

# istat working papers

N.24  
2016

## **The behaviour of respondents while filling in a Web questionnaire: the case of the Italian business R&D survey**

*Mauro Masselli, Alessandra Nuccitelli and Antonio Laureti Palma*



# istat working papers

N.24  
2016

## **The behaviour of respondents while filling in a Web questionnaire: the case of the Italian business R&D survey**

*Mauro Masselli, Alessandra Nuccitelli and Antonio Laureti Palma*

### **Comitato scientifico**

Giorgio Alleva  
Tommaso Di Fonzo  
Fabrizio Onida

Emanuele Baldacci  
Andrea Mancini  
Linda Laura Sabbadini

Francesco Billari  
Roberto Monducci  
Antonio Schizzerotto

### **Comitato di redazione**

Alessandro Brunetti  
Romina Fraboni  
Maria Pia Sorvillo

Patrizia Cacioli  
Stefania Rossetti

Marco Fortini  
Daniela Rossi

### **Segreteria tecnica**

Daniela De Luca   Laura Peci   Marinella Pepe   Gilda Sonetti

## **Istat Working Papers**

The behaviour of respondents while filling in a Web questionnaire:  
the case of the Italian business R&D survey

N. 24/2016

ISBN 978-88-458-1939-1

© 2016

Istituto nazionale di statistica  
Via Cesare Balbo, 16 – Roma



Salvo diversa indicazione, tutti i contenuti  
pubblicati sono soggetti alla licenza  
Creative Commons - Attribuzione - versione 3.0.  
<https://creativecommons.org/licenses/by/3.0/it/>

È dunque possibile riprodurre, distribuire, trasmettere  
e adattare liberamente dati e analisi dell'Istituto  
nazionale di statistica, anche a scopi commerciali,  
a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat),  
marchi registrati e altri contenuti di proprietà di terzi  
appartengono ai rispettivi proprietari e non possono  
essere riprodotti senza il loro consenso.

## The behaviour of respondents while filling in a Web questionnaire: the case of the Italian business R&D survey

Mauro Masselli, Alessandra Nuccitelli and Antonio Laureti Palma

### Sommario

*Le opportunità tecnologiche offerte dalle indagini condotte via Web consentono di monitorare con precisione il comportamento di ciascun rispondente durante la compilazione di un questionario. I file di log generati possono fornire utili informazioni sul processo di risposta: come il rispondente passa da una schermata all'altra, quante volte viene scaricata la versione stampabile del questionario, a quante domande si risponde, quanto tempo richiede il processo di risposta, ecc..*

*In questo contesto, i dati di log prodotti in occasione della rilevazione italiana su Ricerca e Sviluppo nelle imprese per l'anno 2009 sono analizzati al fine di individuare eventuali criticità nella progettazione del questionario e capire come migliorare l'accuratezza delle risposte.*

*Sono proposti alcuni indicatori che descrivono il comportamento delle imprese rispondenti. I risultati derivanti dai dati di log vengono riportati secondo il percorso di navigazione e discussi.*

**Parole chiave:** carico sui rispondenti, CAWI, correlazione, paradati.

### Abstract

*The technological opportunities offered by Web-based surveys allow to track precisely the behaviour of each respondent while filling in a questionnaire. The created log files can provide useful information about the response process: how the respondent navigates from one screen to another, how many times the printable version of the questionnaire is downloaded, how many questions are answered, how long the response process takes, etc..*

*In this framework, log data from the Italian business Research & Development survey for the year 2009 are analysed in order to identify possible critical aspects in the questionnaire design and to understand how to improve the accuracy of responses.*

*Some indicators describing the behaviour of the respondent companies are proposed. Results from log data are reported by navigational path and discussed.*

**Keywords:** respondent burden, CAWI, correlation, paradata.

## Index

	Pag.
<b>1. Introduction</b> .....	7
<b>2. The Italian business R&amp;D survey</b> .....	7
2.1 The Web questionnaire .....	8
2.2 The Web application for the questionnaire .....	8
2.3 Some results from data collection and validation .....	9
<b>3. Indicators of the behaviour of the respondent companies</b> .....	10
<b>4. Main results</b> .....	10
4.1 Merging indicators with ASIA .....	12
<b>5. Concluding remarks</b> .....	14
<b>References</b> .....	15

## 1. Introduction\*

The technological opportunities offered by Web-based surveys allow to track precisely the behaviour of each respondent while filling in a Web questionnaire. The paradata<sup>1</sup> collected in Web surveys can provide factual information about the online response process (Bethlehem and Biffignandi, 2011).

