Istat
Istituto Nazionale
di Statistica

PLANNING A SURVEY ON GENDER-BASED VIOLENCE

PLANNING A SURVEY ON GENDER-BASED VIOLENCE

Edited by Roberta Barletta.

Editorial activities: Nadia Mignolli (coordinator), Claudio Bava, Alfredina Della Branca, Marco Farinacci, Alessandro Franzò, and Manuela Marrone.

Responsible for graphics: Sofia Barletta.

ISBN 978-88-458-2077-9

Istituto Nazionale di Statistica
Via Cesare Balbo, 16 - Roma

STREETLIB

CONTENTS

	Page
Frequently used acronyms	7
Introduction	9
1. Measuring gender-based violence	11
1.1 Main indicators and topics to be addressed	11
1.2 A new framework for statistical data	17
1.2.1 The international panorama about surveys	19
1.2.2 An overview of survey's characteristics	20
1.2.3 The violence panorama as depicted by the surveys: forms of violence, characteristics of violence, consequences, risk factors, disclosure	21
1.2.4 Best practices	23
1.2.5 The questionnaire	26
1.2.6 Critical aspects	27
1.3 Target population: sex	29
1.3.1 Surveys on violence against women	30
1.3.2 Gender-based violence against men	33
References	38
2. Questionnaire design	41
2.1 The general structure of the questionnaire	41
2.2 The screening questions approach	42
2.2.1 The wording	42
2.2.2 How many sections of screening	44
2.2.3 Time frame and reference period	45
2.2.4 Relationship to the perpetrator	46
2.2.5 Frequency and intensity of violence	47
2.3 Follow up questions for victims	47
References	50
3. Qualitative methodologies for questionnaire assessment	51
3.1 Tools and methods for testing a questionnaire: overview	52
3.2 Focus groups	54

Page

3.3 Expert reviews 62
3.4 Cognitive interviews 66
3.4.1 Cognitive interviewing techniques 67
3.4.2 Planning a cognitive test 70
3.4.3 Conclusion 83

References 85

4. Sampling issues 89
4.1 Sampling issues for a victimization survey about violence 89
4.1.1 Target population 89
4.1.2 Sampling frame and coverage 89
4.1.3 Planning the sampling design 91
4.1.4 Sample size 93
4.1.5 Special features for sampling in mixed-mode surveys 95

References 97

5. Modes of interviewing in victimisation surveys about violence: review and suggestions 99
5.1 Interviewing Modes 100
5.1.1 Different available modes 101
5.1.2 Mode effect related problems 105
5.2 Strategies for controlling differences in the interviewing mode 109

References 114

6. Selection and training of interviewers 117
6.1 Skills and attributes of interviewers 117
6.2 Training of interviewers 120
6.2.1 Initial training 121
6.2.2 Continuous training 124
6.2.3 Training methodologies 125
6.2.4 Organisational aspects 126
6.2.5 Training materials 128
6.3 Ethical considerations and support services for respondents 130

References 135

7. The quality of the statistical process 137
7.1 Strategies to reduce refusals 137
7.1.1 The introductory letter 138
7.1.2 The toll-free number 140
7.1.3 Scheduling the mailing 141
7.1.4 Brochure with key statistics and information notes for the respondents 142
7.1.5 Find the best time and location for the interview 142
7.1.6 Contact rules 142

Page

7.2 Possible outcomes of a phone interview 145
7.3 Quality control during the survey process 151
7.3.1 Daily assistance 151
7.3.2 Re-contact calls 151
7.3.3 Quantitative indicators on the quality levels 152
7.3.4 Monitoring computer assisted interviews 153
7.3.5 Monitoring paper and pencil interviews 156
7.3.6 Control charts 156
7.4 Studying the interviewers' effect 160

References 161

8. Dissemination plan 163
8.1 Who is our audience? 163
8.2 What information do we want to disseminate? 165
8.3 How will it be disseminated? 167
8.4 When will we disseminate these results and raise awareness? 171
8.5 What dissemination of new digital data? 171

References 173

FREQUENTLY USED ACRONYMS

ACASI	Audio Computer-Assisted Self-Interviewing
AAPOR	American Association for Public Opinion Research
CAPI	Computer-Assisted Personal Interviewing
CASI	Computer-Assisted Self-interviewing
CATI	Computer-Assisted Telephone Interviewing
CAWI	Computer-Assisted Web Interviewing
CEDAW	Committee on the Elimination of Discrimination Against Women
EIGE	European Institute for Gender Equality
EU	European Union
EUFRA	European Union Agency for Fundamental Rights
GBV	Gender-Based Violence
IPV	Intimate Partner Violence
NSOs	National Statistical Offices
PAPI	Paper And Pen interviewing
UN	United Nations
UNECE	United Nations Economic Commission for Europe
UNODC	United Nations Office on Drugs and Crime
UNSD	United Nations Statistics Division
VAW	Violence Against Women
VCASI	Video Computer-Assisted Self-Interviewing
WHO	World Health Organization

INTRODUCTION

On 11 May 2011, the Council of Europe adopted the Convention on Preventing and Combating Violence Against Women and Domestic Violence, the so-called Istanbul Convention, which entered into force on 1st August 2014. The Convention indicates: "*For the purpose of the implementation of this Convention, parties shall undertake to collect disaggregated relevant statistical data […] (and) shall endeavour to conduct population-based surveys at regular intervals to assess the prevalence of and trends in all forms of violence covered by the scope of this Convention*"[1].

The Convention also defines violence against women as "*all acts of gender-based violence that result in, or are likely to result in, physical, sexual, psychological or economic harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life*". It also defines domestic violence as "*all acts of physical, sexual, psychological or economic violence that occur within the family or domestic unit or between former or current spouses or partners, whether or not the perpetrator shares or has shared the same residence with the victim*".

To meet the requirements of the Convention and better answer policy requirements at the European level, in 2016, during the Working Group on Crime, The European Commission proposed to develop an EU-wide survey on the prevalence of Gender Based Violence (GBV) within the European Statistical Systems to ensure comparability and support Member States to fulfil Article 11 of the Istanbul Convention. The Working Group agreed to establish a Task Force with the objective to develop and test a population-based survey methodology appropriate for collecting representative statistics on gender-based violence in EU Member States and covering the data collection requirements of the Istanbul Convention.

The Task Force was established in 2016 and its Members were Austria, Denmark, France, Germany, Holland, Hungary, Italy, Latvia, Luxembourg, Portugal, Spain, and Sweden[2].

The Task Force activities were also defined:

- review existing experiences of conducting GBV/VAW surveys: good practices, key issues;
- identify the core methodological requirements of a survey on GBV;
- testing surveys on GBV at national level;
- develop a final methodological manual for implementation surveys on GBV at the national level.

In 2016, to support the work of the Task Force, Eurostat called for proposals and signed a grant agreement with Istat in the framework of "*Support for the development of the methodology for a survey on gender-based violence*". Istat supported the Task Force from November 2016 to May 2019.

This work collects, systematises, and, when needed, updates the papers produced by Istat to support the EU Task Force.

Recently, the Directive (EU) 2024/1385 of the European Parliament and of the Council of 14th May 2024 on combating violence against women and domestic violence has been requesting Member States to conduct population-based surveys at regular intervals to assess the prevalence of, and trends in, all forms of violence covered by it (Article 44, paragraph 3).

1 Council of Europe 2011. Convention on Preventing and Combating Violence Against Women and Domestic Violence. http://conventions.coe.int/Treaty/EN/Treaties/html/210.htm.

2 The European Commission's Directorate-General for Justice and Consumers, the European Institute for Gender Equality (Eige), the European Union Agency for Fundamental Rights (EuFra), and the European Institute for Crime Prevention and Control, affiliated with the United Nation (Heuni) also participated in the work of the Task Force.

1. MEASURING GENDER-BASED VIOLENCE[1]

1.1 Main indicators and topics to be addressed

The United Nations Declaration on the Elimination of Violence against Women (1993) defines Violence Against Women (VAW) as "any act of gender-based violence that results in, or is likely to result in, physical, sexual or psychological harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or private life".

The United Nations Fourth World Conference on Women[2], in the Beijing Platform for Action (BPfA) established three strategic objectives and the actions to be taken.

For the strategic objective D.2 - Study the causes and consequences of violence against women and the effectiveness of preventive measures - the actions to be taken by governments, regional organisations, the United Nations, other international organisations, research institutions, women's and youth organisations and non-governmental organisations, as appropriate, are the following:

a. promote research, collect data and compile statistics, especially concerning domestic violence relating to the prevalence of different forms of violence against women, and encourage research into the causes, nature, seriousness and consequences of violence against women and the effectiveness of measures implemented to prevent and redress violence against women;
b. disseminate findings of research and studies widely;
c. support and initiate research on the impact of violence, such as rape, on women and girls, and make the resulting information and statistics available to the public;
d. encourage the media to examine the impact of gender role stereotypes, including those perpetuated by commercial advertisements, which foster gender-based violence and inequalities, and how they are transmitted during the life cycle, and take measures to eliminate these negative images with a view to promoting a violence-free society.

