to project-specific workload as a fixed interviewer characteristic, we also construct the corresponding interviewer-level summary measures of the time-varying measurements. Taking the average of the time-varying measurements for each interviewer essentially corresponds to weighting (sets of) sample units by the amount of time they are in the interviewers’ workloads. Whereas the traditional total-count measure captures case workload over the entire fieldwork period as a whole, these new measures of interviewer workload at the interviewer level capture the active/assigned case workload on a typical fieldwork day. Overall, interviewers worked on average on a total of 25 and 28 sample units over the fieldwork period in Round 6 and Round 7, respectively. On a typical fieldwork day (i.e., according to the temporally weighted average) in either round, the interviewers had on average about five sample units in their ‘active’ case workload and about 11 in their ‘assigned’ case workload (Table 2).

Among the three operationalizations at the interviewer level, the two measuring *typical* workload are most strongly correlated ($r = .59$ in Round 6, $r = .64$ in Round 7). Total case workload is more strongly related to typical assigned case workload ($r = .50$ in Round 6, $r = .57$ in Round 7) than to typical *active* case workload ($r = .22$ in Round 6, $r = .36$ in Round 7).

4.3. Daily Field Efforts and Response Rates

We define the following interviewer field effort and performance measures at the level of interviewer-days. The measures are derived from the ESS contact history data.

Table 2. Descriptive statistics of project-specific workload at the interviewer level.

	ESS6-BE		ESS7-BE	
	Mean	SD	Mean	SD
Total case workload	25.09	13.49	28.11	17.56
Active case workload on a typical fieldwork day	5.20	1.85	5.05	1.94
Assigned case workload on a typical fieldwork day	11.45	3.26	11.54	3.95
<i>N</i>	155		151	

As previously noted, the analysis is restricted to interviewer-days with strictly positive active case workloads ($N = 7,702$ in Round 6 and $N = 9,259$ in Round 7). Some of the dependent variables are defined only on a smaller set of interviewer-days, as indicated below.

Contact activity observed for interviewer j on day t is a binary indicator that is equal to 1 if any personal visits are made by that interviewer on that day of fieldwork. Activity is observed on 35.34% ($N = 2,722$) of interviewer-days in Round 6 and 31.87% ($N = 2,951$) of interviewer days in Round 7.

The **contact effort** observed for (active) interviewer j on day t is the number of personal visits made by that interviewer, relative to his case workload (in tens of cases) on that day. We evaluate the relative number of visits only for interviewer-days on which there was any interviewer activity ($N = 2,722$ in Round 6 and $N = 2,951$ in Round 7). On the average (active) interviewer-day, 3.57 visits ($SD = 3.03$) were made in Round 6 and 3.61 visits ($SD = 3.27$) were made in Round 7. Taking the interviewers' daily case workloads into account, the contact effort on the average (active) interviewer-day was 3.02 visits per ten assigned cases in workload in both rounds ($SD = 2.30$ in Round 6, $SD = 2.49$ in Round 7).

The **contact rate** observed for interviewer j on day t is defined as the relative number of cases visited by interviewer j on day t with which contact with the target respondent was successfully made. The contact rate can thus only be evaluated for days on which at least one visit was made ($N = 2,722$ in Round 6 and $N = 2,951$ in Round 7). On the average (active) interviewer-day, contact was made with 55.44% ($SD = 38.75\%$) of visited target respondents in Round 6 and for 52.25% ($SD = 39.92\%$) of visited target respondents in Round 7.

The **conditional cooperation rate** observed for interviewer j on day t is defined as the relative number of target respondents contacted by interviewer j on day t from which an interview was successfully completed. The cooperation rate can thus only be evaluated for days on which at least one successful contact was made ($N = 2,177$ in Round 6 and $N = 2,220$ in Round 7). On the average (active) interviewer-day, an interview was completed for 52.86% ($SD = 44.62\%$) of contacted target respondents in Round 6 and for 47.29% ($SD = 44.86\%$) of contacted target respondents in Round 7.

4.4. Modelling Approach

For each outcome variable y_{jt} observed for interviewer j on day t , multilevel regression models (logistic for contact activity, linear for contact efforts, contact rates and cooperation rates) are estimated with interviewer-days nested within interviewers and one of the five measures of project-specific workload (at the level of interviewer-days or at the interviewer level), and its quadratic term, as the explanatory variables. The different project-specific workload operationalizations tested are briefly recapitulated in Table 3.

Model 0 includes a random intercept and fieldwork day control variables but none of the interviewer workload variables. This base model is used to estimate the share of the variability in daily field efforts and response outcome rates which may be attributed to

Table 3. Operationalizations of project-specific workload.

