Nonresponse Workshop 2015

Better or Longer? The evolution of weekly number of completed interviews over the fieldwork period in the European Social Survey

Caroline Vandenplas, Geert Loosveldt, Koen Beullens, KU Leuven

Introduction

Extending the fieldwork period by a few weeks increases the number of completed interviews, unless no interview can be realized during the extension. It is however often the case that surveys with longer fieldworks obtain lower response rates, or in other words struggle to achieve a high number of completed interviews, i.e. the minimum number of completed interviews often required in cross-countries surveys. To understand this paradox, we use a principle of classical mechanics in physics:

$$Work = Power * Time.$$

In a survey methodology context, if we consider the fieldwork as a machine designed to produce completed interviews (or returned questionnaires), we can equate 'work' to the achieved number of completed interviews at the end of the fieldwork. The 'time' becomes then the fieldwork period, counted in number of weeks[1], and the 'power' becomes the potential of the fieldwork machine per time unit, or in other words the weekly number of completed interviews. Of course, the produced work and the weekly power depend on the input that is given to the machine. In the case of the fieldwork, the number of sample units that have to be surveyed or processed is equivalent to such an input. As we want to compare different surveys, we will standardize the input to 100. From now on, 'work' will refer to a standardized total number of completed interviews, equal to the total number of completed interviews at the end of the fieldwork divided by the sample size and multiply by 100. 'Power' will be similarly defined as the standardized weekly number of completed interviews, equal to the weekly number of completed interviews divided by the sample size and multiply by 100. This gives the following translation of the principle of classical mechanics in survey methodology:

Standardized final nbr of completed interviews
*= Standardized weekly nbr of completed interviews*nbr of fieldwork weeks.*

Using this principle, it becomes clear that, if the power is kept constant, extending the fieldwork must lead to more produced 'work', i.e. standardized total number of completed interviews. Hence, if in general surveys with a longer fieldwork have lower 'work', the 'power' of these surveys must be lower, in other words the (mean) standardized number of weekly completed interviews must be smaller for these surveys.

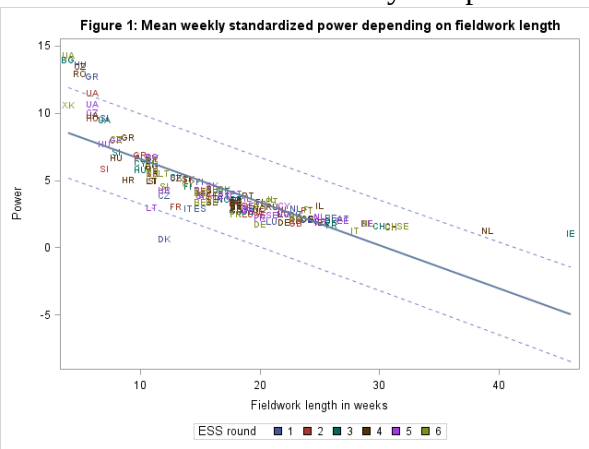
Relation between fieldwork 'power' and fieldwork length in the European Social Survey

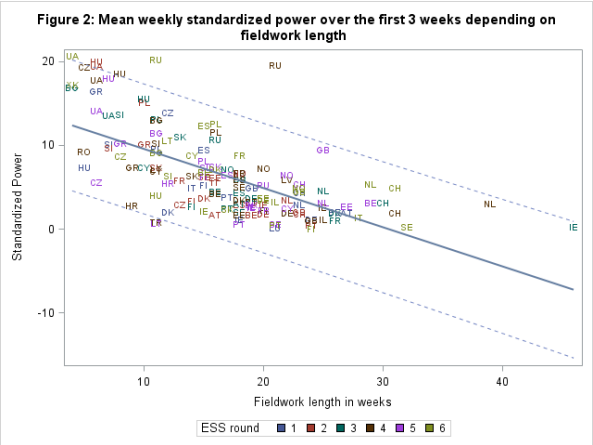
To illustrate the expected relation between the fieldwork length and the fieldwork power, we use the first six rounds of the European Social Survey (ESS). We consider each survey as an unit. The term 'survey' as analyzed unit will designate a combination of an ESS round and a country that participated in this round. In total, there are 149 surveys. To be able to analyze

[1] Note that we here chose 'week' as time unit rather than 'day'. The reasons is to avoid dealing with the difference between weekdays and weekends and such that the multi-level model applied further converges.

such a number of similar surveys (topic, guidelines, etc.) is quite a unique opportunity. For each ESS survey, during the fieldwork period, interviewers are asked to complete a contact form for each sample unit that captures information about each contact attempt: date and time of contact, outcome (contact/non-contact/refusal/ ineligible/appointment/interview, etc.), assigned interviewer, interviewer observation, mode of contact, refusal conversion information, if applicable. These contact forms allow us to calculate the fieldwork length in weeks but also the fieldwork power. Indeed, for each fieldwork week, we can count how many interviews have been completed in that week, take the mean number of completed interviews per week over the fieldwork period and standardize it. So for each survey, we calculated the mean standardized power.