The European Commission's 2010-2015 strategy for equality between women and men prioritised five key areas for action and in the priorities and key actions for 2016-2019, the Commission has reaffirmed its commitment to continue its work to promote equality between men and women, maintaining the focus of gender equality policy on the five existing thematic priority areas:

1. increasing female labour market participation and the equal economic independence of women and men;
2. reducing the gender gaps in pay, earnings and pension and thus fighting poverty among women;
3. promoting equality between women and men in decision-making;

1 Paragraph 1.1 has been drawn up by Roberta Barletta, paragraphs 1.2 and 1.3 have been drawn up by Maria Giuseppina Muratore.

2 The United Nations Fourth World Conference on Women, Action for Equality, Development and Peace, Beijing, China - September 1995.

4. combating gender-based violence and protecting and supporting victims; and
5. promoting gender equality and women's rights around the world.

Further improving the availability, quality and reliability of data on gender-based violence is one of the key actions planned for reaching objective 4: combating gender-based violence.

The European Parliament and the Council have repeatedly called for intensified efforts for collecting and compiling data. Furthermore, the Istanbul Convention introduces the obligation to conduct population-based surveys at regular intervals and to collect disaggregated relevant statistical data of all forms of violence covered by the scope of the Convention[3] in order to study its root causes and effects, incidences and conviction rates, as well as the efficacy of measures taken to implement the Convention.

Developing indicators of the extent of gender-based violence is the first step for monitoring and combating violence.

Review of the proposed indicators

In the BPfA definition violence against women encompasses but is not limited to the following:

- physical, sexual and psychological violence occurring in the family, including battering, sexual abuse of female children in the household, dowry-related violence, marital rape, female genital mutilation and other traditional practices harmful to women, non-spousal violence and violence related to exploitation;
- physical, sexual and psychological violence occurring within the general community, including rape, sexual abuse, sexual harassment and intimidation at work, in educational institutions and elsewhere, trafficking in women and forced prostitution;
- physical, sexual and psychological violence perpetrated or condoned by the State, wherever it occurs.

Acts of violence against women also include forced sterilisation and forced abortion, coercive/forced use of contraceptives, female infanticide and prenatal sex selection.

The Secretary-General's study (United Nations, 2006) addresses, inter alia, the following forms of violence against women: intimate partner violence; harmful traditional practices, including female genital mutilation/cutting, female infanticide and prenatal sex selection, early marriage, forced marriage, dowry-related violence, crimes against women committed in the name of "honour", maltreatment of widows; feminicide; sexual violence by non-partners; sexual harassment and violence in the workplace, educational institutions and in sport, and trafficking in women.

Considering that not all these forms are easy to investigate and not all of these forms are present in all countries, as a common starting point, physical, sexual, psychological and economic violence were identified as areas to be addressed through indicators that measure scope, prevalence, frequency, intensity and severity of violence.

In 2007, an expert group started to develop indicators aimed at studying the prevalence and incidence of violence[4]. Subsequently, the group of the Friends of the Chair of the United

3 The Istanbul Convention in Articles 33 to 40 identified eight forms of violence against women in need of criminalisation when it is "intentional conduct": psychological violence, stalking, physical violence, sexual violence, including rape, forced marriage, female genital mutilation, forced abortion and forced sterilisation and sexual harassment.

4 The expert group was composed by United Nations Division for the Advancement of Women, the United Nations Economic Commission for Europe (UNECE), the United Nations Statistical Division (UNSD) and experts from various countries.

Nations Statistical Commission on Statistical Indicators on Violence against Women was constituted in 2008 with the aim of defining and recommending indicators on violence against women, to be developed at the United Nations level.

The group defined core and optional indicators (United Nations, 2013). The core indicators relate to the issues of physical violence, sexual violence, domestic violence, economic violence, psychological violence and female genital mutilation.

Indicators concerning physical violence, sexual violence and domestic violence separately consider the reference periods of the last 12 months, and within the person's lifetime so far (asked from the age of 15); they are also articulated in terms of the relationship to the perpetrator, severity and frequency.

To better define domestic violence, two indicators regarding economic and psychological violence in intimate partnerships were added. These are very important because they contribute to a holistic view of domestic violence, but also because they are a very good predictor of physical violence.

Core indicators on violence against women (United Nations, 2013)

The core indicators are detailed as follows:

- Total and age specific rate of women subjected to physical violence in the last 12 months by severity of violence, relationship to the perpetrator and frequency
- Total and age specific rate of women subjected to physical violence during their lifetime by severity of violence, relationship to the perpetrator and frequency
- Total and age specific rate of women subjected to sexual violence in the last 12 months by severity of violence, relationship to the perpetrator and frequency
- Total and age specific rate of women subjected to sexual violence during their lifetime by severity of violence, relationship to the perpetrator and frequency
- Total and age specific rate of ever-partnered women subjected to sexual and/or physical violence by current or former intimate partner in the last 12 months by frequency
- Total and age specific rate of ever-partnered women subjected to sexual and/or physical violence by current or former intimate partner during lifetime by frequency
- Total and age specific rate of women subjected to psychological violence in the past 12 months by the intimate partner
- Total and age specific rate of women subjected to economic violence in the past 12 months by the intimate partner
- Total and age specific rate of women subjected to female genital mutilation

Some of these indicators were incorporated into the Global minimum Set of gender indicators (Who, 2014)[5].

The European Institute for Gender Equality (Eige) in the second edition of the Gender Equality Index 2015 (Eige, 2015) presented a first attempt at populating the satellite domain of violence by constructing a composite indicator of direct violence against women. This indicator relied on seven synthetic variables used by the European Union

5 Proportion of ever-partnered women (aged 15-49) subjected to physical and/or sexual violence by a current or former intimate partner, in the last 12 months and proportion of women (aged 15-49) subjected to sexual violence by persons other than an intimate partner, since age 15.

Agency for Fundamental Rights (Fra) in the European survey and agreed upon at the international level (Fra, 2014).

In 2016, Eige suggested further developing the composite indicator on VAW by including a dimension on disclosure/reporting of violence, a dimension on severity of violence against women through the inclusion of variables on health consequences of violence (injuries and long term psychological consequences) and by reflecting multiple. Eige proposed also an additional set of indicators of different forms of violence from administrative or survey data (Eige, 2016).

The United Nations Department of Economic and Social Affairs (UN Desa), in the 2030 Agenda, established goals and targets and recommended the development of comparable indicators across countries - the Sustainable Development Goal Indicators (SDG indicators).

Sustainable Development Goal Indicators (Goals 5 and 16)

Goal 5 - Achieve gender equality and empower all women and girls

5.2.1 Proportion of ever-partnered women and girls aged 15 years and older subjected to physical, sexual or psychological violence by a current or former intimate partner in the previous 12 months, by form of violence and by age

5.2.2 Proportion of women and girls aged 15 years and older subjected to sexual violence by persons other than an intimate partner in the previous 12 months, by age and place of occurrence

5.3.1 Proportion of women aged 20-24 years who were married or in a union before age 15 and before age 18

5.3.2 Proportion of girls and women aged 15-49 years who have undergone female genital mutilation/cutting, by age

Goal 16 - Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective accountable and inclusive institutions at all levels

16.1.3 Proportion of population subjected to physical, psychological or sexual violence in the previous 12 months

In November 2010, the Conference of European Statisticians (Ces) established the Task Force on Indicators of Gender Equality, to improve the monitoring of gender equality in the UNECE region by consolidating and systematising the existing proliferation of gender-relevant statistical indicators.

The volume Indicators of gender equality (UNECE, 2015) contains the result of this work. It presents a set of 115 gender equality indicators recommended for use in countries participating in the work of the Ces.

The indicators are grouped in eight thematic domains inspired by the Beijing Platform for Action and categorised into 42 headline indicators and 73 supporting indicators. The proposals are based on the consideration of policy needs, existing indicator frameworks, relevance to the measurement of gender equality and international availability. Whilst the work of the different United Nations expert groups focussed on measuring violence against women, the indicators of gender equality pertain to both males and females.

The United Nations Economic Commission for Europe defines gender-based violence as “violence directed against a person because of his or her gender or that affects persons of a particular gender disproportionately” (UNECE, 2015). Gender is a conceptual category that refers to social relations between men and women and this means, for instance, that not all violence suffered by men and women has to be included in the indicators. The

violence that should be represented in the data is the violence that is a consequence of different power relationships between men and women. This led the UNECE Task force to consider the importance of selecting those indicators that measure violence against both genders and that allow identification of the extent and profile of violence suffered by women and men respectively.

Studies carried out on violence suffered by women and men highlight that both of them suffer violence (Statistics Canada, 2005; Home Office, 2011). Whilst prevalence rates are indeed not so different between men and women, data on the incidence and consequences, show a very different picture, as studies have pointed out. The percentage of women suffering injuries and serious injuries, who felt their life in danger, who took time off from their everyday activities because of the violence is much higher.

Furthermore, for those women who were in partnerships, violence tended to happen or increased after the separation, while for men this is a very rare situation. This confirms the idea that it is quite important to focus on indicators on severity and on frequency in order to better measure domestic violence for both men and women.

Taking into account the work of the Friends of the Chair group (Foc), the UNECE Expert Group on Measuring Violence against Women (2007-2011) as well as the context of the UNECE Task force, a minimum set of indicators was proposed for measuring violence against women and men.