	At the level of interviewer days	At the interviewer level
Traditional operationalizations	–	Total active case workload = total number of sample units worked on over the fieldwork period as a whole
New operationalizations	Active case workload on day t = number of sample units for which interviewer activity has started and not yet ceased Assigned case workload on day t = number of sample units that have been and remain assigned	Active case workload on a typical fieldwork day = <i>average</i> number of sample units for which interviewer activity has started and not yet ceased Assigned case workload on a typical fieldwork day = <i>average</i> number of sample units that have been and remain assigned

systematic differences between the interviewers, and to evaluate the model fit of subsequent models.

$$y_{jt} = \gamma_{00} + \theta t + \Gamma \text{Day of the week}_t + u_{0j} + \varepsilon_{jt} \tag{0}$$
$$\theta(t) = \theta_1 t + \theta_2 t^2 + \theta_3 t^3 + \theta_4 t^4$$

γ_{00} represents the overall mean intercept, u_{0j} is the interviewer-specific deviation from the overall mean (with zero mean and variance σ_{u0}^2) and ε_{jt} the interviewer-day residual (with zero mean and variance σ_{ε}^2). $\theta(t)$ is a quartic function of fieldwork day t and **Day of the week_{*t*}** is a vector of six dummy variables indicating the day of the week (with Sunday as reference category). A quartic function of time suits the two-phased fieldwork and yields adequate model fits for all four outcome variables. For predicting daily contact rates and daily cooperation rates a linear function of fieldwork day and a cubic function of fieldwork day, respectively, would have sufficed. For these two field performance measures, adding the higher order polynomial terms does not improve model fit. Adding the higher order polynomial terms (up to four) does significantly improve model fit for predicting the probability of any contact activity and the contact efforts made.

Models 1 and 2 additionally include, respectively, a fixed (interviewer-level) measure of project-specific workload (and its quadratic term) and a time-varying (interviewer-day-level) measure of project-specific workload (and its quadratic term):

$$y_{jt} = \gamma_{00} + \beta_1 \text{Ln(Workload}_j) + \beta_2 \text{Ln(Workload}_j)^2 + \theta t \tag{1}$$
$$+ \Gamma \text{Day of the week}_t + u_{0j} + \varepsilon_{jt}$$

$$y_{jt} = \gamma_{00} + \beta_1' \text{Ln(Workload}_{jt}) + \beta_2' \text{Ln(Workload}_{jt})^2 + \theta t \tag{2}$$
$$+ \Gamma \text{Day of the week}_t + u_{0j} + \varepsilon_{jt}$$

where $Workload_j$ is the total case workload (Model 1A), the active case workload on a typical fieldwork day (Model 1B) or the assigned case workload on a typical fieldwork day (Model 1C) for interviewer j , and $Workload_{jt}$ is the active case workload (Model 2B) or the assigned case workload (Model 2C) for interviewer j on day t .

4.5. Robustness Checks

The standard multilevel model assumption of compound symmetry, that is, the different workload measurements are equally correlated, irrespective of their distance in time, is usually not realistic for repeated measures, but the fixed parameter estimates tend to be robust to small misspecifications of the random part of the model (Hox et al. 2018). The empirical autocorrelation functions for the within-interviewer residuals indicate that the residual covariance matrix across measurements (at least for the measurements of contact activity, contact efforts and contact rates) may be better characterized by a (first-order) autoregressive structure. Although explicitly specifying a more complex (autoregressive) within-interviewer correlation structure for these outcome variables does improve the model fit, the workload parameter estimates are not altered in a meaningful way (results not tabulated).

Controlling for fixed interviewer characteristics that may affect both assigned workloads and contact efforts and response outcome rates, prior experience (number of years working as a survey interviewer, in 5 categories) and having another job (binary indicator), does not meaningfully alter the workload parameter estimates either (results not tabulated).

As previously mentioned, the analytic sample contains only interviewer-days with strictly positive *active* case workload. Alternatively, the analytic sample could have been delineated on the basis of strictly positive *assigned* case workloads. Given that the analyses for contact effort, contact rates and cooperation rates further restrict the analytic sample to interviewer-days with any interviewer activity, the results for these outcomes are unaffected by this analytic choice. For the first outcome measure, the binary indicator of any interviewer activity, the analytic sample would have been somewhat larger if interviewer-days with zero active case workload (but non-zero assigned case workload) were also included. Using this extended analytic sample, the model fit relative to the base model (Model 0) is more pronounced for all workload operationalizations (the models with workload operationalized as a fixed interviewer characteristic as well as the models with workload operationalized as a time-varying interviewer characteristic), but the workload parameter estimates suggest a similar relationship between workload and interviewer activity over the relevant range (results not presented). As may be expected given the additional interviewer-days with zero active case workload, the estimated regression curve is somewhat steeper at the very lowest workload levels when workload is operationalized as the daily *active* case workload but is close to unaltered when workload is operationalized as the daily *assigned* case workload. The reported results are based on the more conservative delineation of the analytic sample.