The collection of paradata can happen on the server side and/or on the client side.

In the first case, paradata are collected at the server on which the Web survey resides and essentially describe server events, typically “visits” to Web pages. Usually, every visit to each Web page of the questionnaire is logged, together with an identifying code and a timestamp. This enables to track the survey progress and the behaviour of each respondent who started the survey. In the other case, paradata are collected at the client side or at the respondent device level and describe events within a Web page, such as mouse clicks or changes of answers. Whereas server-side paradata can be collected independently of future actions by a respondent (the moment a visit occurs, these data are recorded and available), client-side paradata need to be transmitted from the respondent’s computer to the server (Heerwegh, 2002).

In this framework, detailed server-side paradata related to the Italian R&D (Research & Development) survey for the year 2009 are analysed in order to identify possible critical aspects in the questionnaire design and to understand how to improve the accuracy of responses.

The paper is organized as follows. In section 2, after introducing some characteristics of the survey and of the innovative Web application developed for the R&D questionnaire, a few results from data collection and validation are provided. On the basis of the available paradata, a set of indicators describing the behaviour of the respondent companies is proposed (section 3). The main results are reported by navigational path and discussed in section 4; in particular, with reference to intramural<sup>2</sup> R&D performers, the resulting indicators are merged with statistical information coming from the Italian register of active enterprises. Finally, some conclusions are drawn, making a suggestion on how to reduce the burden on respondents (section 5).

## 2. The Italian business R&D survey

The Italian business R&D survey produces statistics on R&D expenditures and on personnel involved in R&D activities. The survey – which is conducted on the Web by the Italian National Statistical Institute (Istat) – is based on the methodological recommendations provided by the “Frascati Manual” (OECD, 2002), the main source of theoretical and practical guidelines to produce official R&D statistics at the international level.

The Istat business R&D survey is an annual census of the population of the Italian companies that can be identified as “potential R&D performers” in the reference year.

In order to identify such a target population, Istat is regularly collecting information from a range of statistical and administrative data sources. They include: previous Istat R&D surveys, other Istat business surveys with R&D-related questions, the Italian register of active enterprises, the Italian register of R&D performing institutions (managed by the Ministry of University and Research), data on national and European Union funding to research projects, patent databases, business reports, data from the Italian Tax Agency.

For a better interpretation of the results reported in the present paper, it is worth mentioning some peculiar aspects of the response process in economic surveys, like that on business R&D. In

---

\* An early version of this paper was presented by Masselli M. and Nuccitelli A. at the Conference on New Techniques and Technologies for Statistics, Brussels, March 5-7, 2013.

<sup>1</sup> Paradata – this term was coined by Couper (1998) – are data about the process by which survey data are collected.

<sup>2</sup> A company or statistical unit may have expenditures for R&D either within the unit (intramural) or outside it (extramural). A more detailed definition of intramural expenditures can be found in OECD (2002).

such surveys, respondents tend not to be answering questions for themselves as individuals, but as representatives of their businesses, and should be familiar with tables, matrices, and numerical information. Besides, these surveys may need to be completed by multiple respondents and the release of data may require approval by the company. In the case of periodic surveys, it is important to take into account of previously reported data and of changes in the organizational structure, such as through a merger and/or acquisition. Finally, printouts of Web questionnaires are frequently used to support the preliminary process of identifying what information needs to be provided. Respondents often use paper forms as rough drafts before attempting to enter the data and answer the sequence of questions on multiple topics that appear on successive screens of a Web survey (Morrison *et al.*, 2008).

## 2.1 The Web questionnaire

In the Web questionnaire of the Italian business R&D survey, the questions are numbered consecutively from beginning to end, and each question is placed on a single Web page (or screen)<sup>3</sup>. The respondents can freely move back and forth through the questionnaire and make corrections as needed.

