

Are proxy answers different in some situations?

LFS Methodology, Budapest May 2019

Francesca della Ratta-Rinaldi, Filomena de Filippo, Alessia Sabbatini; ISTAT

In some surveys as LFS, it is required to interview all the members of the family, and it is often necessary to use the proxy interview, asking a family member to answer in place of another family member when he/she is unable to do so, under penalty of non-completeness of the family interview.

A challenging aspect is how to assess the loss of reliability in proxy responses, in the absence of cognitive tests or direct feedback with the interviewee.

To improve collection strategies, the analysis could help to identify the kind of respondents more "risky" in terms of reliability: for example, by providing surveyors with a vademecum on the type of proxy respondents to be "avoided" or accompanied, and by disseminating indicator data on non-sampling errors for questions characterized by a high incidence of partial non-responses.

1. Introduction

In surveys where it is necessary to interview all the members of the household, it is often unavoidable to use the proxy interview, asking one of the family members to respond on behalf of another when he/she cannot answer, otherwise the completeness of the interview the family unit would be compromised.

In the design of the questionnaire, the possibility of using proxy responses is always taken into account, in the awareness that some questions (for example those of a more subjective type) could present additional difficulties.

Table 1 Rates of family and individual proxies per technique. Years 2015-2018

	Household interviews			Individual interviews		
	capi	cati	total	capi	cati	total
I 2015	39.5	20.8	29.7	26.2	11.6	18.4
II 2015	39.3	23.3	30.8	26.2	13.3	19.1
III 2015	38.4	28.5	33.2	26.0	17.0	21.1
IV 2015	37.9	22.8	29.8	25.7	13.1	18.7
I 2016	38.5	23.0	30.1	26.2	13.4	19.0
II 2016	37.5	27.2	32.0	25.6	16.9	20.8
III 2016	38.3	31.1	34.5	26.4	19.4	22.5
IV 2016	38.2	27.0	32.2	26.2	15.9	20.5
I 2017	40.4	34.6	37.3	27.7	21.2	24.1
II 2017	41.5	42.5	42.0	28.7	26.6	27.6
III 2017	41.3	46.2	43.9	29.0	29.6	29.3
IV 2017	40.2	48.0	44.4	28.1	30.5	29.5
I 2018	39.0	50.2	45.0	27.1	32.5	30.1
II 2018	39.6	56.4	48.3	27.6	37.2	32.8
III 2018	38.9	60.6	50.3	27.2	40.3	34.4
IV 2018	39.6	61.8	51.3	27.7	41.4	35.2

Source: Istat, Labour force survey

In the most recent period, the use of proxies has grown: considering the last four years, we note that the share of families in which at least one proxy was used came from 29.7% in the first quarter of 2015 to 51.3% in the fourth quarter 2018. For individual interviews, the percentage of proxies compared to the total interviews was 18.4% in the first quarter of 2015 and reached 35.2% in the last quarter of 2018.

As for the technique, if in the past the use of the proxy was more frequent in the Capi interviews, due to the difficulty of reaching the home several times to administer the interview, in the last two years the use of the proxy has also grown in the Cati interviews.

The reasons given for the use of the proxy are the unavailability or the temporary absence of individual components, the illness or the lack of interest in the interview. In recent years, the justified proxy for illness and the temporary absence of the interviewee have decreased. On the contrary, the reasons for the unavailability of the components ("difficult to find, never at home") and the lack of interest in the interview have increased (the last two passed from the first quarter of 2015 to the fourth of 2018 from 31 to 42.7% and from 6.9 to 9.9%, respectively).

Undoubtedly, proxy interviews expose to the risk of obtaining less reliable information. It is therefore necessary to keep their number as low as possible and pursue strategies to keep them under control, such as the training of interviewers, communication and awareness campaigns aimed at citizens (informative letters or campaigns that focus on civic duty/right to statistical information), use of data from administrative archives to reduce the length of the interview.

A first evaluation on the quality of the proxy interview data was carried out on the main sections of the questionnaire and in particular on the records modified in the ex-post control and correction phase. This to verify if there were differences in the number of corrections between direct and proxy interviews, then if the number of probabilistic and deterministic imputations performed were different. Besides not having found substantial differences, the corrections on proxy interviews are not necessarily more numerous than those made on direct interviews. Except for the section on employment (main activity), the corrections do not exceed or exceed by just 1%. For the employment section, the imputed records are 4.4% for direct interviews and 4.7% for those in proxy. The only trend that seems to be emerging is that ex-post correction interventions are more frequent when the proxy interview is provided by a relative who does not live in the family or another person, a circumstance that occurs in very rare cases.