Minimum set of indicators (UNECE, 2015)

Physical violence

4.1 Proportion of women/men aged 15-49 years, subjected to physical violence in the last 12 months (*Violence survey required*)

4.1.1 Proportion of women/men aged 15-49 years, subjected to physical violence in the last 12 months, by frequency (every day or nearly every day, once or twice a week, once or twice a month, less than once a month), severity and relationship to perpetrator (*Violence survey required*)

4.2.1 Proportion of women/men aged 15-49 years, subjected to physical violence since age 15, by frequency (once, few times, many times), severity and relationship to perpetrator (*Violence survey required*)

Sexual violence

4.3 Proportion of women/men aged 15-49 years, subjected to sexual violence in the last 12 months (*Violence survey required*)

4.3.1 Proportion of women/men aged 15-49 years, subjected to sexual violence in the last 12 months, by frequency (every day or nearly every day, once or twice a week, once or twice a month, less than once a month), severity and relationship to perpetrator. *(Violence survey required)*

4.4.1 Proportion of women/men aged 15-49 years, subjected to sexual violence since age 15, by frequency (once, few times, many times), severity and relationship to perpetrator. (*Violence survey required*)

Psychological and economic violence

4.5.1 Proportion of women/men aged 15-49 years, subjected to psychological violence in the past 12 months by the intimate partner *(Violence survey required)*

4.6.1 Proportion of women/men aged 15-49 years, subjected to economic violence in the past 12 months by the intimate partner (*Violence survey required*)

Violence by the partner

4.7 Proportion of ever-partnered women/men aged 15-49 years, subjected to sexual and/or physical violence by current or former intimate partner in the last 12 months (*Violence survey required*)

4.7.1 Proportion of ever-partnered women/men aged 15-49 years, subjected to sexual and/or physical violence by current or former intimate partner in the last 12 months, by frequency (every day or nearly every day, once or twice a week, once or twice a month, less than once a month) (*Violence survey required*)

4.8.1 Proportion of ever-partnered women/men aged 15-49 years, subjected to sexual and/or physical violence by current or former intimate partner since age 15, by frequency (once, few times, many times) (*Violence survey required*)

4.9 Proportion of women/men aged 15-49 years, killed by a current or a former partner in the last 12 months (*Administrative sources*)

Under reporting

4.10 Proportion of women/men aged 15-49 years, subjected to sexual and/or physical violence in the last 12 months, who did not report the crime to the authorities/to the police (*Violence survey required*)

Violence in childhood

4.11.1 Proportion of women/men aged 16-49 years, subjected to sexual violence in childhood (occurring before sixteen years old) (*Violence survey required*)

Conclusions

Indicators should cover requirements of the Istanbul Convention. However, it is clear that not everything could be covered with a population survey and also limitations of sample size needs to be taken into account.

A set of indicators can been proposed in order to estimate the prevalence of Gender-Based Violence (GBV), taking into account the experiences and recommendations of European and international organisations.

About the forms of violence, the survey should include physical and sexual violence by any perpetrator and psychological and economic violence by intimate partner. For answering to the requests of the EC directive (2006/54/EC) sexual harassment at work should be included as core topic. It includes unwanted verbal, non-verbal or physical conduct of sexual nature which occur with the purpose or effect of violating the dignity of a person, in particular when creating an intimidating, hostile, degrading, humiliating or offensive environment.

Since the phenomenon is becoming more and more relevant, specific questions about stalking behaviours should be also included, at least with reference to those perpetrated by intimate partners.

Physical and psychological violence suffered during childhood by the parents is not necessarily associated with gender, but it represents a risk factor for being victim or perpetrator of violence in adulthood. For this reason, in a dedicated survey on violence, sexual, physical and psychological violence suffered by the respondents and by their intimate partners in childhood, could be additional topics and indicators useful for studying risk factors and multiple victimisation.

For covering other type of violence, as forced and early marriage, forced abortion, forces sterilisation, female genital mutilation, special sampling and even special survey could be required.

About the time frame of violence, even if indicators covering lifetime prevalence might be problematic, since respondents might not remember what has been happened during their lifetime, several experts supported lifetime prevalence indicators at least when the survey is done for the first time. Furthermore, in the perspective of regularly updated surveys to add the time frame of the last three or five years could allow one to compare prevalence in the period of time between two consecutive data collections.

Frequency and severity[6] of violence as well as disclosure, seeking assistance from health agencies and acts of reporting violence to the authorities should be included in the survey.

With regard to the personal characteristics of respondents, data on respondent's disability should be added to the list proposed in the United Nations Guidelines. Disability, indeed, is a strong risk factor for being a victim of violence.

With regard to the characteristics of the perpetrator other than intimate partner, at least age and ethnicity or citizenship should be added when the perpetrator is someone known by the victim.

The economic circumstances to figure out economic power, independence and economic imbalances could be important additional issues to include as background information.

Including attitudes and opinions on violence, instead, can be problematic. The answers to these questions, indeed, could affect the questions on experiencing the violence or vice versa. For this reason it is recommended not to include attitudes and opinions in this kind of survey.

1.2 A new framework for statistical data

Violence against women and gender-based violence are very complex phenomena and need of good tools and good definitions to be studied.

In 2011 the Council of Europe Convention on preventing and combating violence against women and domestic violence (the so-called Istanbul Convention), provided a common framework for violence against women as:

- a violation of human rights and a form of discrimination against women and shall mean all acts of gender-based violence that result in, or are likely to result in, physical, sexual, psychological or economic harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life;
- where "domestic violence" shall mean all acts of physical, sexual, psychological or economic violence that occur within the family or domestic unit or between former or current spouses or partners, whether or not the perpetrator shares or has shared the same residence with the victim;
- and "gender" shall mean the socially constructed roles, behaviours, activities and attributes that a given society considers appropriate for women and men;
- "gender-based violence against women" shall mean violence that is directed against a woman because she is a woman or that affects women disproportionately;
- "women" includes girls under the age of 18.

6 Unfortunately, there is not a single indicator to represent severity, so multiple variables are required. Among them, injuries suffered and type of injuries are very important, as well as the perception of one's own life being in danger, the victim's perception of seriousness, the victim's incapability to live her daily life because of violence. Furthermore, some acts such as rape, attempts to strangle or suffocate, to burn on purpose or the use of a knife or gun, are severe in and of themselves.

The Istanbul Convention in article 11 represents also a strong push to collect relevant statistical data at regular intervals and to conduct population-based surveys to assess the prevalence of and trends in all forms of violence covered by the scope of the Convention. But the attention to statistical data has its roots early in the years, in the United Nation Convention on the Elimination of all Forms of Discrimination against Women (CEDAW).

The need and the sensitivity for statistical data is in fact increased in the years, as demonstrated in the CEDAW General Recommendations (GR). The first in 1989 (GR12) emphasises the need to study the prevalence and incidence of VAW, the second in 1992 (GR19) focusses on "*encourage the compilation of statistics and research on the extent, causes and effects of violence, and on the effectiveness of measures to prevent and deal with violence*", while the 35 General Recommendation of CEDAW (2017) asks to "*establish a system to regularly collect, analyses and publish statistical data on the number of complaints about all forms of gender-based violence against women, including technology mediated violence, the number and types of protection orders issued, the rates of dismissal and withdrawal of complaints, prosecution and conviction rates as well as time taken for disposal of cases. The data should be disaggregated by type of violence, relationship between the victim/survivor and the perpetrator, as well as in relation to intersecting forms of discrimination against women and other relevant socio-demographic characteristics, including the age of the victim. The analysis of the data should enable the identification of protection failures and serve to improve and further develop preventive measures*". (GR 35, 2017: 49)

Furthermore, since 2015, UN asked data to monitor VAW for Agenda 30 through the Sustainable Development Goals (SDGs) that focus VAW in goal 5 "Achieve gender equality", and have some indicators in goal 16, "Peace, justice and strong institutions", and goal 11 "Sustainable cities and communities".

To combat and prevent VAW, the need is indeed to have a comprehensive and exhaustive data system to approach such complex topic. The Italian dedicated website on VAW[7], that Istat developed in agreement with the Department of Equal Opportunity (Agreement "Banca-Dati sulla Violenza di Genere" 2017) answering to the National Action Plans against Violence 2015-2017, 2017-2020, 2021-2023 is an effective response to the International recommendations.

Indeed, the Law 53 of 2022, on the statistical measurement of violence against women, obliges Istat to coduct the VAW survey every 4 years (art.2).

All data sources to study violence against women and gender-based violence are vital, but the main important and challenging data source are population surveys in order to measure the prevalence of violence and the dark figure of violence against women.

Population surveys are the best tool to measure and to monitor violence against women. Surveys allow to estimate the true number of victims of violence, including the unreported violence, that are the main part of violence it-self.

Violence against women surveys answer to many questions, for instance "how is widespread VAW in our country?", or "how is it changed over time?" or "is it decreased or increased?", or "is it different in some part of the Country", or "which forms and types of violence are most prevalent?" or "how much is violence reported to the police or to the law enforcement?"

More specifically, the main objectives of violence against women surveys are to measure:

– the prevalence/magnitude of violence and the characteristics of the partner and non-partner violence in terms of frequency of events, dynamics and peculiarities of different violence forms;

7 https://www.istat.it/en/violence-against-women; https://www.istat.it/it/violenza-sulle-donne.