The relation between fieldwork power and fieldwork length for all ESS surveys is illustrated in figure 1. Per extra fieldwork week, the 'power' decreases by -0,32. Surveys that have longer fieldwork hence produce a lower number of completed interviews per 100 sampling units per week. This could be due to a "tail" effect: in case of longer fieldworks, the last few weeks, when very few interviews are completed, could drive the mean of the standardized number of weekly completed interviews down. However, similar results are obtained if the power is calculated as the mean of the standardized number of weekly completed interviews over the 3 first weeks of the fieldwork only. The fieldwork power decreases with fieldwork length, -0,47 in standardized number of weekly completed interviews per week increase. We can thus conclude that surveys for which the fieldwork last longer also produce a lower standardized number of weekly completed interviews at the start of the fieldwork.





Distribution of the weekly power over the fieldwork length

Extending the fieldwork period does not seem to be a very fruitful strategy to obtain more completed interviews. As indicated in the basic model, the other factor influencing the fieldwork work is the power. We need to understand this other factor better. By doing so, we may be able to discover which fieldwork characteristics can increase the final (standardized) number of completed interviews (fieldwork work).

So far we have made the very naïve assumption that the power remains constant over the fieldwork period. This is rarely the case, even in physics. Looking at the evolution of the power over the fieldwork in all surveys of the sixth round of the ESS (figure 3), two type of scenarios seem possible. First scenario: the power is high in the first week and then decreases. The decrease of the power first get larger (steeper curve) to then get smaller (flatter curve) to form a tail (Belgium, Cyprus, Germany, Denmark, Great Britain, Lithuania, Norway, Russia). Second scenario: the power first increases at the beginning of the

fieldwork to reach a maximum and finally decrease. Once the power starts to decrease, the loss in power first gets larger to then get smaller and form a tail. In figure 4, the shapes of both scenarios are presented.

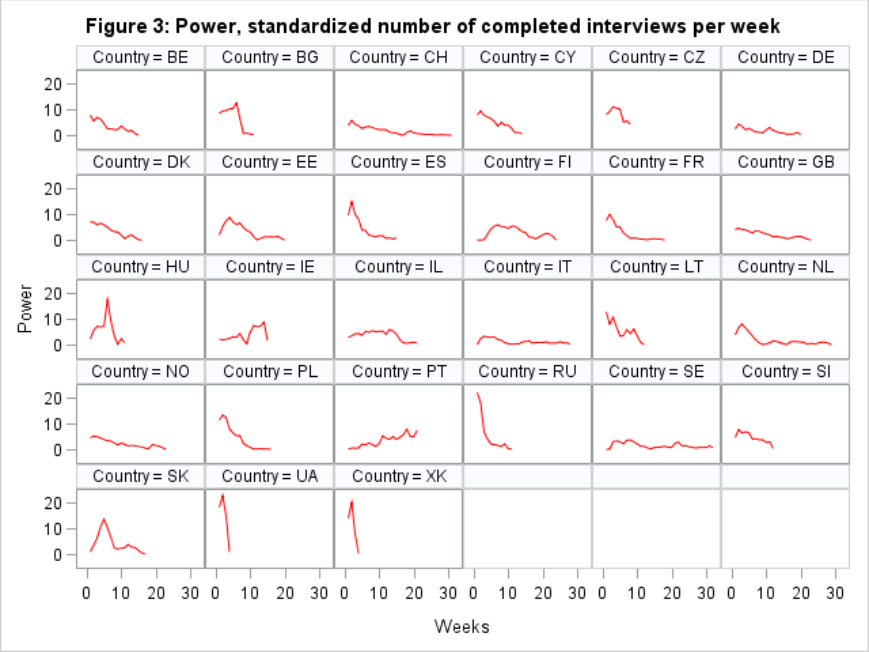
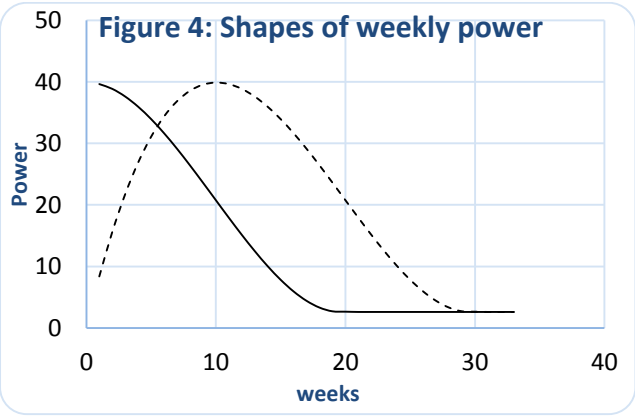