Inconsistencies (or errors) within a question or between different questions are pointed out and described in detail at the top of each page.

The section concerning the company information and the printable version of the questionnaire are placed at the end, after the last question.

Depending on the answer given to the first question, the path to be followed is highlighted in bold text on each Web page:

- a) intramural R&D performers in the reference year should complete all the questionnaire (25 questions and the section concerning the company information);
- b) extramural R&D performers in the reference year should complete only questions 1, 3, and the section concerning the company information;
- c) future intramural R&D performers (not involved in R&D in the reference year) should complete questions 1, 5, 19, and the section concerning the company information;
- d) companies not involved in R&D should complete only question 1 and the section concerning the company information.

The underlying software architecture is very innovative, in the sense that the electronic form of the questionnaire – delivered via the Web on the respondent device – is physically distinct from its “smart” component, which is located in a Web server on the premises of Istat.

During the survey, a set of paradata is recorded and used to monitor the fieldwork activities and to guide the intervention decisions during the data collection. Specific tools were developed to display the flow of process information and to facilitate decision-making.

## 2.2 The Web application for the questionnaire

The Web application developed for the R&D questionnaire responds to the basic requirement to cut the response time due to the complexity of the survey. In practice, respondents start to fill in the questionnaire but interrupt their activity whenever they lack information. This forces respondents to explore all the questions first and to resume the response process later. For this reason, the Web application must be able to store any answer at any time and to browse the questionnaire without routing constraints.

<sup>3</sup> It is worth noting that the informational value of server-side paradata increases as fewer questions are placed on a single Web page. As Web pages contain more questions, the level of detail of server-side paradata decreases, since respondent’s actions within a particular Web page are not recorded.

To satisfy this requirement, the Web application includes two main components: a routing-tool – to support the correct logical route sequence of questions to be answered – and an editing-tool<sup>4</sup> – to check for inconsistencies and give feedback to respondents.

In order to decouple the statistical concept code from the software source code, the Web questionnaire was developed as a completely metadata-driven application. Moreover, in order to use a common language, all metadata related to statistical concepts were expressed using the first-order logic language. Therefore, the Web application was developed following the Model-View-Controller (MVC) paradigm, according to which the software application is divided into three interconnected parts, so as to separate internal representations of information from the ways information is presented. The central component, the “model”, consists of application data, business rules, logic, and functions. The “view” is any questionnaire visualization or information release. The third part, the “controller”, accepts input and converts it into commands for the model or the view.

The controller is also responsible for the real time tracking of the behaviour of respondents, which is the source for paradata. These are stored in a dedicated log table for each respondent company and differentiated by clickstreaming and violation of statistical concepts.

Paradata enable real time monitoring of the effective behaviour of each respondent company and can be used to assist the user while filling in the questionnaire, establishing a direct contact (for instance, by telephone) with him during the data collection. Furthermore, in this software architecture, using feedback from real time paradata, statistical concepts can be improved or corrected while the application is online. This is possible because all business rules are located in the application server, while the client contains only the view and is “thin” in terms of software.

Since the risks of browser incompatibility are minimized, this kind of configuration proves to be very useful, especially when a large sample of heterogeneous respondents is involved in a Web survey.

### 2.3 Some results from data collection and validation

With reference to the year 2009, the target population for the Italian business R&D survey consisted of 18,945 companies. Istat collected data via the Web from 9,551 companies.

As displayed in table 1, about two thirds (66.8 per cent) of the questionnaires were error-free and were considered validated; the remaining questionnaires (33.2 per cent) needed to be reviewed – after data collection – by a direct contact with the companies. Besides, intramural R&D performers in 2009 (path *a*) represented about 58 per cent of the total respondent companies; companies not involved in R&D activities (path *d*) were slightly more than one third. Percentages of questionnaires related to extramural R&D performers (path *b*) or to future intramural R&D performers (path *c*) were very low.

**Table 1. Main results from data collection and validation by path**

PATH	n. of error-free questionnaires	n. of reviewed questionnaires	total
<i>a</i>	2,938	2,555	5,493
<i>b</i>	263	63	326
<i>c</i>	146	227	373
<i>d</i>	3,036	323	3,359
total	6,383	3,168	9,551

Source: Elaboration based on Istat data

<sup>4</sup> It includes more than 400 checking rules.