- the place of occurrence;
- victims' characteristics and their reactions to the violence event;
- the violence awareness: *i.e.* victims' perception of violence as crime
- the physical, psychological and economic consequences of the suffered violence;
- the perpetrators' characteristics, especially for domestic violence;
- the seeking for support services (shelters) or healthcare services;
- the incidence of the dark figure of violence, and reasons why violence is not reported to police;
- the children witnessing of violence;
- the history of intimate partner violence: long term consequences;
- the risk and protective factors both at an individual and at social level;
- the intergenerational transmission of violence;
- the social and economic costs of violence (direct and indirect).

Studying domestic violence implies adopting methodological and procedural dedicated tools that take into consideration all relevant aspects with regard to the variables measured, way of posing questions, use of specific terms and specific training for female interviewers. Dedicated surveys on violence against women are designed in such a way that eases women so that they can build up confidence with the interviewer who is trained to deal with these cases. Not all forms of violence women suffer from, such as psychological and economical violence are acknowledged by women themselves; in this respect the instrument used to collect data needs to address the problem in all its aspects.

Warning in implementing a dedicated survey measuring 'sensitive data' is essential in a cultural context that even if changed and in constant evolution, it is still biased with regard to gender roles within the family context and domestic violence issues.

1.2.1 The international panorama about surveys

It is not easy to create an exhaustive inventory of Violence against women Surveys and Gender-Based Violence Surveys and the COVID 19 emergence situation has stimulated the implementation of new dedicated studies on violence against women during the pandemic.

In 2017, Istat counted 136 researches on the topic carried out by different organisations (National statistical offices, Ministries, Research centres, NGO etc.) of 88 countries. In particular, 19 studies were carried out by European Countries at 28, the first survey was conducted in 1988 in Belgium. In addition, there is the "Multi-country Study on Women's Health and Domestic Violence against Women" carried out in many countries and supported by the World Health Organization (Who) and the United Nations Fund for Population Activities (Unfpa). While in Europe, the European Fundamental Rights Agency (Eufra) carried out the VAW Survey in 2012 that involves the 28 EU Countries. Furthermore, in 2019 Eufra carried out a survey on crime and violence in order to measure some aspects of physical and sexual violence that affect men and women, always in Europe.

Also Eurostat is currently coordinating a European survey on gender-based violence and other forms of interpersonal violence. Some EU Countries have already conducted the survey in 2020 and others in 2021. Results are expected in 2023.

Existing surveys differ from each other in several points: contents, methodology, structure and target. The varieties of existing surveys are complex and interesting. For example, there are *ad hoc* survey, module and singles questions addressing violence; some surveys are periodical, some occasional and some other annual.

Furthermore, some methodological and strategic choices are not so dissimilar between countries. This is the case of questionnaires' structure, the violence definitions used, the high attention put on quality, ethical and safety considerations.

1.2.2 An overview of survey's characteristics

The most part of the surveys are dedicated surveys, mainly quantitative, one-shot or occasional or repeated at least every 4/5 years. While in some countries specific modules about VAW or GBV are inserted in other kind of surveys such as health survey or victimisation surveys, only few studies regards general social survey or demographic survey.

But dedicated surveys differ from each other too: VAW or GBV surveys cover sometimes the entire range of possible kinds of violence, while other times they address only some questions regarding physical violence/physical assault and sexual violence.

Who conduct the surveys?

National Statistical Offices (NSOs), alone or in partnership with other agencies, are very often responsible of the studies carried out in Europe. When conducted in partnership, NSOs share with other partners their competencies about data collection, methodology and the sample design.

Private institutions, Non-Governmental Organisations (NGO) and universities are the surveys 'responsible too, while in other cases the responsibility is shared between ministries, governmental departments, and universities, in collaboration with national statistical offices or sometimes the United Nations.

Periodicity

As noted above, most of the survey are occasional or periodically updated. The very few annual data collections are module or simple questions on physical violence inserted in health or living conditions surveys and few dedicated surveys are periodically updated, as well as some dedicated modules.

Nevertheless, a positive aspect emerges analysing the studies by single countrie. A country starting to address the violence topic in an accurate way, repeats and improves the study and, very often, includes it in a road map that seems to lead to the definition of a periodicity and its implementation.

A push for a better periodicity come from the ratification of the Istanbul Convention, at least for the European Countries.

Since this social phenomenon do not change so fast, in comparison to other topics, the best periodicity seems to be every 4 years in several.

Target population

Most of the studies collect data on VAW mainly following a gender perspective, while only few include questions regarding men too.

Generally, the questionnaire for men and women is the same; the problem that arises in these cases is that the violence suffered by men is not specifically gender-related. Results show data by type of violence and type of perpetrators; some of them focus also

on frequency and severity. The question if the violence suffered is gender-based is not satisfactorily solved, not only when victims are men, but also for women this aspect is not completely well solved. Identify gender-related violence suffered by men is particularly urging in order to apply a harmonised solution in all the GBV surveys.

The majority of the studies interview persons from the age of 15 or 16 years. In some cases, adult people are interviewed (from the age of 18 or 20 years).

A superior threshold can also be found, this is very different from a survey to others, and it varies from the age of 49 of the Who studies to 84 years while in some cases there is not a superior limit at all. Generally, the target population are persons resident in the country, but sometimes there are specific sample for foreigners and in two or three studies, social groups suffering specific risks of marginalisation are considered (homeless, prisoners, prostitutes).

Methodology

Also about methodology, and in particular the interview technique mode, countries present very different approach. Many of them adopted mix-modes, where more combinations are used: Cati and Capi, Capi and Casi; Cawi plus Capi plus Cati are the most frequent.The most common interview technique is Capi, followed by Cati and only a few surveys are exclusively online or postal.

The response rate fluctuates from 30% to more than 90%. Some surveys deliver data regarding only the refusal rate and others the total non-response rate due to ineligibility or list coverage problems, anyway the rates are very different. The sample designs are very different, but a random sample has been planned almost everywhere. No random sample are considered especially for qualitative studies or when the focus is on marginalised target, as prisoners, homeless, prostitutes. The sample units are persons and households. The sample size goes from 1.000 to more than 70.000 units.

Reference period

Reference periods for indicators are generally related to life time and the last 12 months. Very few studies deliver only the last 12 months.

The life time period considers diverse situations. In some surveys life-time is a mix of “since the age of 16 teen”, or “since the age of 15 teen” for non-partner violence and “during the relationship”, if the questions addresses intimate partner violence. While in other cases, the lifetime includes also the childhood period. Some studies deliver also data for the last five or three years, in addition to the life time and the last 12 months.

1.2.3 The violence panorama as depicted by the surveys: forms of violence, characteristics of violence, consequences, risk factors, disclosure

Forms of violence

Almost all studies collect data on physical and on sexual violence. Psychological violence is asked in the half of studies, economic violence less frequently. Violence in childhood is asked in one third of cases, followed by sexual harassment. Stalking instead is asked very rarely.

Some studies highlight the importance to collect data on violence in labour sphere, violence related to school, violence in the community.

The context and dynamic of the violence episodes

All studies collect data about the frequency of violence, how often it happens, even if they differ about the way to ask this information.

The localisation, the place of occurrence is asked too, as well as the kind of perpetrator (partners, relatives, friends, acquaintances...).

Other information regard physical violence suffered during pregnancy, the alcohol consumption by perpetrator and victim (only in some studies for the second one) in the dynamics, the use of contraceptives during the sexual abuse, the arms involvement, the victims' reactions to violence (also in terms of physical violence used to response to assaults or other strategies used during the violence episodes).

About consequences

Many evidences are showed by different surveys about severity of violence and violence consequences.

Women's physical consequences are addressed almost everywhere as injuries suffered, type of injuries (included miscarriage) and the need for medical assistance; the risk of acquiring a sexually transmitted disease (STD) and disability as a consequence of violence are less frequently addressed.

The mental health effects of violence on women and the psychological consequences of sexual violence are also addressed. The progress and impact of domestic violence and the elements to individuate a post-traumatic stress disorder are considered, as well as the alcohol use and others coping strategies.

From the economic point of view, the job loss, the finance value of emotional harm, the cost of violence, also linked to the separation or the leaving home due to violence are considered. Nevertheless, not many studies collect data on these aspects.

In case of children in the family, data on presence and impact of domestic violence on children is collected.

Perception of violence

Victims' perception of seriousness is collected and it represents the subjective side of severity.

Data on fear are also very important. This aims to understand the perception of victim about her own life in danger during the violence's event, and it is an important factor for the risk's assessment.

Other aspects regard awareness about domestic violence. This aspect, usually measured through the victims' perception of violence as crime, is very important in order to monitor the arising in victim's awareness and the social change.

Disclosure

The disclosure issue is one of the main topic addressed in surveys: to whom the survivors open themselves (if they do), reporting to the police and satisfaction with police response, seeking help outside of the criminal justice system.

Also experiences with support services are investigated, as well as contact with healthcare professionals and health services.

Furthermore, in some studies the attention is also on the evaluation and the response effectiveness of services (public and private), health and social services, shelters and crisis centres.