Figure 4 illustrate scenario 1 (full line) and scenario 2 (dash line).



The shapes of the weekly power make clear that it is not constant over time. In this situation, we can define the work as the area underneath the curve, or in a mathematical formula:

$$Work = \int_{t=0}^{t=end} Power(t)dt$$

Expressed in the context of survey fieldwork, the standardized number of completed interviews at the end of the fieldwork is the sum of the standardized number of completed interviews that have been realized in each week over the fieldwork period:

$$Standardized\ final\ nbr\ of\ completed\ interviews$$

$$= \sum_{w=1}^{last\ week\ of\ fieldwork} Standardized\ nbr\ of\ completed\ interviews\ in\ week\ w.$$

The model of the standardized weekly number of completed interviews

To confirm and explore our intuition about the shape of the evolution of the fieldwork power, we apply a repeated measurement multi-level model on the data, with the surveys (country*round) as level 1 units and the weekly power as repeated measurements. To mimic the shapes in figure 4, we need a cubic expression. Obviously, the relation between the fieldwork weeks and the power is not linear. A quadratic expression could maybe estimate the first scenario, starting high and then a slowing down decrease that forms tail, but could not simulate an eventual acceleration in the decrease at the beginning of the fieldwork before the slowing down or the second scenario. We are actually interested in 4 characteristics of the fieldwork: the first week power (starting point=intercept), the decrease or increase in power in the first week (β_1), the acceleration or deceleration of the decrease (scenario 1) or increase (scenario 2) in power at the start (β_2) and the start of the tail (- $\beta_2/3\beta_3$).

We apply a repeated measurement multi-level model on the data, with the surveys (country*round) as level 1 units, and the weekly power as repeated measurements. For the model to converge, we limit the analyses to 35 weeks[2]. The studied model is the following where *s* denotes the different surveys and *w* the fieldwork weeks (w=0 for week 1),:

$$P(s,w) = \beta_0 + \beta_1 w + \beta_2 w^2 + \beta_3 w^3 + \varepsilon_{s,w}$$
$$\beta_0 = \gamma_{00} + u_0$$
$$\beta_1 = \gamma_{10} + u_1$$
$$\beta_2 = \gamma_{20} + u_2$$
$$\beta_3 = \gamma_{30}$$

With $\varepsilon_{s,w} \sim N(0, \sigma^2)$ and level 2 covariance matrix $\begin{bmatrix} \sigma_0^2 & \sigma_{0,1} & \sigma_{0,2} \\ \sigma_{0,1} & \sigma_1^2 & \sigma_{1,2} \\ \sigma_{0,2} & \sigma_{1,2} & \sigma_2^2 \end{bmatrix}$. Table 1gives the estimates for the fixed effects ($\gamma's$). Note that we chose for the intercept (β_0), the linear (β_1) and quadratic (β_2) coefficients to be random to cover the possible scenarios. The first week power cannot be expected to be the same for every survey, therefore the intercept (β_0) is a random effect. The linear coefficient (β_1) makes the difference between scenario 1 (negative) and scenario 2 (positive). The quadratic coefficient (β_2) determines whether the power decreases/increases with an acceleration or deceleration at the start. The cubic coefficient (β_3) has been kept fixed to keep the model as simple as possible.