### 3. Indicators of the behaviour of the respondent companies

On the basis of the paradata collected to support the management and monitoring of the business R&D survey, a set of indicators can be built to study *a posteriori* the behaviour of each respondent company while filling in the Web questionnaire.

The proposed indicators can be considered as proxies of the burden on the company:

- $I_j$ : number of times question  $j$  is visited ( $j = 1, 2, \dots, 25$ );
- $I_{26}$ : length of the actual navigational path (in terms of number of visited questions);
- $I_{27}$ : number of times the section concerning the company information is visited;
- $I_{28}$ : gross connection time, in days (the time span between the first and last timestamps);
- $I_{29}$ : connection time, in minutes (the sum of each time span between the first and last daily timestamps);
- $I_{30}$ : number of days in which the connection occurs;
- $I_{31}$ : number of logins;
- $I_{32}$ : number of save operations;
- $I_{33}$ : number of downloads of the printable version of the questionnaire;
- $I_{34}$ : number of times in which the respondent company goes back when completing the questionnaire;
- $I_{35}$ : number of errors made in completing the questionnaire.
- $I_{36}$ : number of variables for which the value provided by the respondent company is different from the validated one<sup>5</sup>.

### 4. Main results

In table 2, the mean and 90<sup>th</sup> percentile of the indicators proposed in section 3 are reported by path.

With reference to path *a* (columns 2 and 3), the gross connection time spent performing the task of participating in the survey is about 17 days, while the actual connection lasts nearly 3 hours spread over 2.3 days (mean values).

The completion of the questionnaire requires on average 4.5 logins and approximately 35 save operations. It should be noted that for a tenth of the companies responding to path *a* – slightly more than 500 intramural R&D performers – the connection time is very long (more than 6 hours).

On the average, the length of the actual navigational path is nearly four times the length of the questionnaire; the most visited questions are 16 “*Personnel involved in intramural R&D activities by sex, occupation and qualification*” (7.9 visits), 2 “*Expenditures for intramural R&D activities by type of costs*” (7 visits), and 15 “*Personnel involved in intramural R&D activities by sex and age category*” (6.8 visits). In particular, for a tenth of the companies responding to path *a*, the actual navigational path is at least 161 questions long.

About 241 errors are made and the most frequent ones are those related to questions 2, 4 “*Intramural R&D expenditures by source of funds*”, 14 “*Personnel involved in intramural R&D activities by co-operation partner*”, 16, and 17 “*Personnel involved in intramural R&D by occupation and type of R&D activity*”. The number of variables for which the value provided by the respondent company is different from the validated one is 6.8 (mean values).

<sup>5</sup> Obviously,  $I_{36} = 0$  for companies whose questionnaire was error-free.

Due to the shorter length, the other paths (*b*, *c*, and *d*) show lower values of the indicators, when compared to path *a* (columns 4-9 of table 2). On the average, the actual connection lasts approximately one hour for the companies responding to paths *b* and *c*, while companies not involved in R&D spend about half an hour only to complete the first question and the section concerning the company information.

Besides, values reported in column 4-9 show that a few respondents to paths *b*, *c*, and *d* tend to explore the questionnaire by selecting questions not included in their own path.

When completing the questionnaire, intramural R&D performers (path *a*) go back an average of 4.2 times. As displayed in table 3, the most frequent turning questions are 25 “*R&D of foreign branches and subsidiaries*” and 16.

**Table 2. Indicators of the behaviour of the respondent companies by path (mean values and 90<sup>th</sup> percentiles)**