Moreover, on previous occasions of analysis in which the survey data were linked with administrative data, there was no effect of the proxy on the conflicting data, which instead are due more to differences in the definitions of the concepts (Istat, 2019).

Starting from the assumption that proxy interviews are inevitable, it is important to develop methodologies for monitoring and evaluating the quality of the data obtained. This is necessary especially in the case of ad hoc modules that "use" the sample of a single quarter or a single wave and ask specific questions in which the quality of response in proxy can become critical (as in the case of the module on Retirement or the next one on Job skills). Since it is not possible to perform cognitive tests or obtain other direct feedback with the interviewed, it is useful to reflect on what other clues it is possible to investigate to ascertain the loss of information and potential distortions of proxy responses.

This work analyses the responses provided in proxies in some questions of the labour force survey questionnaire, focusing in particular on:

- Percentage of "don't know" answers provided in the proxy questionnaire, which is considered a first quality measure of the proxy interview. In particular, some features of the person providing the answers (henceforth the respondent) and of the one in whose name the answers are provided (henceforth the "replied") are compared. The hypothesis is that the more the two subjects "resemble" (both employed, for example) or are on a horizontal relationship axis (wife/husband rather than father/son), the more the answers will be reliable. In other words, the more the respondent in proxy will share some experiences

with the one for whom he provides the answer, the less frequent the unreliable answers will be, including the "don't know".

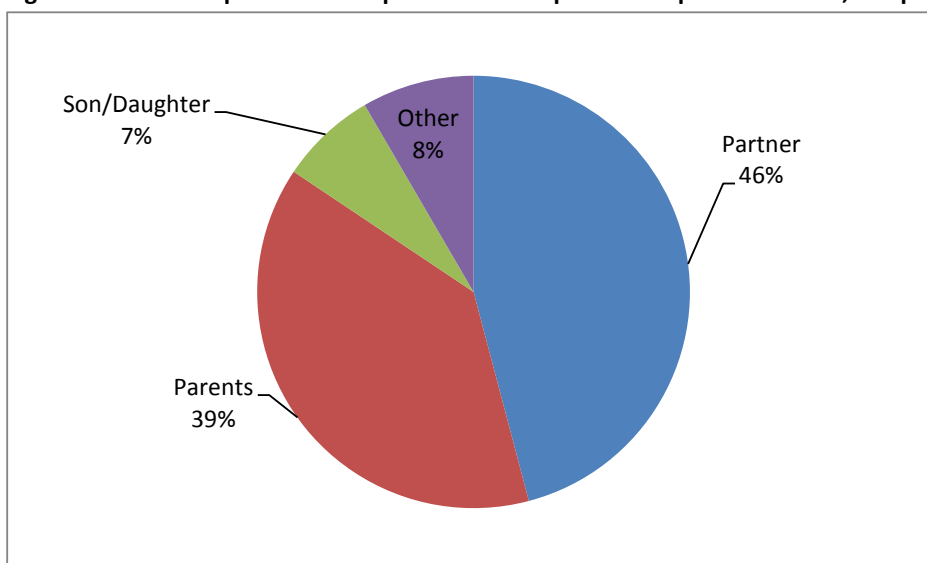
The features on which this verification will be performed are:

- the employment status of respondent and *replied*, in the hypothesis that if the respondent is in employment has fewer difficulties in responding to questions on employment, while an unoccupied respondent may have less ease in providing an answer even for do not supply an inaccurate one;
- The parental relationship between respondent and *replied*, in the hypothesis that the spouse is informed about the employment of his partner more than the sons/daughters for the parents or also the parents for the sons/daughters and therefore the answers have a different degree of reliability.
- Any other evidence of less careful proxy responses. The hypothesis to be checked is whether in the proxy interviews there is a greater concentration in the modal mode or the first modes in the list, or in the simplest ways that allow avoiding additional questions, or in the "lazy" or simpler mode (for example in the case of the timetable "in the week worked as usual");
- Differences in the distribution of proxy responses compared to direct ones, especially when the working condition is similar (for example, when both are self-employed in the module 2017).

The data used for the analysis are those of the second quarter of 2017 which contained the ad hoc module on self-employment (the last module addressed to all waves in a single quarter in Italy). These data were preferred both because the questions of the modules are less known to the interviewee than those on the core questionnaire (the questions are asked only once, while in the standard questionnaire many questions are repeated and therefore more familiar in interviews after the first wave). Furthermore, because they explore "detailed" themes and have a more "subjective" character than those of the core questionnaire.