5. Results

We start by estimating the share of the variability in daily field efforts and response outcome rates which may be attributed to systematic differences between the interviewers

(interviewer-level variance component) and to within-interviewer differences across fieldwork days (interviewer day residual variance component). Table 4 presents the variance components and intraclass correlation coefficients (ICCs) estimated from the base model (Model 0) for the four outcome variables. The ICCs suggest that only a very small share (daily contact activity and cooperation rates) to a moderately large share (daily contact efforts and contact rates) of the observed variability is due to systematic differences *between* interviewers. Much residual variance remains to be explained at the level of interviewer days.

Table 5 shows an overview of the differences in the Akaike information criterion (AIC) for the three models with project-specific workload operationalized as a fixed interviewer characteristic and the two models with project-specific workload operationalized as a time-varying interviewer characteristic, each compared to the base model (Model 0). We first consider the three fixed (interviewer-level) workload operationalizations. These operationalizations have very limited predictive power, and none consistently performs well across the outcome variables and for both survey rounds. For predicting daily contact activity, a marginal improvement in model fit compared to the base model (Model 0) is achieved by including any fixed workload operationalization in Round 7, but no improvement is achieved by including either fixed operationalization in Round 6. For predicting daily contact efforts, some improvement in model fit is achieved in both rounds by including total case workload or typical assigned case workload as fixed workload operationalizations. Typical assigned case workload appears to outperform total case workload as well as typical active case workload in terms of predictive power for this measure of field efforts. For predicting daily contact and cooperation rates, some improvement in model fit is achieved in both rounds by including typical active case workload and total case workload, respectively. Model fits for both outcome rates gain

Table 4. Estimated variance components for fieldwork effort and outcome variables.

	ESS6-BE	ESS7-BE
<i>Daily contact activity</i>		
Interviewer-level variance	0.1361***	0.1729***
ICC	0.0397	0.0499
<i>Daily contact effort</i>		
Interviewer-level variance	0.5415***	0.4897***
Residual variance	4.6477	5.3025
ICC	0.1044	0.0845
<i>Daily contact rate</i>		
Interviewer-level variance	0.0157***	0.0154***
Residual variance	0.1318	0.1409
ICC	0.1068	0.0986
<i>Daily cooperation rate</i>		
Interviewer-level variance	0.0041*	0.0039*
Residual variance	0.1907	0.1927
ICC	0.0209	0.0200

Note: Statistical significance of the random intercept is tested by a likelihood ratio test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5. Model fit statistics (ΔAIC) relative to base model (Model 0) for alternative operationalizations of project-specific workload.

	Daily contact activity (ref. = no personal visits made)		Daily contact effort		Daily contact rate		Daily cooperation rate	
	ESS6-BE	ESS7-BE	ESS6-BE	ESS7-BE	ESS6-BE	ESS7-BE	ESS6-BE	ESS7-BE
<i>Fixed interviewer characteristic models</i>								
Model 1A (Total case workload)	–	– 5.70	– 13.10	– 25.14	– 3.25	–	– 6.73	– 3.11
Model 1B (Active case workload on a typical fieldwork day)	–	– 3.10	–	–	– 15.14	– 12.16	–	– 4.39
Model 1C (Assigned case workload on a typical fieldwork day)	–	– 2.38	– 68.25	– 86.88	– 6.75	–	– 4.40	–
<i>Time-varying interviewer characteristic models</i>								
Model 2B (Daily active case workload)	– 646.79	– 706.80	– 325.12	– 162.76	– 52.48	–	– 108.93	– 68.47
Model 2C (Daily assigned case workload)	– 109.41	– 207.04	– 162.15	– 369.49	– 19.61	–	– 12.61	– 7.17
N	7,702	9,259	2,722	2,951	2,722	2,951	2,177	2,220

Note: Values for models with no statistically significant improvement in model fit compared to the base model (Model 0), as indicated by the likelihood ratio test ($p < .05$) are suppressed.

slightly from including any of the other fixed workload operationalization in one of the two survey rounds, but not in the other.

The predictive power of the two time-varying workload operationalizations is much stronger. At least for predicting daily contact activity, contact efforts and cooperation rates, the models with the time-varying workload operationalizations yield considerable improvements in model fit compared to the base model (Model 0) in both rounds. Only for predicting daily contact rates, the time-varying workload operationalizations are not much more convincing than the fixed (interviewer-level) workload operationalizations. Some improvement in model fit for this outcome variable is achieved by including either time-varying workload operationalization in Round 6, but no improvement is achieved by either operationalization in Round 7. Of the two time-varying operationalizations, daily *active* case workload frequently exceeds daily *assigned* case workload in terms of predictive power.