INDICATORS	path a		path b		path c		path d	
	mean	90 <sup>th</sup> percentile						
<i>l</i> <sub>1</sub>	2.7	5.0	2.2	4.0	2.9	5.0	2.2	4.0
<i>l</i> <sub>2</sub>	7.0	13.0	0.9	3.0	0.9	2.0	0.5	1.0
<i>l</i> <sub>3</sub>	4.3	9.0	3.4	6.0	0.6	1.0	0.4	1.0
<i>l</i> <sub>4</sub>	4.8	9.0	0.4	1.0	0.5	1.0	0.3	1.0
<i>l</i> <sub>5</sub>	4.3	8.0	0.4	1.0	3.8	7.0	0.3	1.0
<i>l</i> <sub>6</sub>	4.0	7.0	0.4	1.0	0.4	1.0	0.2	1.0
<i>l</i> <sub>7</sub>	3.8	7.0	0.4	1.0	0.4	1.0	0.2	1.0
<i>l</i> <sub>8</sub>	2.8	6.0	0.2	1.0	0.3	1.0	0.2	1.0
<i>l</i> <sub>9</sub>	2.5	5.0	0.2	1.0	0.3	1.0	0.2	1.0
<i>l</i> <sub>10</sub>	3.3	6.0	0.2	1.0	0.3	1.0	0.2	1.0
<i>l</i> <sub>11</sub>	2.6	5.0	0.2	1.0	0.2	0.0	0.2	0.0
<i>l</i> <sub>12</sub>	2.5	5.0	0.2	1.0	0.2	0.0	0.2	0.0
<i>l</i> <sub>13</sub>	2.5	5.0	0.2	1.0	0.2	0.0	0.2	0.0
<i>l</i> <sub>14</sub>	6.1	12.0	0.5	1.0	0.4	0.0	0.3	0.0
<i>l</i> <sub>15</sub>	6.8	14.0	0.4	1.0	0.3	0.0	0.2	0.0
<i>l</i> <sub>16</sub>	7.9	16.0	0.3	1.0	0.3	0.0	0.2	0.0
<i>l</i> <sub>17</sub>	5.8	12.0	0.3	0.0	0.2	0.0	0.2	0.0
<i>l</i> <sub>18</sub>	5.0	10.0	0.2	0.0	0.2	0.0	0.1	0.0
<i>l</i> <sub>19</sub>	2.4	5.0	0.2	1.0	2.7	5.0	0.2	0.0
<i>l</i> <sub>20</sub>	2.1	4.0	0.2	1.0	0.1	0.0	0.1	0.0
<i>l</i> <sub>21</sub>	2.0	4.0	0.2	1.0	0.1	0.0	0.1	0.0
<i>l</i> <sub>22</sub>	2.0	4.0	0.2	0.0	0.1	0.0	0.1	0.0
<i>l</i> <sub>23</sub>	2.0	4.0	0.2	0.0	0.1	0.0	0.1	0.0
<i>l</i> <sub>24</sub>	1.9	4.0	0.1	0.0	0.1	0.0	0.1	0.0
<i>l</i> <sub>25</sub>	2.1	4.0	0.1	0.0	0.1	0.0	0.1	0.0
<i>l</i> <sub>26</sub>	93.3	161.0	12.3	26.0	15.9	27.0	7.2	17.0
<i>l</i> <sub>27</sub>	3.7	7.0	2.9	6.0	3.4	7.0	2.9	6.0
<i>l</i> <sub>28</sub>	16.7	51.0	10.8	35.0	15.3	48.9	8.4	32.7
<i>l</i> <sub>29</sub>	170.5	399.0	57.4	200.0	62.4	168.0	34.3	78.0
<i>l</i> <sub>30</sub>	2.3	4.0	1.7	3.0	1.8	3.0	1.5	2.0
<i>l</i> <sub>31</sub>	4.5	9.0	3.0	6.0	3.4	7.0	2.6	5.0
<i>l</i> <sub>32</sub>	35.4	63.0	4.9	8.0	4.8	8.0	2.0	3.0
<i>l</i> <sub>33</sub>	2.3	4.0	1.6	3.0	1.8	4.0	1.4	3.0
<i>l</i> <sub>34</sub>	4.2	8.0	0.9	3.0	1.4	3.0	0.4	1.0
<i>l</i> <sub>35</sub>	241.3	590.0	29.6	32.0	29.0	37.0	13.1	14.0
<i>l</i> <sub>36</sub>	6.8	18.0	0.9	2.0	1.5	4.0	0.2	0.0

Source: Elaboration based on Istat data

The relatively high number of turns that occur at the last question – nearly a quarter – indicates that most of the companies responding to path *a* (about 62 per cent) get to the end of the questionnaire before correcting inconsistent answers to earlier questions or, at least, before providing definitive answers.