Only the interviews in which the respondent in a proxy is a resident member of the family were considered, as all the information in the questionnaire is available, for a total of 25,104 interviews. Proxy interviews were answered in 46% of cases by the partner (of which 54.6% were women), the parent by 39% (58.2% mothers). Son/daughter and other family relationships are rarer (Figure 1).

Figure 1 Relationship between respondent and replied. 2nd quarter of 2017, sample data (Percentage values)



Source: Istat, Labour force survey

2. The interpretation of "don't know"

The indicator on which we have focused our attention on the data quality analysis is the percentage of “do not know” answers assigned to each question (the “don't know” answer is only allowed in proxies and only for certain questions) although, obviously, it does not solve all the problems related to the reliability of the answers. The "don't know" involves a partial non-response. Therefore a high proportion of responses with this option makes the data less solid.

The analysis of the distribution of "don't know" for the questions of the ad hoc module generally shows higher percentages than the more structural questions present in the core questionnaire (which confirms our hypothesis). In particular the questions with higher frequencies of "don't know" are those on: the difficulties of self-employment (Hoc 8 - SEDIFFIC), the percentage of gain obtained from the main client (Hoc 2; MAINCLNT) and the desire to change working status (from independent to employee; Hoc16_1-PREFSTAP) with about 7% of "don't know" in the responses provided in the proxy. In the core questionnaire the answers "do not know" in proxies are less (within 3%), except for some questions such as income (for this question everyone can answer "don't know" or choose not to answer, but the percentage of “don't know” in direct interviews is about half of those in proxy), the number of employees in the workplace (SIZEFIRM; incidence of “does not know” higher than 25%), the set of questions on satisfaction in different aspects of work (between 9 and 11%) and the year in which *replied* started the first job and current job (respectively 15.9 and 3.4% of "don't know").

In almost all cases there is a lower percentage of "don't know" if the respondent and the *replied* are both in employment (Table 2). There are different situations for 1) remuneration: when both are in employment, the share of “does not know” is lower, while the share of those who prefer not to answer is slightly higher; 2) in the questions concerning job satisfaction even if in this case the difference is less pronounced. In the ad hoc module, the question on the percentage of earnings obtained from the main client is the one with the greatest difference between in employment and not in employment persons.

Table 2 Rates of "do not know" answer for some questions based on the employment status and the family relationship of the respondent and *replied*. II quarter of 2017

	Status		Relationship with replied				Total	
	Not in empl. /in empl.	Bot employed	Partner	Parent	Son/ daughter	Other		
Core questionnaire								
Number of people working in the workplace (SIZEFIRM)	32.5	24.5	26.1	31.0	29.2	30.3	28.0	
Year start work with curren employer (YSTARTWK)	4.8	2.3	3.1	2.9	7.7	4.9	3.4	
Year stat first job	17.3	14.8	17.6	11.9	19.7	16.6	15.9	
Income (INCDEC)	Don't know	17.1	11.7	11.4	16.8	23.4	19.1	14.1
	Not responding	8.1	8.4	8.9	6.9	7.7	10.0	8.3
	Work in general	9.5	8.7	8.7	8.7	13.7	11.1	9.0
	Earning	10.4	9.2	9.0	10.0	13.7	12.8	9.7
	Work relations	10.5	9.3	9.4	9.5	14.4	12.1	9.8
	Career	13.0	10.5	10.9	12.0	14.6	14.1	11.6
Job satisfaction	Working time	10.3	9.3	9.2	9.7	14.6	11.8	9.7
	Work stability	10.5	9.3	9.1	10.1	14.4	12.3	9.8
	Kind of activity	9.9	9.1	8.9	9.4	14.2	11.6	9.4
	Times from home to work	10.0	9.0	9.0	9.3	14.2	11.4	9.4
	Interesting job	10.8	9.3	9.5	10.0	14.4	12.3	10.0
Ad hoc module 2017								
Percentage of income from the main customer (MAINCLIENT)	9.9	5.2	7.2	6.9	8.5	5.6	7.1	
Difficulty in self-employment (SEDIFFIC)	8.3	6.4	6.1	10.3	9.3	6.5	7.2	
Desire to change working status (PREFSTAP)	9.3	6.1	6.8	9.3	3.9	11.3	7.4	

Source: Istat, Labour force survey

In the ad hoc module, aimed at self-employed workers only, when the respondent is also a self-employed, the "don't know" rate is systematically even lower, confirming the possibility of responding easily when the working situation is similar.