Risk factors and intergenerational transmission

Under this topic, information regards the history of family violence are considered the neglecting and punishment experiences in childhood, violence in family, as victim or as witness and the sexual child abuse are considered.

Some risk factors are also related to abuses suffered by the woman's partner in childhood and violence in partner family. Other risk factors are related to the partner behaviours, as alcohol use/abuse, partner's physical or verbal violent behaviour also outside the family and so on.

Risk factors associated to health problems are measured too, as well as violence in the educational system, in workplaces and in the community.

Variables regarding health, socio economic status, the overall context

The information collected concerns the socio-demographic characteristics of the surveyed population, as sex, age and civil status, the housing characteristics of the households, the marriage history of women, their health situation and disability, the women independence (autonomy, use of free time) and the economic situation.

Some studies put more attention at on the health aspects, not only to the general health, but also to women behaviours (including alcohol or drug consumption), women's sexual and reproductive health and behaviour, perceptions of sexual autonomy.

In one study the focus was on problems of children in the family and on social discrimination.

Some studies include attitudes towards gender roles, attitudes towards violence, and prejudices about violence against women.

1.2.4 Best practices

Some best practices can be identified on the methodological approach in carrying on surveys on violence topic. In the following paragraphs some of them are shortly described.

The informative letter sent before the survey

The use of the advance letter is very useful to stimulate a positive attitude towards the survey among selected interviewed people.

In one study there was a complex way to use advance letter: it was sent a letter in general terms, about the purpose of the project, the survey approbation by "Institution"

and privacy guarantee, the fact that the participation was voluntary, the time for contacting, and participants could opt not to participate by contacting the responsible agency. Then, another letter was sent with more details, communicating who contact in case of problems or for speaking about feeling that might arise when completing the questionnaire.

A critical point seems the fact that this letter explicitly said that the persons "could opt not to participate by contacting the responsible agency ", since it can be counterproductive. In the direct contact between the interviewer and the interviewee, many uncertain situations can be in fact recuperated.

In another study four reminder letters were sent.

When using online questionnaire, the letter also provide respondents with username and password for the login to ensure privacy and secret access to the internet.

Interview mode

Since the considerable amount of literature already available on this issue, only some notes on particular aspects will be provided here[8].

In one study, the telephone interviewing was chosen partly because this procedure allows for follow up questions for those who report exposure to violence.

Respondent were asked if they were willing to be contacted again, the 90% answered positively with no significant difference between those exposed to violence and the non-exposed ones. This was considered as indicator that the procedure was acceptable for participants.

Multiple techniques are used according to respondents' preference: Casi, Audio-Casi, self-completed paper questionnaire, Capi. A study used both qualitative and quantitative technique to help disclosure and help women to feel much more at ease.

Some studies underline the goodness of use several interviewing modes, the respondent in fact in this cases could choice the most suitable for him/her. Nevertheless, in this situation, it is not completely clear the impact of the use of multiple modes on the results: response rates, non-response rates and prevalence rates, since many factors are involved.

Since the mix-modes are quite used, a great attention has to be put on the impact of several techniques on the data results.

Target population

Some more vulnerable groups in our society can present differences in the level of risk, as well as different difficulties in disclosing sexual violence and seeking for help, for these reasons it is very important to have a specific glance at them.

Some studies include specific sample for migrants and others conduct researches on marginalised groups: homeless, prostitute, people with learning disability, prisoners and patients in psychiatric settings. The groups described were considered with different methodologies and levels of involvement from the classic VAW surveys. The reason for the use of different procedures and techniques in interviewing these groups concerns both pragmatic issues and the awareness that there are differing perspectives on abuse in marginalised groups. Another important topic regards the possibility to understand if violence occurred before or after the marginalisation, if it is a cause or a consequence.

8 For a better comprehensive analysis of this topic see chapter 5.

Attention to methodology is essential, one-to-one interviews with such a vulnerable group requires extensive skill and resources, including time.

They require also specific questions (for instance type of perpetrator), specific forms of violence, as forced or arranged marriage and above all specific analysis.

In one study refugees have been interviewed too. This was only a very difficult experiment, that did not really achieve its task due to the many problems faced (from the language issues to the cultural barrier and the fear for own destiny).

About the "age" of survey population, a study underlines the difficulty in comparisons if the age of the target population is not the same everywhere, since the time of exposure to violence is diverse, as well as the memory mechanism of recalling at different age and, it is possible to add, the disclosure level could be different for different generations.

One possibility, in this case, can be at least to perform analysis by the same age group.

Quality

Problems on data reliability, validity and generalisability were considered in the studies: limits to data quality have to be not ignored, nor exaggerated, but taken into consideration.

Many surveys put a lot of attention to quality and quality monitoring. Tools are different and vary from the quantitative to the qualitative ones. Some researches adopted a dedicated monitoring system.

In some studies, data regarding reasons for refusals as well as interviewee availability, sincerity and difficulties in answering questions are collected, as perceived by interviewers, at the end of each interview. The information about the context are also very useful to monitor quality and to understand better the survey results.

Another survey used feedback questions asked directly to respondents. Respondents were asked for feedback regarding the main questions: about the sensitiveness of the covered topics, the burden, the memory problems, the difficulty in understanding, the misleading and event duplication.

Attention is given to memory problems too: the less serious event can be forgot, the most serious one can be removed or there can be reticence in speaking about it. These are the main reasons for asking the interviewees to take their time to answer and for asking separately about each violence form.

Ethical safety procedures

It is pointed out that a system to take care of individuals who may be in need of support after completing the interview is required. Even if few individuals needed or wanted such assistance, for those who did need it, a follow-up system is important.

Sometimes leaflet for the respondents have been produced. These are useful to reduce the possible negative consequences of the survey. General information on intimate partner violence and information about who contact in case of need, where to found which kind of help and the victim support agencies are provided.

Other measures regard the use of appointments (call back or other visits) to ensure privacy during the interview and the help-line that can be called if the respondent feeling distress during an interview.

Effects on interviewer and possible helps in order to cope with the burn-out

First, a good training and a good support is essential to take care of interviewers. The recruitment aspect is also basilar to define if an interviewer is appropriate to carry out this kind of survey on so sensitive topics.

About the burn-out, in one study at the end of each survey, the interviewer was given a chance to rate how the interview had gone (from "very difficult" to "very smoothly") and to note how she was feeling (from "not upset at all" to "very upset"). This and other similar strategies are useful to let go off emotions.

In studies the multifaceted role of interviewers is underlined: combination of working as a salesperson (to gain maximum participation), a confidant (to ensure a high level of trust in order to allow the participant to be honest), and a neutral researcher (to ensure that the results are not biased in any way). The last one seems to be the most difficult aspect due to the emotional distress linked to traumatic stories reported. A help is the possibility to give the help-line number or the shelters references to victims.

Documentation

Some studies are very accurate when presenting their results. They explain choices and reasons behind them; they describe the process and the monitoring phase; they document the quality reached.

Furthermore, some studies present rates both on total non-response and item non-response indicators and perform item non-response analysis.

Finally, some researches conduct interesting comparisons at analysis level between survey's data and data from shelters.

1.2.5 The questionnaire

Structure and design

The use of screening is common in most of the studies.

Screenings for each kind of perpetrator is the best choice, at least for non-partner, current partner and former partner.

It is basilar the event-based description of violence. It is useful for physical violence to graduate items in a way similar to the conflict tactics scale.

The importance of wording is underlined in some studies. The formulation has a great importance because it can help the disclosure. But it has to be used carefully, since it can lead to underestimation or overestimation.

About the violence details (dynamic, consequences, etc.), surveys use to collect data regarding the last incident, since the number of incidents can be so large that not all of them can be described in detail in a usual interview. The use of the last incident may underestimate the victims' health consequences, if the victims have had many victimisation experiences and the majority of them are less severe.

Asking details of the most severe incident is the other option, but this method might overestimate the severity of the consequences, and remembering the details of the most severe episode might be difficult to the respondent, if the most severe incident has happened a long time before the interview. Furthermore, there is a problem of subjectivity

in determining which is the most severe episode. These facts have a strong impact in order to have a representative overview of the phenomenon, especially when comparing data from different surveys.

About contents

Violence definition of sexual harassment at work: in one study, sexual harassment is considered in a broader view as a behaviour that fall along the continuum of sexual abuse and violence. In this sense, discrimination can be only one of the possible consequences of this experience.

Some examples of sexual harassment at work: received crude and offensive remarks; humiliated by sexual gestures; made uncomfortable by sex talk; shown pornography; felt uncomfortable by someone exposing themselves; been touched in uncomfortable way; stalked; treated badly for refusing sex; bribed to engage in sex; threatened if not sexually co-operative.

Socio-demographic variables and not only

Some studies collect less common background variables on sociodemographic information: they add psychosocial condition during childhood, aspects of nurture, school satisfaction, relations with parents and friends, antisocial behaviour, self-harming behaviours and ages at which respondents first encountered alcohol, drugs and sex, and include the homosexual partner.

About childhood violence

Some researches collect data on physical violence (hair pulling, pinching, slapping) suffered in childhood: sex of perpetrators frequency and length of the abuse, the precise age of the abuse, if the person thinks that the abuser did the same to other boys/girls too.