Table 1: Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
γ_{00}	7,530	0,541	147,000	1,392	<,0001
γ_{10}	-0,503	0,124	147,000	-4,060	<,0001
γ_{20}	-0,006	0,007	147,000	-0,790	0,432
γ_{30}	0,001	0,000	2.032,000	3,890	0,0001

The value of the overall intercept γ_{00} means that in the first week, taking the mean over all surveys, 7,5 interviews from the 100 sampling units are or in other words 7,5% of the total sample has been converted to completed interviews. The negative sign of the mean value of the linear coefficients points to a decrease of the power in the first week of the fieldwork period. The negative sign of the mean value of the quadratic coefficients (non-significant) means that at the start of the fieldwork, the decrease in power first accelerate to slow down later. Finally, the cubic coefficient is significantly positive. We can determine when the tail

[2] Only Ireland in round 6 (46 weeks) and the Netherlands in round 4 (39 weeks) exceeded 35 weeks of fieldwork.

starts (when the decrease in power stops accelerating and starts slowing down) by combining both the cubic and quadratic coefficient, namely at week 2 (0,006/(3*0,001))**.** Under this model, we can also calculate that 50% of the sample is converted in completed interviews after 10 weeks.
The covariance parameters are all significantly (alpha <0,001) different from 0, showing that surveys differ a lot as well at the start of the fieldwork as in the way they evolve during the fieldwork.(table not shown).

Fieldwork characteristics influencing the evolution of the fieldwork power.

If the standardized number of completed interviews at the end of the fieldwork is equal to the area under the curve in figure 4, we want to maximize this area, in other words, we want the curve to be as high as possible (high power) and as broad as possible. This means that we want to keep the power as high as possible for as long as possible. Of course, there are limits to the possible height and width of the curve. Constrains such as budget, number of data that the fieldwork agency can process in a week, the expected total number of realized interviews or practical consideration such as interviewers travel or even the number of released addresses put an upper bound to the height and width of the fieldwork power curve. Our hypotheses are the following:

- The height of the curve: We expect that, for each week, the power heavily depends on the number (also standardized per 100 sampling unit) of interviewers that are active that week. We also assume that the survey climate as measured by the percentage of completed interviews after one contact attempt influence the power of the first week, and hence the intercept, as well as the decrease of the power at the start. Finally the type of frame is expected to have an effect on the standardized number of completed interviews at the start of the fieldwork period, interviewers needing to first perform the second stage if no individual sampling frame is available.
- The width of the curve: The rapidity with which a maximum weekly power is reached (scenario 2) or the decrease in weekly power grows (scenario 1) depends on careful timing of contact attempts and the percentage of refusal conversion, such strategy can keep the power high or prevent it to drop quickly. To give a measure of how well-timed the contact attempts are, a contact score and refusal conversion score were constructed for each survey. The percentage of non-contacts that have been visited four times, once in the evening, once at the weekend and for which the attempts were spread over two weeks[3] are calculated and added up, for a maximum of 400 points, standardized by dividing by 4. A refusal procedure score was also computed. All sample units that at least refused once, called initial refusers, were considered. The percentages of initial refusers for which at least two interviewers were assigned over the whole fieldwork period, the contact attempt following the initial refusal was at a different time of the day (morning, afternoon, evening), the contact attempt following the initial refusal was another day of the week, the contact attempt following the refusal conversion happened after a 'cooling' period of at least 7 days were calculated. These percentages were then added up, for a maximum of 400 points and standardized by dividing by 4.

To test these hypotheses, the following model was analyzed:

$$P(s,w) = \beta_0 + \beta_1 w + \beta_2 w^2 + \beta_3 w^3 + \beta_4 * standardized\ nbr\ of\ interviewers + \varepsilon_{s,w}$$

[3] The ESS guidelines for contact procedure state for each sampling units four contact attempts should be made from which at least one in the evening and one at the weekend, unless contact is made before this is needed. Moreover the contact attempts should be spread over two weeks.

$$\beta_0 = \gamma_{00} + \gamma_{01} * \%\ interviews\ completed\ after\ 1st\ contact\ attempt + frame\ type + u_0$$
$$\beta_1 = \gamma_{10} + \gamma_{11} * \%\ interviews\ completed\ after\ 1st\ contact\ attempt + u_1$$
$$\beta_2 = \gamma_{20} + \gamma_{21} * contact\ score + \gamma_{21} * refusal\ score + \gamma_{21} * conversion\ attempts + u_2$$
$$\beta_3 = \gamma_{30}$$

With $\varepsilon_{s,w} N(0, \sigma^2)$ and level 2 covariance matrix $\begin{bmatrix} \sigma_0^2 & \sigma_{0,1} & \sigma_{0,2} \\ \sigma_{0,1} & \sigma_1^2 & \sigma_{1,2} \\ \sigma_{0,2} & \sigma_{1,2} & \sigma_2^2 \end{bmatrix}$. The estimated fixed effect of this model can be found in table 2.