The analysis of the distribution of "don't know" by the relationship between the respondent and the *replied* shows that when the partner responds the percentage of "don't know" is generally lower than in other cases. The only two exceptions concern the year in which the first job started as in this case are parents who answer for sons/daughters that use less often "don't know". And, as for the analysis by occupational condition, the income, where even if the partner responds less often "don't know", at the same time chooses slightly more than the others not to answer.

In the ad hoc module questions, the sons/daughters seem more informed if their parent wants to change status, but more often they don't know the year in which they started working with the employer, or the number of people working in the same workplace. Parents, on the other hand, look more like the partner but differ in the percentage of "don't know" especially regarding: the difficulties of self-employment, the number of people working in the workplace, the desire to change working status. Regarding job satisfaction, if the employment condition does not seem to influence the "don't know" rates, the parental relationship seems to affect instead: the partners tend to provide a different answer than "don't know" more than parents and finally sons/daughters. The items on which the horizontal relationship seems to produce the greatest effect are satisfaction for career, hours worked, stability and the type of work performed.

In general, women are a little more likely to respond "don't know", but not when they are daughters as for sons the "don't know" prevail.

3. Less attention in proxy?

Another indicator for data quality may be considered some empirical evidences that can show the inclination of the proxy respondent to provide answers that match with the modal or with the most "comfortable" response, conventional or "lazy", approaching the most usual situation and not necessarily the real one. In fact, in addition to the use of the "don't know" answer, proxy respondents could choose more often standardised or generic answers.

As in the case, for example, of the difference between usual and actual working hours, proxy respondents declare, more often than the *replied*, they worked "as usual" during the reference week (question C32 HOURS, 83.2% for proxy vs 77.9% for direct interviews). At the same time, for some of the questions for which many response options are proposed, in proxy interviews can be noted the tendency to choose the most desirable ones or those presented first. For question C59a (FINDMETHOD), in which is asked how the current work was found, the most frequent answer option - through relatives or friends - is chosen more often in proxies than in the direct interview (32.9% vs 29.6%). Even for the questions on reasons for a fixed term and part-time jobs (TEMPREAS and FTPTREAS), the frequencies of the first answer option are higher in proxies. In the first case (reasons for temporary jobs) the first two options - training or probationary period - are both around 12% in proxies compared to 6.7 and 9.9% in direct interviews; regarding the reasons for part-time, the first option (training) is chosen by 9.2% of respondents in the case of proxies and by 2.8% in direct interviews. However, we found that the concentration on some options can be explained by parental relationship: it occurs more often in the case of parents who respond for their sons or daughters, and it is reasonable that the reasons for the part-time or temporary job are study (or recent entrance in the labour market) and that the job was found through the family network.

Instead, the more frequent use of the option "having worked as usual" in proxy interviews in the reference week is confirmed, with slight differences considering the parental relationship.

In the ad hoc module, when both the respondent and the *replied* are self-employed, the answers provided are closer to what one would expect from an "genuine" independent worker. It is important to remember that one of the aims of the 2017 ad hoc module was to distinguish the "genuine" self-employed from those with subordination signals (dependent self-employed, DSE; Istat 2018). When there are two or more independent workers in the family, the answers that are more typical for DSEs are less frequent. While when the parents answered for their sons/daughters, the response profiles become the one typical among DSEs, as these figures are very common among the younger who often still live in the family. Except for this difference, which is consistent with the analyses on DSE produced (Istat, 2018), there are no other findings on some particular options in proxy interviews compared to direct ones.

Nevertheless, and more generally, to prevent response set phenomena, it could be useful to introduce items' rotation.

Regarding the similarity of the answers when both - respondent and *replied* - are self-employed, it is necessary to understand if this similarity arises from a projection of one's condition on the other or from a real sharing/knowledge that makes the answers similar. For this reason, in the next section we will focus on the association of responses between respondent and *replied*.

4. The similarity of the answers provided in your interview and on that date in a proxy

Sharing some experiences - being in employment, especially if with similar professional status, or being part of a couple - ensures a lower use of the "don't know" response, and therefore a better reliability of the proxy response. However, it is equally true that the opposite risk could occur: that the respondent projects his assessments on the *replied* by choosing answers closer to his thinking than to the actual condition of the person for whom he responds. In other words, it is necessary to ensure during the interview that the interviewee always remembers not to answer for himself but in the place of another person.