The inferred relationships between project-specific workload, operationalized in a time-varying manner (daily active case workload and daily assigned case workload), and the four outcome variables are discussed in the following subsections. Because the inclusion of the quadratic term complicates interpreting the parameter estimates directly, we present prediction plots for each outcome variable (Figures 4 to 7) for one hypothetical fieldwork

Fig. 4. Predicted probability of contact activity as a function of time-varying project-specific workload, evaluated on a Monday, the 50th fieldwork day.

Fig. 5. Predicted contact effort per case in workload as a function of time-varying project-specific workload, evaluated on a Monday, the 50th fieldwork day.

day. These prediction plots demonstrate the dominant direction and curvature of the regression curve. The choice of fieldwork day presented (Monday, the 50th fieldwork day) is arbitrary but, in the absence of any interaction effects between fieldwork day and project-specific workload, only affects the fieldwork-day-specific intercept, not the shape of the curve. In order to avoid extrapolations beyond the normal workload ranges, the predicted field efforts and response outcome rates are presented over the range defined by the 5% and 95% workload quantiles (3 to 22 cases in assigned case workload, 1 to 12 cases in active case workload).

5.1. Contact Activity

The probability of contact activity on a given day tends to increase with project-specific workload (Figure 4). The results for the two survey rounds are very similar. The predicted probability of any interviewer activity on a given day is about 10 percentage points higher when an interviewer has an assigned case workload of 10 cases rather than 5, and again 5 to 6 percentage points higher when an interviewer has an assigned case workload of 15 cases rather than 10. The relationship is stronger for daily *active* case workload than for daily *assigned* case workload but, as previously noted, the strength of the association may

Fig. 6. Predicted daily contact rate as a function of time-varying project-specific workload, evaluated on a Monday, the 50th fieldwork day.

be inflated by active case workloads increasing as more sample units are actually being attempted. The predicted probability of any interviewer activity on a given day is 17 to 18 percentage points higher when an interviewer has an active case workload of 10 cases rather than 5.

5.2. Contact Effort

Contact efforts tend to decrease with daily *assigned* case workload (Figure 5). The relationship between project-specific workload and contact effort is in the opposite direction for the daily *active* case workload operationalization. As for predicting contact activity, the positive association for daily active case workload may well be an artefact of this particular workload operationalization. In both survey rounds, the negative relationship between project-specific workload (assigned case workload operationalization) and contact effort is strongest at very low workload levels but more gradual for higher workload levels. Beyond an assigned case workload of about ten cases the negative relationship flattens out almost completely and contact efforts per case remain fairly constant across increasingly large workloads.

5.3. Contact Rate

Contact rates are inconsistently, if at all, related to project-specific workload (Figure 6). The parameter estimates of the daily *assigned* case workload model suggest a weak negative association in Round 6 but a weak positive association in Round 7. The estimates of the daily *active* case workload model suggest a weak, approximately linear negative association in Round 6, but no association in Round 7. Altogether these results indicate that daily workload and contact rates are not really related.

5.4. Cooperation Rate

Cooperation rates tend to decrease with project-specific workload (Figure 7). The results for the two survey rounds are again fairly similar. Over the relevant range of daily workload levels, the negative relationship is approximately linear, and much steeper for daily *active* case workload than for daily *assigned* case workload. The predicted cooperation rate on a given day is 5 to 6 percentage points lower when an interviewer has an assigned case workload of 10 cases rather than 5, and again 2 to 3 percentage points lower when an interviewer has an assigned case workload of 15 cases rather than 10. The predicted cooperation rate on a given day is 10 to 14 percentage points lower when an

Fig. 7. Predicted daily cooperation rate as a function of time-varying project-specific workload, evaluated on a Monday, the 50th fieldwork day.

interviewer has an *active* case workload of 10 cases rather than 5. Although the negative relationship between project-specific workload (assigned case workload operationalization) and cooperation rates is strongest at very low workload levels and much more gradual for higher workload levels, even these apparently small differences are far from negligible in view of commonly observed cooperation rates.

6. Conclusions and Discussion

Interviewer characteristics are usually assumed fixed over the fieldwork period. In this article we show that since interviewers' project-specific workloads can vary strongly over the fieldwork period, a more fine-grained, time-varying operationalization can be useful to explain differences in interviewers' field efforts and performance. We proposed two approaches to measure project-specific workload for interviewers on each day in the fieldwork: (1) the number of sample units which have been and remain assigned on any day t (assigned case workload), and (2) the number of sample units for which interviewer activity has started and not yet ceased on any day t (active case workload). We have used these operationalizations in an examination of workload correlates in two rounds of the European Social Survey in Belgium (results not presented).