When a turn occurs, the most visited questions<sup>6</sup> are 15, 16 and 2.

**Table 3. Number of times in which intramural R&D performers go back when completing the questionnaire, by turning question**

TURNING QUESTION	number of times in which intramural R&D performers go back	
	absolute frequency	%
2	953	4.09
3	1,250	5.37
4	1,275	5.48
5	1,154	4.96
6	763	3.28
7	566	2.43
8	240	1.03
9	329	1.41
10	340	1.46
11	338	1.45
12	165	0.71
13	469	2.01
14	717	3.08
15	1,659	7.13
16	3,198	13.74
17	1,764	7.58
18	1,176	5.05
19	449	1.93
20	312	1.34
21	148	0.64
22	130	0.56
23	267	1.15
24	230	0.99
25	5,388	23.14
total	23,280	100.00

Source: Elaboration based on Istat data

#### 4.1 Merging indicators with ASIA

In order to evaluate if possible difficulties in completing the questionnaire may be attributed or associated to some dimensional business characteristics, the resulting indicators are merged by company with statistical information coming from the business R&D survey itself and from ASIA (year 2009), the Italian register of active enterprises.

With reference to intramural R&D performers, table 4 shows the Pearson correlation coefficients between the indicators of the respondent behaviour and the following business characteristics:

- *ratio of number of researchers to number of employees;*
- *ratio of R&D expenditure to turnover;*
- *number of employees;*
- *turnover.*

Somewhat surprisingly, no significant correlation is found.

The weakness of the relationships between the above indicators and business characteristics is also confirmed by the results from the application of canonical correlation analysis<sup>7</sup>.

<sup>6</sup> For the sake of brevity, the number of times each question is visited, when a turn occurs, is not reported.

<sup>7</sup> Besides, results from the application of analysis of variance indicate that neither the type of economic activity nor the geographical region of the respondent companies has a significant effect on some of the proposed indicators. For the sake of brevity, details about the performed analyses are not reported.

**Table 4. Pearson correlation coefficients between the proposed indicators and business characteristics, with reference to companies responding to path a**

INDICATORS	correlation coefficients			
	n. of researchers / n. of employees	R&D expenditure / turnover	n. of employees	turnover
<i>I</i> <sub>1</sub>	0.02	0.02	0.02	0.00
<i>I</i> <sub>2</sub>	0.05	0.01	0.00	-0.01
<i>I</i> <sub>3</sub>	0.07	0.02	0.01	-0.01
<i>I</i> <sub>4</sub>	0.08	0.02	0.02	-0.01
<i>I</i> <sub>5</sub>	0.04	0.02	0.03	0.01
<i>I</i> <sub>6</sub>	0.03	0.03	-0.01	-0.01
<i>I</i> <sub>7</sub>	0.01	0.04	0.01	0.00
<i>I</i> <sub>8</sub>	0.02	0.03	0.00	0.01
<i>I</i> <sub>9</sub>	0.01	0.03	0.02	0.00
<i>I</i> <sub>10</sub>	0.00	0.03	0.00	0.00
<i>I</i> <sub>11</sub>	0.01	0.05	0.02	0.01
<i>I</i> <sub>12</sub>	0.02	0.04	0.01	0.00
<i>I</i> <sub>13</sub>	0.03	0.04	0.02	0.00
<i>I</i> <sub>14</sub>	0.05	0.01	0.01	-0.01
<i>I</i> <sub>15</sub>	0.04	0.01	0.02	0.03
<i>I</i> <sub>16</sub>	0.01	0.00	0.05	0.02
<i>I</i> <sub>17</sub>	0.01	0.01	0.03	0.01
<i>I</i> <sub>18</sub>	0.01	0.00	0.06	0.03
<i>I</i> <sub>19</sub>	0.00	0.03	0.05	0.03
<i>I</i> <sub>20</sub>	0.00	0.04	0.03	0.01
<i>I</i> <sub>21</sub>	0.00	0.02	0.04	0.02
<i>I</i> <sub>22</sub>	0.02	0.02	0.06	0.05
<i>I</i> <sub>23</sub>	0.02	0.02	0.06	0.03
<i>I</i> <sub>24</sub>	0.01	0.02	0.04	0.02
<i>I</i> <sub>25</sub>	0.00	0.02	0.06	0.05
<i>I</i> <sub>26</sub>	0.03	0.02	0.03	0.01
<i>I</i> <sub>27</sub>	-0.01	0.00	0.07	0.03
<i>I</i> <sub>28</sub>	0.01	0.00	0.04	0.03
<i>I</i> <sub>29</sub>	0.01	0.00	0.10	0.04
<i>I</i> <sub>30</sub>	0.02	0.01	0.10	0.06
<i>I</i> <sub>31</sub>	0.00	0.00	0.10	0.06
<i>I</i> <sub>32</sub>	0.04	0.00	0.05	0.02
<i>I</i> <sub>33</sub>	-0.02	0.01	0.03	0.01
<i>I</i> <sub>34</sub>	0.04	0.00	0.00	0.02
<i>I</i> <sub>35</sub>	0.01	0.00	0.07	0.02
<i>I</i> <sub>36</sub>	0.01	0.01	0.01	0.00