To clarify this aspect, some variables have been analysed, such as the quantitative set on the job satisfaction of different aspects and some questions from the ad hoc module on self-employment.

Since 2013, in the Italian questionnaire, some questions on satisfaction with different aspects of the job have been included, which are answered using a score from 0 to 10 (della Ratta et al.; 2013). This tool, as reported by the interviewers, creates some initial problem of understanding for the interviewees, because he must use a tool (the scale) never used up to that point in the questionnaire. The use of the scale, especially in a long and complex questionnaire such as the labour force survey, poses some problems in terms of response set and tendency to provide homogeneous scores (Gobo, 2003). This especially in the telephone interview, even if the data analysis for categories of employees it shows a significant variability and an undoubted interest (for example see data analysis on the module 2017: della Ratta, Sabbatini, 2019).

We examined (in this case the analysis is limited to those who are both in employment) the association between the scores given to the different aspects of the job by the respondent in his interview and those he assigned for the *replied* in the proxy interview. This to verify the existence or not of an association between the scores assigned for the satisfaction for his own job and for that of the person for which he answers. An association emerges when both are self-employed, especially about job stability, business and general satisfaction (Cramer's V of 0.49, 0.43 and 0.43; Table 3). This association becomes even stronger

when the partner responds and when the activity is carried out within a similar sector of activity (the value of Cramer's V for stability in these cases reaches 0.54)¹.

Table 3 Association between classes of scores assigned to satisfaction for the different aspects of the job provided by the respondent in his direct interview and in proxy for the replied by professional status. 2nd quarter of 2017 (Cramer's V)

SATISFACTION FOR:	Both self-employed	Both employee	Different condition
Work in general	0.43	0.17	0.19
Earning	0.42	0.24	0.17
Work relations	0.4	0.21	0.18
Career	0.43	0.27	0.14
Working time	0.37	0.18	0.18
Work stability	0.49	0.19	0.1
Kind of activity	0.38	0.19	0.17
Times from home to work	0.31	0.21	0.21
Interesting job	0.36	0.2	0.18

Sign. <0.0001

Source: Istat, Labour force survey

Considering the data of the ad hoc module, when both the respondent and the *replied* are self-employed there is a strong concentration of the answers on the main diagonal, which suggests the similarity of the two answers given. The similarity in the answers was analyzed both in the questions with many options, such as those on the reasons of self-employment (REASSE) or on the difficulties (SEDIFFIC), and in the simpler ones such as the question on the presence of associates (BPARTNER) or the degree of autonomy at work (JOBAUTON). The strongest association - measured with Cramer's V - is found in the question about the presence of associates, in the question of difficulties and autonomy concerning the order in which to carry out the activities (Table 4). In this case, since the number of cases is small, it is not possible to investigate the relationship with the sector of activity and the parental relationship, as it was instead done for the questions on the satisfaction of the core questionnaire.

Table 4 Association between answers provided by respondent in his direct interview and in proxy for the replied in some questions of the ad hoc module by professional status. 2nd quarter of 2017 (Cramer's V)

2017 AD HOC MODULE	Both self-employed	Both employee	Different condition
REASSE (*)	0.27		
SEDIFFIC (*)	0.46		
BPARTNER	0.49		
JOBAUTON 1 (order)	0.33	0.25	ns
JOBAUTON 2 (contents)	0.27	0.29	ns

Sign. <0.0001

* For these variables 72% of the cells have expected counts less than 5 and chi-Square may not be a valid test.

Source: Istat, Labour force survey

In the case of questions about job autonomy, it is possible to check the association for all employees. The association does not exist when the occupational condition of the respondent and the *replied* is different (employee / self-employed and viceversa); it is present but weaker when both are employees. Even the distribution of the answers, given as an example for the variable on job autonomy, highlights this type of relationship, which is stronger when both are self-employed (Table 5).

¹ We check this result also in a sample of direct interview: in this case, the association is weaker: it doesn't exist when respondents share different condition, it is weaker but exists when both are self-employed (probably because often people in the same household share the business). In the ad hoc module we found a stronger association with the BPARTNER variable (0.55) but the sample we build was mainly composed by couples.