Other related topic

Some surveys focus also on attitudes towards violence. They investigate attitudes and prejudices; public perception of the prevalence of sexual violence, public perception of media role in speaking about violence.

1.2.6 Critical aspects

About women and men violence

Studies that collect data on both women and men do not stress enough gender-based concepts. Analysing data, prevalence is not so different between men and women; in some cases, the physical violence suffered by man is even higher; some differences arise in the intensity/frequency or seriousness of violence, about the context/location where violence happens or regarding the perpetrator typologies.

Moreover, studies show a problematic aspect concerning the men awareness about women's violence against men and its impact on the underestimation of prevalence rates. The reason is linked to the fact that men are ashamed of disclosing their victimisation or they do not look at it as important and for this reason do not notice it. While on the other side, men seem more ready to answer to questions on sexual violence.

Other studies put the attention on the possible underestimation of violence suffered by men, highlighting that surveys are unable to capture the most severe and the less severe cases of violence. From one hand, it is assessed that the less severe cases remain hidden because in the everyday life of men a certain kind of violence is quite so common and it is not really considered to be violence (for instance violence in public places that can be defined as an argument/conflict or a fight, or physical/corporal punishment by the boy's parents). On the other hand, about the most severe cases, it is underlined that men do not want to see themselves as defenceless victims. There are taboos that prevent men to reveal their violence's stories. Such cases are felt to be shameful, and therefore men do not want to remember them or talk about them. This group of events comprises, for instance, sexual harassment, rape, and partner violence.

These observations show that there is a lack of conceptualisation of the gender related violence topic and highlight the need to go more in depth in studying men as victim of violence.

Attention to definition that can be too much strictly linked to the law

A study highlights the direct relationship between the number of violence forms explored in the questionnaire and the higher violence rates.

For instance, it demonstrates that using other forms of undesired sexual acts and coercion in the definition, it is possible to calculate prevalence for a broader definition of violence as well, including undesired or forced sexual activities using psychological or moral pressure. The study conclusion is that "This shows quite clearly how wide the range of prevalence can be, depending on the scope of the definition". This is true, but at the same time it is very important to collect data that are linked to *real or supposed real* (*as culturally specifically defined*) violent behaviours.

Others problematical aspects regard the low response rate presented by some studies and the non-comparability of studies, the generalised lack of information on marginalised groups.

About marginalised group: sometimes samples do not cover all the population and are not able to give good estimates for women living in rural and remote areas and not speaking fluently the official country language. Some surveys present also coverage list problems. In one study is underlined the problem to reach women living in hostels and emergency shelters.

About the *non-comparability*, it is underlined as a problem also within countries which promote more than one survey, when the strategies adopted to carry out any single survey are very different. This is the case of results, for instance, of indicators that are not comparable because they are created on the basis of different questions in input: in a survey they were the synthesis of a lists of 17 acts of physical abuse and 24 acts of sexual abuse; in another survey although the indicators covered the same forms of abuse, they were collected in far less detail (only three questions on physical abuse and one generic question on sexual abuse).

Another aspect regards the *reference period*. It is recommended to focus not only on the last 5 years, especially in order to examine victimisation from former partners. Some of the consequences of the acts of violence can last a long time and still deploying their effects (on

health, or from a judicial point of view) even if violence happened more than five years ago.

Attention to *definition of intimate partnership*: if the partner violence focusses on marital and common-law (de facto) relationships, it does not conform to current accepted definition of 'intimate partner violence' which also includes dating relationships.

About *disclosure problems*: there can be a risk of underestimation. It is likely that not all women fully disclose their experiences regarding domestic abuse to the interviewer out of fear, shame or denial. Furthermore, women may be more likely to report abuse suffered in a past marriage or relationship than abuse in a current marriage.

About *item non-response*: some partial non-responses were likely to have been caused by the difficult structure and the length of the questionnaire.

About the *response rate*: a study reports that the response rate for the telephone interviews and the internet survey correspond to the response rates achieved on average on other studies with these methods while, due to refusals, the response rate in the face-to-face interviews was lower than on average face-to-face interviews. It was mainly due to the short period allocated for the fieldwork, and to the fact that the interviews were conducted just before Christmas. It is important indeed to choice the right period and timing for the interviews.

1.3 Target population: sex

The Istanbul Convention (Council of Europe, 2011) introduces the obligation to conduct population-based surveys at regular intervals and to collect disaggregated relevant statistical data of all forms of violence covered by the scope of the Convention in order to study its root causes and effects, incidences and convictions rates, as well as the efficacy of measures taken to implement the Convention (article 11).

In the preamble and in articles 1, 2 and 3, the scope of the Convention is clear and the focus is on violence against women. Only "gender-based violence against women" is defined: "Violence that is directed against a woman because she is a woman or that affects women disproportionately" and it is assessed that "women and girls are exposed to a higher risk of gender-based violence than men" and that "domestic violence affects women disproportionately".

The American Psychological Association (APA) noted that violence against women remains fundamentally a learned behaviour that is shaped by socio-cultural norms and role expectations that support female subordination and perpetuation of violence (APA, 1999).

Violence can be understood both as a consequence of sexism and as a type of sexist behaviour causing other subsequent consequences. Brownmiller (1975) was one of the first to argue that sexism was a central cause of violence against women when she said that rape was a form of male dominance serving to keep women in a state of fear. Consistent with this, young women report that fear of rape is one of their most prominent fears (Hickman & Muehlenhard, 1997).

Violence against women is a vast social phenomenon both in terms of prevalence and of the multitude of forms it takes. Estimates about the prevalence of violence against women vary according to the type of violence examined and the type of reports examined. Yet, the data point out the pervasiveness of male-perpetrated violence against women, especially when considered in cumulative form across various forms of violence.

Furthermore, violence against women may result in extreme consequences: the femicide[9]. In Italy, for instance, 126 women were killed in 2022: 48.4% were murdered by their partners and 34.1% by relatives (including parents and children). In 2022, estimated femicides where 106 on 126 homicides with a women as a victim; in 2021, 104 on 119, in 2020, 104 on 116. Data are stable over time[10].

In order to comply with article 11 of the Istanbul Convention, referring to data collection and research, the focus should be on women since men are less affected than women are by gender-based and domestic violence.

The gender of the target population was one of the main issues debated in the Eurostat Task Force on the development of a survey on gender-based violence: should the survey be addressed only to women or both to men and women? Aside from the decision of the EU Task Force, where most of the countries expressed the desire to interview also men, the issue calls for reflection.

If we assume that the definition of GBV against women have a corresponding meaning in the definition of GBV against men, we should investigate "the violence that is directed against a man because he is a man or that affects men disproportionately". The question is: can we investigate this kind of violence with the same questions usually asked to women or, instead, a different questionnaire should be planned?

1.3.1 Surveys on Violence against women

The questionnaire for surveys on gender-based violence against women usually covers a series of threatening and violent acts and situations by the current and/or former partner. The average tool of measurement of intimate partner violence (IPV) is an adaptation of the famous 'Conflict Tactics Scale' (CTS), with which both members of a couple in family studies were usually interviewed (Straus, 1979). The CTS is based on the implicit assumption of equivalence between acts of violence perpetrated by a male on a female and vice versa[11].

Since the eighties, a lengthy controversy about the nature of male victimisation by female partners has been developed. According to Straus, empirical evidences show similar (or even worse) rates of victimisation for battered men, victims of their female partners.

Looking from the opposite perspective, this point was conceptualised by scholars who differentiate by average male violence (some families suffer from occasional outbursts of violence from either husbands or wives and this is the so-called "common couple violence") and patriarchal terrorism (in which families are terrorised by systematic male violence) (Johnson, 1995). The authors argued that the distinction between common couple violence and patriarchal terrorism is important because it has implications for the implementation of public policy, the development of educational programmes, intervention strategies, and theories of interpersonal violence.

---

9 "Femicide is generally understood to involve intentional murder of women because they are women. Femicide is usually perpetrated by men, but sometimes female family members may be involved. Femicide differs from male homicide in specific ways. For example, most cases of femicide are committed by partners or ex-partners, and involve ongoing abuse in the home, threats or intimidation, sexual violence or situations where women have less power or fewer resources than their partner". WHO (2012) Understanding and addressing violence against women. In: https://apps.who.int/iris/bitstream/handle/10665/77421/WHO_RHR_12.38_eng.pdf?sequence=1.

10 Italy, in adoption of the UN Statistical Framework to measure femicide, Istat started to estimate femicide. https://www.unodc.org/documents/data-and-analysis/statistics/Statistical_framework_femicide_2022.pdf.

11 The Index of Spouse Abuse (ISA), instead, is constructed for studying men violence against women (McIntosh, S.R., W.W. Hudson, 1978).

In this perspective, the process of changing attitudes must be located within a project of changing familial, organisational, communal, and societal norms that support violence against women. Interventions must address not only those attitudes that are overtly condoning violence against women but also the wider clusters of attitudes related to gender and sexuality that normalise and justify this violence. The efforts to address violence-supportive attitudes must also work to provide an alternative, a set of norms and values centred on nonviolence and gender equality.