The results indicate that none of the operationalizations at the interviewer level, whether the traditional total-count workload measure or the proposed operationalizations of *typical* (active or assigned) case workload, consistently help to explain the observed systematic differences in efforts and response outcome rates between interviewers in the contact and recruitment task. The remaining variability in task efforts and response outcome rates within interviewers, across fieldwork days, is not completely random. This variability can partially be explained by differences in interviewers' workloads at this level, across fieldwork days. Capturing temporal workload variation directly, the time-varying operationalizations, daily active case workload and daily assigned case workload, fairly consistently outperform the interviewer-level *typical* workload operationalizations that are derived from them, as well as the traditional total-count workload operationalization. The results are remarkably similar for the two survey rounds studied, supporting their external validity.

Although the results show that *active* case workload frequently has a better predictive power than *assigned* case workload when predicting interviewers' efforts and performance in the contact and recruitment task, the positive associations with contact activity and efforts observed for the active case workload operationalization may be biased by reverse causality. An interviewer's field efforts to a large extent drive the number of active cases he or she has. Active case workload itself would thus by construction be positively related to contact activity and efforts. *Assigned* case workload is much less sensitive to variations in interviewers' field efforts over the fieldwork period and can thus be regarded as the conceptually more adequate operationalization of project-specific workload. Unfortunately, the assigned case workload operationalization requires additional case assignment paradata where contact history paradata suffices for the *active* case workload operationalization. Case assignment paradata may be even less readily available than contact history paradata. Without access to additional case assignment data, it may be possible to derive an approximate assigned case workload measure from contact history

data by identifying groups of sample units with overlapping recorded interviewer activity, suggestive of simultaneous assignment. Our attempt to derive such an approximation has not altogether been successful for these two survey rounds of the European Social Survey in Belgium.

Further research on the impact of interviewer workload in particular, and research aimed at unpacking the fieldwork process more generally, would benefit from contact history paradata being supplemented with case assignment and conversion paradata. Timestamped assignments to interviewers, returns to field offices, and transitions in incentive schemes, recruitment modes and other features of the data collection that may be dynamically altered over the course of the fieldwork, would allow investigations into aspects of the fieldwork process that are not captured by contact history paradata and currently remain obscure in many settings. Interviewer workload patterns over the fieldwork period is one such aspect. Other aspects of the fieldwork process into which valuable insights may be gained are the progression of sample units being initially issued, the application of different reassignment and conversion strategies, and the extent to which sample units are (temporarily) shelved by field supervisors because they are costly to pursue or interviewers are (temporarily) unavailable, or by interviewers because they are remote or located in disadvantaged areas.

Although research on the fieldwork process would benefit from case assignment and conversion paradata supplementing contact history paradata, simply collecting more paradata is a poor strategy if adequate paradata quality cannot be maintained. The proposed approach of measuring workload more precisely presumes higher accuracy of the underlying (contact history and case assignment) paradata than is required for the traditional-total count measure. Comprehensive quality assessments of this type of paradata are rarely published (West and Sinibaldi 2013), but some studies have demonstrated that (unsuccessful) contact attempts, especially attempts that may not be unambiguously considered proper attempts by the interviewers (e.g., ‘drive-by’ visits, Biemer et al. 2013; Wagner et al. 2017), are commonly underreported. We observed small numbers of cases for which the recorded date of the first contact attempt preceded the recorded date of assignment, indicative of entry errors in the contact history paradata and/or the assignment paradata. Especially when contact histories are not recorded by the interviewers in a timely manner, we may worry about their accuracy. The quality of case assignment paradata may likewise suffer from underreporting and inaccurate recording by field managers. As for other types of paradata, quality assurance and control procedures should safeguard the quality of contact history and case assignment paradata.

The results related to the secondary study objective, to apply the different workload operationalizations in a further exploration of the interviewer workload-performance relationship, are tentative, but emphasize some interesting mechanisms that should invite further research. The performance of face-to-face survey interviewers in the contact and recruitment task does appear to be affected by the number of sample units that require their attention over time. However, these workload effects are not unidirectional and straightforward.

The hypothesis that when interviewers carry larger workloads their contact effort per case is reduced (Hypothesis 1) is only partially supported. The initial evidence suggests that when interviewers carry heavier workloads they tend to be *more* likely to make any

personal visits at all. This positive result suggests that interviewers are more strongly motivated to make a round trip of visits when they are carrying larger workloads than when they are carrying small workloads. When sample units are geographically clustered, the fixed cost of travel to the cluster may only be worth the investment if many can be visited on the same round trip. The result substantiates the comments from several of the interviewers in the study (collected via a post-fieldwork interviewer questionnaire) that assignment sizes are *too small* rather than too large. To which extent interviewers may experience larger workloads as a greater potential income (Wuyts and Loosveldt 2016), rather than as a burden, is an important question to answer if the mechanisms through which interviewer workload affect fieldwork outcomes are to be well understood. Whereas larger workloads were related to a higher probability of any interviewer activity, they were also related to smaller numbers of personal visits made relative to the number of cases in workload. Interviewers may be induced to spread their contact efforts over more working days, rather than reduce their contact efforts, when they have more cases to work on. The negative association is most pronounced at lower workload levels, and almost completely flattens out for moderate to large workloads. The observed compression of the negative effect of workload on contact effort challenges the presumed dominant role of time availability constraints in the contact and recruitment task.