Table 5 Degree of autonomy in the sequence in which to carry out the tasks by the response provided by respondent in his direct interview and in the proxy for the replied by the professional status. 2nd quarter of 2017 (percentage values and absolute values)

Replied (proxy)	Both employee			Respondent (direct)			Different condition		
	Yes	No	Total	Both self-employed			Yes	No	Total
				Yes	No	Total			
Yes	50.3	25.4	40.0	86.9	48.2	82.1	57.1	62.7	58.3
No	45.5	70.0	55.7	11.3	44.6	15.5	39.1	32.9	37.7
Don't know	4.2	4.6	4.4	1.7	7.2	2.4	3.8	4.4	4.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Total (N)	2,177	1,537	3,714	582	83	665	1,377	386	1,763

Source: Istat, Labour force survey

The difference seems to go against the explanation based on the projection of oneself in proxy responses: if the risk exists, it should occur both in the case of similarity of status (i.e. in the case of two employees) and in the case of diversity. The fact that similarity occurs mostly among self-employed workers seems an element in favour of the reliability of the data, due probably to the sharing of the same job culture that often bonds self-employed workers, especially when they work together or share a family tradition of self-employment.

5. Conclusions

Given the strong increase in proxy interviews in recent years, it is important to ensure that the quality of the data remains high as proxies expose them to the risk of obtaining less reliable answers.

To verify this exposure, some analyses have been carried out which unanimously indicate that the proxy interview data is not affected by distortions such as to compromise its reliability.

We saw that in the ex-post correction phase the number of probabilistic and deterministic imputations performed for proxy interviews was not very different from direct ones. Post correction interventions are more frequent when the proxy interview is provided by a relative or person who does not live in the family, but these are very rare cases.

We found that, except in special cases such as income, the percentage of "don't know" answers in proxy interviews, which result in a partial information loss, is less when the respondent and the *replied* share the same employment condition (both in employment), and even more when they are both self-employed, or when they are a couple. These elements could be useful in the training of interviewers, for example pushing them to choose the most "appropriate" family members to provide proxy responses, using specific vademecum on the type of proxy respondents to "avoid".

However the similarity of condition between respondent and *replied* could expose to the risk of an excessive projection and a semi-cloning of the answers. The similarity of some answers between respondent and *replied* in the most subjective questions is certainly an element of reflection to be deepened in the future with specific assessment tools (i.e. cognitive tests). In the meantime, prevention strategies can be implemented to allow interviewers, for example, to reiterate through personalised wording that in the proxy interview they are answering for another person.

To avoid the concentration of answers on specific modalities, it may be useful to introduce the rotation of the items in the computer-assisted questionnaire. Finally, information on the incidence of partial non-responses by type of responsive proxy could enrich the dissemination of indicators on non-sample errors.

References

- Cannavò L., Frudà L. (2007). *Ricerca sociale. Tecniche speciali di rilevazione, trattamento e analisi*. Carocci, Roma
- della Ratta-Rinaldi F., Marzilli E., Pintaldi F., Pontecorvo M.E., Tibaldi M., (2013). "Measuring job satisfaction in Italian Labour force survey" LFS Methodology, Gdansk, May 2013
- della Ratta-Rinaldi F., Sabbatini A (2019). "Una lettura del lavoro autonomo in chiave territorial". *Quaderni di rassegna sindacale*, N. 2, going to press
- Gobo G. (2003). *Le risposte e il loro contesto. Processi cognitivi e comunicativi nelle interviste standardizzate*. Milano, Angeli
- Grassia M.G., Pintaldi F., Quattrociochi L. (2004). "The electronic questionnaire in ISTAT's new CAPI/CATI Labour Force Survey". *Proceedings of the European Conference on Quality and Methodology in Official Statistics (Q2004)*, Mainz, Germany, May 2004
- Groves R.M., Fowler jr. F.J., Couper M.P., Lepkowski, J.M., Singer E., Tourangeau R. (2004). *Survey methodology*. Wiley, Hoboken New Jersey
- Istat (2018). I lavoratori indipendenti. [Statistica Focus](#), 5 novembre
- Istat (2012). *Linee guida per la qualità dei processi statistici*. Roma
- Pintaldi F., Rapiti F.M. (eds) (2019). *Il mercato del lavoro 2018. Verso una lettura integrata*. Ministero del Lavoro e delle Politiche Sociali, Istat, Inps, Inail, Anpal, Roma 2019
- Ortenzi A., Ranaldi R., Spizzichino A., Vergura E. (2015). "An analysis of the proxy response in the Italian LFS". LFS Methodology, Prague, April 2015
- Ortenzi A., Ranaldi R., Spizzichino A., Vergura E. (2015). "Un'analisi della risposta proxy nella rilevazione forze di lavoro". *Rivista Italiana di Economia, demografia e statistica*. Vol LXIX, n. 4, ottobre 2015