Thinking that violence has the same meaning does not help in planning the best policies for violence against men that generally require safety policies, and the best ones for women, that generally require cultural and social policies. It is true that the National Statistical Offices (NSOs) do not have to plan policies, this is not their duty, but they have to produce unbiased data useful for policymakers. NSOs have the responsibility of collecting good data and properly disseminating them. The information has to be complete and meaningful.

The Italian example from "Homicides Victims, 2019"[1]

Data show that homicides decreased strongly from the Nineties, especially those with male victims. In 1990, one woman was killed for every five men, while in 2019 the proportion is of one woman for less than two men (111 women compared to 204 men). This is due to the decrease of mafia and organised crime homicides where the victims are mostly men.
However, this information is not complete and not meaningful unless other data is also taken into consideration, as follows:
in 2019, 68 women (61,3%) were killed by their actual or previous partner. In the same year, only 11 men (5,4%) were killed by their partners and this distribution is basically stable over time.

1 Italian National Institute of Statistics - Istat. 2021. "Perpetrators and victims of homicide. Years 2018-2019", https://www.istat.it/en/archivio/254589. For year 2022, data are available in Italian https://www.istat.it/it/archivio/291266.

For research purposes, it would be ideal to have scales which use the same items to assess both female-to-male and male-to-female violence. Nevertheless, when interviewing men with the same questions, there is a high risk of reducing gender differences and supporting unrealistic interpretations of gender similarity. Especially if the results are communicated or interpreted in a simplistic way, for example, by focussing on prevalence alone. The questionnaire needs to include elements of severity, frequency and consequences of violence that have to be taken into account to discover the differences.

Although it is demonstrated that a woman's initiative in domestic fights is often traceable either as self-defence or as a reciprocal pattern or even as offense - see, for instance, Graham-Kevan and Archer, 2005 - it is very rare that it can bring about serious damages. There is always a bias in counting something as an act of violence without taking into account its gravity.

Furthermore, the severity behind the same act of violence can be very different for men and women, as well as its cultural meaning. For instance, to refer to how "frightening" the physical violence was, it does not help because the feeling of fear is linked to different cultural representations and different vulnerability profiles of men and women.

Violent acts also are likely to differ in meaning and implications depending upon the gender of both perpetrator and victim (Arias and Johnson, 1986; Makepeace, 1986; Stets and Pirog-Good, 1989). The act can be the same, but the causes of violence against men and women are different as are the motivations and the consequences; therefore, different

questions are needed, as well as different indicators to be disseminated. Otherwise, the risk is the underestimation of gender imbalance in domestic violence and in violence in general and even the underestimation of the violence suffered by men.

A male sample can be useful in order to compare data between women and men, but this kind of study would concern the volume and extent of violence, but would not detect the gender-based violence against women and men.

An example thereof is the US National Violence Against Women Survey that compares the prevalence and consequences of violence perpetrated against men and women by spouses and opposite-sex cohabiting partners, by using data from a telephone survey of 8,000 men and 8,000 women (Tjaden and Thoennes, 2000).

The study found that married/cohabiting women reported significantly more rapes, physical assaults, and stalking perpetrated by their partners than married/cohabiting men, regardless the time considered was the respondent's lifetime or the 12 months preceding the survey. Women also reported more frequent and longer-lasting victimisation, fear of bodily injury, lost work time, injuries, and use of medical and mental health services, and justice system services. This survey does not address gender-based violence against women and men and the male sample is only used for comparison purposes.

The existence of female violence against men does not cancel the differences between violence inflicted by and suffered by men and women. In fact, the same act of domestic violence can have different meanings when it involves different genders, and it should probably be revealed with different instruments. Otherwise, there is a false hypothesis of equivalence and symmetry.

Indicators should also measure the context of violence suffered by men and women, since they might prove to be fundamental in order to understand gender differences in violence victimisation and perpetration. Measuring the context means also looking at the reasons for the violence/ "motivation for violence". This is not an easy issue to address in a population survey, since only the victim's point of view is considered. Nevertheless, this variable is included in some surveys and it is considered as a good approximation of a real situation.

Below there is an example of a question asked in the Italian national victimisation survey about the motivation of threat[12]. Other examples can be derived by surveys on discrimination and bullying.Only by analysing this information, violence against both men and women can be fairly well represented. The drawback is that the cross-tables between several variables (for instance perpetrator type, form of violence, severity, who starts the violence, place of occurrence) entail a great number of dimensions to be crossed. In order to do so the sample needs to be sufficiently large; otherwise, the resulting data may not be representative, and consequently the data collected on men may be useless as it generally presents low numbers regarding sexual violence, IPV, etc.

Summarising, in order to fully understand the differences between men and women, the questionnaire needs to include:

- separate measures of power and dominance within relationships, through the use of existing standardised and validated scales[13];

12 The Italian national victimisation survey, the Citizens' Safety Survey, is a periodical survey carried out every 5/6 years by Istat on a sample of 50,000 interviews on individuals 14 years old and over. In the survey, the motivation is asked only about threats.

13 See, for example, "The Relationship Power Inventory" (Farrell et al., 2015) and "The Controlling Behaviours Scale" (Graham-Kevan, 1999).

According to you, for what reasons were you threatened?
(*Multiple* answers possible)
1. To embarrass / humiliate me / out of envy
2. To get money
3. For problems of the offender (he /she was drunk, crazy, drug addict, etc.)
4. For condominium disputes
5. For problems at work
6. For political differences
7. For problems of love, relationship
8. For sexual intentions
9. For racial prejudices
10. For anti-gay / anti-lesbian behaviour
11. The threats originated from a fight / dispute started in traffic
12. Because the respondent is elderly
13. Because the respondent is disabled
14. Does not know, does not remember
15. Family arguments
16. Other (specify)

- the frequency of victimisation and perpetration of different violent acts (repetitive nature), with response categories adjusted to each type of violent act (number of times each act of psychological, physical, sexual, other types of violence happened during the defined period);
- the severity of the acts of violence (in terms of their likelihood to cause psychological and/or physical harm);
- identification of who initiates each violent episode and type of relationship between victims and perpetrators for each referred violent episode (especially for intimate partner violence);
- information on the dynamic (where did it happen);
- information on the motivation;
- a specific set of items to measure when violence is used in self-defence.

Furthermore, some questions should be asked only to men and others only to women, in order to detect the peculiarities of violence experienced by male and female victims.

Great care in how to address the issue of violence must be taken. Studies show a problematic aspect concerning men's awareness about women's violence against men.

There are taboos and stereotypes that prevent men from revealing the stories of the violence against them. At the same time, men are less prompt to report even minor episodes of violence since, in their everyday life, a certain kind of violence is common and it is not really considered violence.

1.3.2 Gender-based violence against men

Certainly, some forms of violence against men, in which there is a power imbalance, exist, but this kind of violence, *i.e.* sexual and physical violence aimed at controlling and keeping a man in a state of subordination, is limited to very specific situations and it is perpetrated mainly by males against males. It can happen in the army, in jail, in war situations against enemies and prisoners, and similar situations. This kind of violence disproportionally affects men and it is directed against them because they are men, then

it can be defined as "gender-based violence against men", but it cannot be collected by a population survey where a small sample of men are interviewed or with a questionnaire focussed on intimate partner violence, sexual harassment or stalking.

These situations are so rare that an *ad hoc* questionnaire and sample should be envisaged.

The neutral approach is not useful. A survey with a real gender-based approach should use "the gender lens" to formulate questions and should include different questions for men and women, because they experience different forms of GBV, of GB discrimination and different social roles and behaviours.

Unfortunately, no consolidated methodology incorporating the gender lens has yet been developed.

The Italian experience

Just as an example, in the last Italian national survey on victimisation, a module on sexual harassment and sexual blackmail at work took also men into consideration for the first time. About 40,000 interviews were carried out in 2015-2016 with Capi and Cati techniques on individuals aged 14-65.
The same questions on sexual harassment were used for women and men:

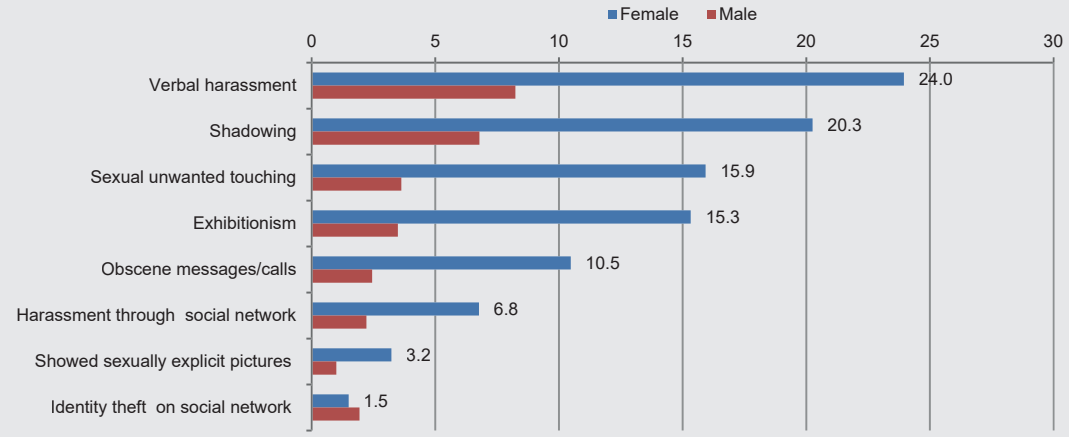
- verbal harassment
- sexual unwanted touching
- shadowing
- exhibitionism
- obscene messages/calls
- sent/showed porno/sexually explicit pictures, photos that make the person feel offended
- harassment through social networking websites
- identity theft on social networks with the aim of offending other people

Since in the pilot survey (2015) very little data was collected on sexual blackmail at work for men, it was decided not to explore further this topic in the full-fledged survey, because data would not be representative in spite of the sample size (around 20,000 male respondents).