Source: Elaboration based on Istat data

## 5. Concluding remarks

On the basis of paradata collected during the R&D survey, a set of indicators is proposed to study the behaviour of respondents while filling in the Web questionnaire. These indicators are analysed by navigational path in order to identify possible critical aspects in the questionnaire design and to understand how to improve the accuracy of responses. It is worth noting that the questionnaire – in particular, its contents – can be modified only to some extent, as final statistics have to be internationally comparable.

Obviously, any reference to the complexity of the questionnaire and to the associate burden on respondents concerns mainly intramural R&D performers.

The most problematic questions – that is, those with the most visits or errors – are characterized by a high number of either data items or inter-relationships among data items, but they can not further improved.

The analysis of the turning questions shows that most of the intramural R&D performers get to the end of the Web questionnaire before correcting inconsistent answers to earlier questions or, at least, before providing definitive answers. So, providing respondents with the printouts at the beginning of the survey – rather than after the last question – could prevent errors and too long navigational paths.

Finally, no significant correlation between the set of indicators and the business characteristics considered in this paper is found. Unfortunately, other variables, like characteristics of individuals answering the questionnaire or information on the organization of the requested data within each company, are not available. As a suggestion for future work, it will be interesting to study the extent of the relationship between indicators of the respondent behaviour and business characteristics for other economic Web surveys.

## References

- Bethlehem J. and S. Biffignandi. 2011. *Handbook of Web surveys*. Hoboken, New Jersey: John Wiley & Sons, Inc..
- Brancato G., Macchia S., Murgia M., Signore M., Simeoni G., Blande K., Körner T., Nimmergut A., Lima P., Paulino R. and J. H. P. Hoffmyer-Zlotnik. 2006. *Handbook of recommended practices for questionnaire development and testing in the European statistical system*. European Commission Grant Agreement 200410300002. Available at: [http://www.istat.it/it/files/2013/12/Handbook\\_questionnaire\\_development\\_2006.pdf](http://www.istat.it/it/files/2013/12/Handbook_questionnaire_development_2006.pdf) (accessed March 2015).
- Couper M. 1998. Measuring survey quality in a CASIC environment. In: *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 41-49.
- Heerweegh D. 2002. Describing response behavior in Websurveys using client side paradata. Paper presented at the International Workshop on Websurveys, Mannheim, Germany, October 17-19.
- Morrison R. L., Stokes S. L., Burton J., Caruso A., Edwards K. K., Harley D., Hough C., Hough R., Lazirko B. A., and S. Proudfoot. 2008. *Economic Directorate guidelines on questionnaire design*. U.S. Census Bureau. Available at: [http://www.census.gov/srd/Economic\\_Directorate\\_Guidelines\\_on\\_questionnaire\\_Design.pdf](http://www.census.gov/srd/Economic_Directorate_Guidelines_on_questionnaire_Design.pdf) (accessed March 2015).
- OECD. 2002. *Frascati Manual. Proposed standard practice for surveys on research and experimental development*. 6<sup>th</sup> edition. Paris.