The hypothesis that when interviewers carry larger workloads their contact and cooperation rates are reduced (Hypothesis 2) is also only partially supported. Overall, the link between interviewer workload and contact efforts is stronger than the link between workload and outcome rates. This may be explained by efforts being completely within the survey interviewers' control. That contact rates were not consistently related to project-specific workload suggests that contact schedules are no less productive when interviewers carry larger workloads than when they have only small numbers of cases to work on. It is possible that time constraints do not severely limit the number of visits during the more productive hours. Alternatively, the number of these visits may be considerably limited but interviewers may take likely accessible at home patterns and contact histories of different target respondents into account when planning their visits, for example by dedicating evening visits to target respondents that are of working age and/or previously could not be contacted during the day. In this sense, large workloads may encourage interviewers to optimize their contact schedule within the given time constraints, offsetting the direct negative effects of these time constraints. That cooperation rates, but not contact rates, were consistently negatively related to interviewer workload suggests that when interviewers carry larger workloads they interact with contacted target respondents in a different, less (immediately) successful way at the doorstep, possibly more easily accepting reluctance or postponement. The negative association, however, is most pronounced at lower workload levels, and flattens out for moderate to large workloads.

A common limitation of any operationalization of interviewer workload in terms of case counts (whether in the aggregate or time-varying) is that assignments are assumed equivalent in the amount of work they imply. For many reasons the amount of work may vary across sample units, and the amount of work for a single sample unit need not even be constant over time. A first obvious reason is that not all sample units are equally accessible and cooperative. A second reason in the context of face-to-face survey data collection is that travel distance to and among assignment clusters (a function of the geographical

distribution of sample units relative to the geographical distribution of interviewers) may vary considerably and is also likely to weigh heavily on interviewers' experienced workloads. Marginal travel costs may also shift as different (clusters of) sample units move in and out of interviewers' workloads. The recent study of interviewer travel behavior by [Wagner and Olson \(2018\)](#) is worth mentioning in this context. They examine administrative travel data at the level of interviewer days and observe that interviewer travel between clusters is a relevant factor to both field outcomes and fieldwork costs.

A plausible implication is that the amount of work for one sample unit is larger when the number of assigned sample units is small (and mostly dispersed, hard-to-contact and uncooperative sample units remain). This would weaken our conclusions for the workload effects at low workload levels. We previously noted that the observed negative association of assigned case workload with the number of personal visits made relative to the number of cases in workload, and with the daily cooperation rate, tends to flatten out for moderate to large workload levels. The observed negative associations at low workload levels may to some extent be driven by the particular composition of the remaining workload.

A second limitation of this study is the limited range of observed interviewer workloads. We observe an average assigned case workload of 12 sample units across interviewer-days, and an assigned case workload of only up to 22 cases for 95% of interviewer-days. A plausible explanation for why we observe mostly moderate workload levels lies in the batch allocation strategy that was used in the two survey rounds studied, against the backdrop of the explicit workload restrictions advocated by the standards and specifications of the European Social Survey. In the European Social Survey, a maximum total of 48 sample units is in principle allowed to be assigned per interviewer over the course of the fieldwork, including any reassignments as well as initial assignments ([Stoop et al. 2010](#)). If workload only has strong harmful effects on interviewers' fieldwork efforts and outcomes at higher levels, these cannot be observed. Excessive workloads that may negatively affect interviewers' performance may thus have been systematically avoided by design ([Blom et al. 2011](#)).

One should also keep in mind that the question about reasonable interviewer workloads is not only relevant to the fieldwork process and nonresponse error, but also (and possibly even more so) for measurement error. The extent to which interviewers' individual systematic effects on responses to survey questions result in inflated standard errors of survey estimates depends on the total number of respondents interviewed by each interviewer ([Kish 1962](#); [Beullens et al. 2016](#)).

In summary, when interviewers carry heavier project-specific workloads, we observe (1) a higher probability of making any visits at all, but, up to a certain point, a reduced number of visits if any are made, and (2) a similar probability of attempted target respondents being successfully contacted, but, again up to a point, a reduced probability of contacted target respondents actually participating.