Table 2: Fixed effects.

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-0,8248	0,6472	31	-1,27	0,212
% completed interviews at 1st attempt	0,07135	0,0097	1952	7,36	<,0001
Standardized nbr interviewers	2,0890	0,04271	1952	48,91	<,0001
individual frame (0)	0,7291	0,1463	1952	4,98	<,0001
Week	-0,2284	0,1238	31	-1,85	0,0745
Week*% completed interviews at 1st attempt	-0,00399	0,001	1952	-3,98	<,0001
Week^2	0,01321	0,00598	31	2,21	0,0346
Week^2*contact score	-0,00003	9,90E-03	1952	-2,87	0,0042
Week^2*refusal score	3,8E-05	2,1E-05	1952	1,78	0,0757
Week^2*percentage conversion attempts	-0,00011	2,8E-05	1952	-4,05	<,0001
Week^3	1,5E-05	9,1E-05	1952	0,16	0,8712

All explanatory variables significantly influenced the shape of the weekly standardized power.

The effect of the survey climate is as expected positive, for 14% more completed interviews at the first contact attempt the first week power increases by 1. Surveys with no individual frame seem to have more completed interviews in the first week than surveys that have an individual frame, which is contrary our expectations.

The number of active interviewers (by 100 sampling units) has, not surprisingly, the largest influence on the weekly power, one more interviewer (for 100 sampling units) in a given week would increase the standardized number of weekly completed interviews by 2,1.
The decrease at the start of the fieldwork in standardized power (scenario 1) is steeper for countries with a better survey climate (higher number of completed interview at the first attempt) or the increase is slower (scenario 2). This also may seem counterintuitive but perhaps a good survey climate means that the fieldwork starts well with a high power but the effect of the survey climate quickly disappears.

The contact score (1 to 100) also has the expected effect, it makes the quadratic coefficient smaller, slowing down the lost in power as does the percentage of conversion attempts. The effect of the refusal score is however unexpected, a well-planned refusal conversion (high score) would accelerate the decrease in power.

Conclusion

- Extending the fieldwork does not seem to be a very fruitful strategy to increase the number of completed interview. It is often a reaction to a too low number of completed interview and not a strategy planned before the start of the fieldwork.
- In general, the (standardized) number of completed interviews per week over the fieldwork period follows a cubic curve, with, first (an increase followed by) an accelerating decrease in power that finally slows down to eventually form a tail.
- The number of interviewers active in a given week has the highest effect on the increase of the weekly power, about 2,1 completed interview per week for one interviewer per 100 sample units.
- The survey climate (as measured by the percentage of completed interviews at the first contact attempt) influences the height of this curve or the starting power, standardized number of interviews achieved in the first week and also the decrease right after the first week, faster decrease if good survey climate. The latter is contrary to our expectations.
- The type of sampling frame (individual or not) has an unexpected effect on the number of completed interviews in the first week (standardized number), non-individual frames resulting in more completed interviews (standardized number). This raises question about the right implementation of the sampling rules for the second stage.
- A good strategy (contact procedure, percentage conversion attempts) slows down the decrease in power. However the refusal conversion score has an unexpected effect, higher refusal score accelerate the decrease in power.
- All covariance parameters are still significant, we hence only explain a small part of the differences between surveys. More knowledge about the planning of the fieldwork and the applied strategy would be needed to further analyze and understand the fieldwork characteristics that influence the shape (height and width).
- Better or longer? It seems that prolonging the fieldwork period to reach higher response rate is not a very fruitful strategy. It is better to have a good start (enough active interviewers) and a well-planned strategy for non-contact and refusal to keep the number of completed interview per week as high as possible for as long as possible.

Questions/discussion

- Other fieldwork characteristics that may influence the shape?
- Any speculations on the correct use of the contact forms. For example, countries that systematically under-report noncontact at the first attempt, might appear to have a better survey climate
- It is relatively hard to 'read' the fieldwork process from the contact files. It may be worthwhile to get a more narrative type of strategic information (e.g. by interviewing the fieldwork management).
- Any recommendations for fieldwork? (better noncontact/refusal policy, more interviewers)
- What about countries with two peaks due to a plan increase in fieldwork effort resulting in a sudden increase in power like Germany, or Lithuania in round 6?
- In the analysis, we used a lot (149) of surveys. But what do the results mean for single surveys? Can we estimate the shape of the power for one survey? Can it be used as a performance indicator for that survey in particular?