Overall, our results warrant no serious concern (at moderate workload levels) but some caution about possible harmful workload effects on nonresponse error. Given that interviewers' project-specific workload is a factor that can be controlled by survey designers and field supervisors, further studies on the risk of workload effects are called for. Larger workload levels than studied here deserve particular attention. On the other hand, the observed association with the likelihood of making any personal visits at all

encourages further investigation into minimum efficient workloads and the possible benefits and risks of continuous replenishment of workloads to a manageable level. There may be motivational mechanisms driving task performance up with increasing workload (Delasay et al. 2019). In particular at low workload levels, increasing workloads may be experienced by engaged and responsive interviewers as a greater challenge and a more worthwhile investment of time and effort. An inverted-U pattern may emerge from different workload mechanisms (with effects in different directions) dominating over low-versus high-workload ranges. Such a pattern has been commonly hypothesized for task performance as a function of subjective workload measures and stress (Westman and Eden 1996; Muse et al. 2013) and has been observed using objective workload measures in field studies in the restaurant and grocery industry (Tan and Netessine 2014; Bruggen 2015).

Our results for contact efforts and contact rates do not unambiguously support the commonly held belief that larger workloads allow interviewers less time and effort to expend for individual sample units, and therefore necessarily harm response rates. Interviewers are likely to adjust their contact activities on the basis of a careful (though not necessarily deliberate) balancing exercise that depends on preferences about work time organization and anticipated benefits and costs of additional efforts, as well as overall time availability. The result for cooperation rates highlights that not only interviewers' contact activities, but also interviewers' behavior at the doorstep can be affected.

7. References

- Beullens, K., G. Loosveldt, K. Denies, and C. Vandenplas. 2016. "Quality Matrix for the European Social Survey, Round 7." Leuven, Belgium: Centre for Sociological Research, KU Leuven.
- Biemer, P.P., P. Chen, and K. Wang. 2013. "Using Level-of-Effort Paradata in Non-Response Adjustments with Application to Field Surveys." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176(1): 147–168. DOI: <https://doi.org/10.1111/j.1467-985X.2012.01058.x>.
- Blom, A.G. 2012. "Explaining Cross-Country Differences in Survey Contact Rates: Application of Decomposition Methods: Cross-Country Differences in Survey Contact Rates." *Journal of the Royal Statistical Society, Statistics in Society* 175(1): 217–242. DOI: <https://doi.org/10.1111/j.1467-985X.2011.01006.x>.
- Blom, A.G., E.D. de Leeuw, and J.J. Hox. 2011. "Interviewer Effects on Nonresponse in the European Social Survey." *Journal of Official Statistics* 27 (2): 359–377. Available at: <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/interviewer-effects-on-nonresponse-in-the-european-social-survey.pdf> (accessed April 2020).
- Botman, S.L. and O.T. Thornberry. 1992. "Survey Design Features Correlates of Nonresponse." In *Proceedings of the Survey Research Methods Section: American Statistical Association*, August 9–13, 1992. 309–314. Alexandria, VA: American Statistical Association. Available at: http://www.asasrms.org/Proceedings/papers/1992_048.pdf (accessed April 2016).
- Bruggen, A. 2015. "An Empirical Investigation of the Relationship between Workload and Performance." *Management Decision* 53(10): 2377–2389. DOI: <https://doi.org/10.1108/MD-02-2015-0063>.

- Calderwood, L., I. Plewis, S. Ketende, and T. Mostafa. 2016. "Evaluating the Immediate and Longer Term Impact of a Refusal Conversion Strategy in a Large Scale Longitudinal Study." *Survey Research Methods* 10(3): 225–236. DOI: <https://doi.org/10.18148/srm/2016.v10i3.6275>.
- Delasay, M., A. Ingolfsson, B. Kolfal, and K. Schultz. 2019. "Load Effect on Service Times." *European Journal of Operational Research* 279(3): 673–686. DOI: <https://doi.org/10.1016/j.ejor.2018.12.028>.
- European Social Survey (2014). *ESS Round 6 Data from Contact forms, edition 2.0*. Bergen, Norway. Norwegian Social Science Data Services, Norway—Data Archive and distributor of ESS data.
- European Social Survey (2016). *ESS Round 7 Data from Contact forms, edition 2.1*. Bergen, Norway. Norwegian Social Science Data Services, Norway—Data Archive and distributor of ESS data.
- Groves, R.M. and M. Couper. 1998. "How Survey Design Features Affect Participation." In *Nonresponse in Household Interview Surveys*, 269–293. New York, NY: John Wiley & Sons.
- Hox, J.J., M. Moerbeek, and R. van de Schoot. 2018. *Multilevel Analysis: Techniques and Applications*. Third edition. Quantitative Methodology Series. New York, NY: Routledge.
- Kirchner, A. and K. Olson. 2017. "Examining Changes of Interview Length over the Course of the Field Period." *Journal of Survey Statistics and Methodology* 5: 84–108. DOI: <https://doi.org/10.1093/jssam/smw031>.
- Japac, L. 2005. "The Concept of Interviewer Burden." Presented at the International Workshop on Household Survey Nonresponse, Tällberg, Sweden.
- Japac, L. 2008. "Interviewer Error and Interviewer Burden." In *Advances in Telephone Survey Methodology*, edited by J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. de Leeuw, L. Japac, P.J. Lavrakas, M.W. Link, and R.L. Sangster, 187–211. New York, NY: John Wiley & Sons.
- Kish, L. 1962. "Studies of Interviewer Variance for Attitudinal Variables." *Journal of the American Statistical Association* 57(297): 92–115. DOI: <https://doi.org/10.2307/2282442>.
- Loosveldt, G. and K. Beullens. 2013. "'How Long Will It Take?' An Analysis of Interview Length in the Fifth Round of the European Social Survey." *Survey Research Methods* 7(2): 69–78. DOI: <https://doi.org/10.18148/srm/2013.v7i2.5086>.
- Loosveldt, G., A. Carton, and J. Billiet. 2004. "Assessment of Survey Data Quality: A Pragmatic Approach Focused on Interviewer Tasks." *International Journal of Market Research* 46(1): 65–82. DOI: <https://doi.org/10.1177/147078530404600101>.
- Muse, L.A., S.G. Harris, and H.S. Feild. 2003. "Has the Inverted-U Theory of Stress and Job Performance Had a Fair Test?" *Human Performance* 16(4): 349–364. DOI: https://doi.org/10.1207/S15327043HUP1604_2.
- Nicoletti, C. and N.N. Buck. 2004. "Explaining Interviewee Contact and Co-Operation in the British and German Household Panels." 2004–2006. Working Papers of the Institute for Social and Economic Research. Colchester, United Kingdom: University of Essex.

- Olson, K. and A. Peytchev. 2007. "Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes." *Public Opinion Quarterly* 71(2): 273–286. DOI: <https://doi.org/10.1093/poq/nfm007>.
- Singer, E., M.R. Frankel, and M.B. Glassman. 1983. "The Effect of Interviewer Characteristics and Expectations on Response." *Public Opinion Quarterly* 47(1): 68–83. DOI: <https://doi.org/10.1086/268767>.
- Stoop, I., J. Billiet, A. Koch, and R. Fitzgerald. 2010. *Improving Survey Response: Lessons Learned from the European Social Survey*. Hoboken, NJ: John Wiley & Sons.
- Tan, T.F. and S. Netessine. 2014. "When Does the Devil Make Work? An Empirical Study of the Impact of Workload on Worker Productivity." *Management Science* 60(6): 1574–1593. DOI: <https://doi.org/10.1287/mnsc.2014.1950>.
- Wagner, J., K. Olson, and M. Edgar. 2017. "The Utility of GPS Data in Assessing Interviewer Travel Behavior and Errors in Level-of-Effort Paradata." *Survey Research Methods* 11(3): 219–233. DOI: <https://doi.org/10.18148/srm/2017.v11i3.6794>.
- Wagner, J. and K. Olson. 2018. "An Analysis of Interviewer Travel and Field Outcomes in Two Field Surveys." *Journal of Official Statistics* 34(1): 211–237. DOI: <https://doi.org/10.1515/jos-2018-0010>.
- Watson, N. and M. Wooden. 2009. "Identifying Factors Affecting Longitudinal Survey Response." In *Methodology of Longitudinal Surveys*, edited by P. Lynn. Chichester, United Kingdom: John Wiley & Sons.
- West, B.T. and A.G. Blom. 2016. "Explaining Interviewer Effects: A Research Synthesis." *Journal of Survey Statistics and Methodology* 5(2): 175–211. DOI: <https://doi.org/10.1093/jssam/smw024>.
- West, B.T. and J. Sinibaldi. 2013. "The Quality of Paradata: A Literature Review." In *Improving Surveys with Paradata: Analytic Uses of Process Information*, edited by F. Kreuter. Hoboken, NJ: John Wiley & Sons.
- Westman, M. and D. Eden. 1996. "The Inverted-U Relationship between Stress and Performance: A Field Study." *Work & Stress* 10(2): 165–173. DOI: <https://doi.org/10.1080/02678379608256795>.
- Wuyts, C. and G. Loosveldt. 2016. "Workload-Related Interviewer Characteristics and Unit Nonresponse in ESS Belgium." Presented at the International Workshop on Household Survey Nonresponse, Oslo, Norway.

Received September 2018

Revised June 2019

Accepted September 2019