About sexual harassment, the results show that 43.6% of the women, almost 9 million, and only 18.8% of the men (almost 4 million) experienced at least one form of sexual harassment in their lifetime; 15.4% of the women (≅3 million) and 6.4% of the men (≅1 million) experienced sexual harassment in the last 3 years.

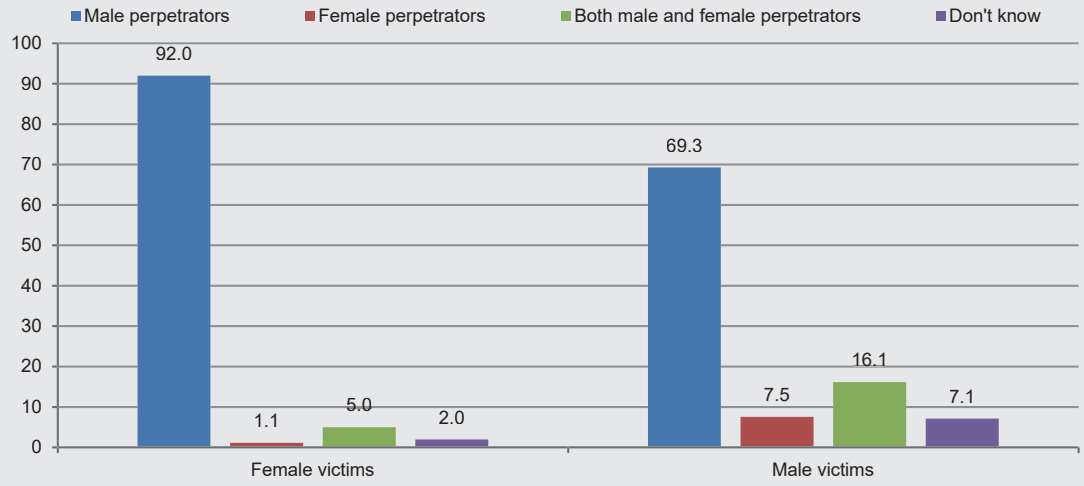
Perpetrators of sexual harassment were mostly men: they were the perpetrators for 97% of the female victims and for 85.4% of the male victims (Fig. 1.2).

Figure 1.1 - Women and men aged 14 to 65 subjected to sexual harassment during their lifetime, by type of harassment



Source: Istat, Citizen safety survey, 2015-2016

Figure 1.2 - Women and men aged 14 to 65 subjected to sexual harassment during their lifetime, by gender of the perpetrator



Source: Istat, Citizen safety suvey 2015-2016

What is more, women and men perceived the severity of the physical harassment differently: 76.4% of the women considered it very or fairly serious compared to only 47.2% of the men. These results, however, do not tell us much about the significance of the harassment for men. For instance, what is behind the experience of “shadowing” or “exhibitionism” for a male and for a female? What kind of situation they have on their minds while answering?

The Estonian Experience

The same concerns also came out from the survey on "Estonian men's attitudes and behaviour" carried out by the University of Estonia in 2014, and financed by the Government Office of Republic of Estonia.

> *"Unfortunately, the questionnaire (the same one used for women) did not perform well for men.* […] *The missing values were higher for men, especially for questions about sexual violence. The idea was also that the topics were too different for men and women; there was a different perception of the same topic. Men perceived sexual violence as a women's issue, not a men's issue.*
>
> *Men suffer above all emotional abuse in childhood (humiliating and controlling behaviours) and only a few times in adulthood. It seems that the pattern of the prevalence of psychological violence rate is completely different for men and women, since the psychological violence is almost exclusively during childhood for men.*
>
> *Their prevalence rate for physical violence is higher than women's rate, especially when children. But men suffer different kinds of violence: they fight at school, in the pub, they are often involved as victim and perpetrator in the same circumstance.* […] *The problem pointed out regarded the different meanings of violence for men and women.".*

In this perspective, the question about "what comprises gender-based violence suffered by men" arose and the suggestion was that it is possible perhaps to study violence suffered by men in the public and private sphere, but not GBV, since there is not an unequal power manifestation concerning men, as defined in the Istanbul Convention.

Data resulting from pilot surveys, carried out in the context of EU GBV violence survey's development, are very useful to assess the performance of the questionnaire on women and men. Obviously, data resulting from the test are not based on a representative sample, nor do represent the actual situation of violence in each country, because of the small number of interviews. Nevertheless, it was decided to use some of the indicators as a proxy of possible future situations.

The average of the countries' aggregated indicators is calculated by Eurostat taking account of the data sent by Bulgaria, Estonia, Latvia, Malta, and Slovenia and, in some cases, Poland and the Netherlands.

Analysing data, the prevalence rate of physical violence is quite similar for women and men, and often it is even higher for men (more than in half of the countries). Differences emerge when analysing the intensity/frequency or the seriousness of violence or the relationship to the perpetrator. In more detail:

- physical and sexual violence perpetrated by an intimate partner is three times higher for women over the lifetime, sexual violence by a partner is twice as much for women over the last five years;
- one-time physical violence is more prevalent for men than for women (almost in all countries), while repeated physical violence is more prevalent for women (for all the countries), especially when the perpetrator is the intimate partner;
- the percentage of women who experienced physical or sexual violence and felt that their lives were in danger was higher compared to men, especially when the perpetrator was the intimate partner;

- more women than men received injuries as a consequence of their intimate partner violence (physical or sexual);
- women experienced at least one act of psychological violence (sometimes, often or all the time), twice as many as the men;
- the prevalence rate for severe psychological violence (at least three different behaviours out of all listed acts) is three times more prominent in women than in men;
- women experienced controlling behaviour, combined with physical or sexual violence by their intimate partner, four to five times more often than men did.

Since the most severe forms of gender-based violence are the result of dominance and coercive control, specific variables of dominance and control must be added in the questionnaire, along with variables of violent acts reported by both men and women. Otherwise, theories and empirical approaches that identify male dominance and control as determinants of the most severe forms of violence would not be further testable.

Both women and men are victims of violence, but they experience different kind of violence: different perpetrators, different dynamics, different causes and roots, different consequences, different perceptions.

Unfortunately, there is a lack of conceptualisation of the topic of gender-related violence. For this reason, if the attention is on gender-based violence against women as it is reflected in the main purposes of the Istanbul Convention, it is suitable to interview only women. If the attention, instead, is on violence and domestic violence, then it is possible to interview men too, being aware that is something different from gender-based violence. In this case particular attention is required. It is necessary to adapt the questionnaire, differentiate the questions for men and women and interpret the results properly.

REFERENCES

American Psychological Association - APA. 1999. *Resolution on male violence against women*. Washington, DC, U.S.: APA. https://www.apa.org/about/policy/male-violence.

Arias, I., and P. Johnson. 1986. "Evaluations of Physical Aggression in Marriage". Paper presented at the *20th Annual Convention of the Association for Advancement of Behavior Therapy*. Chicago, IL, U.S., 13-16 November 1986.

Bell, K. M., A. E. Naugle. 2008. "Intimate partner violence theoretical considerations: Moving towards a contextual framework". *Clinical Psychology Review* 28: 1096-1107.

Bradbury-Jones, C., J.V. Appleton, M. Clark, and E. Paavilainen. 2017. "A Profile of Gender-Based Violence Research in Europe: Findings From a Focused Mapping Review and Synthesis". *Trauma, Violence, & Abuse*, Volume 20, Issue 4: 470-483. https://doi.org/10.1177/1524838017719234.

Brownmiller, S. 1975. *Against Our Will: Men, Women and Rape*. New York, NY, U.S.: Simon & Schuster.

Carlson, B.E., and A. Pollitz Worden. 2005. "Attitudes and Beliefs about Domestic Violence: Results of a Public Opinion Survey II. Beliefs about Causes". *Journal of Interpersonal Violence*, Volume 20, Issue 10: 1197-1218.

Council of Europe. 2011. "Council of Europe Convention on preventing and combating violence against women and domestic violence (Istanbul Convention)". *Council of Europe Treaty Series - CETS*, N. 210. Strasbourg, France: Council of Europe. https://rm.coe.int/168008482e.

Ehrensaft, M.K., P. Cohen, J. Brown, E. Smailes, H. Chen, and J.G. Johnson. 2003. "Intergenerational transmission of partner violence: a 20-year prospective study". *Journal of Consulting and Clinical Psychology*, Volume 71, Issue 4: 741-753.

European Institute for Gender Equality - EIGE. 2017. "Gender Equality Index 2017. Measuring gender equality in the European Union 2005-2015". *Report*. Luxembourg: Publications Office of the European Union. https://eige.europa.eu/publications/gender-equality-index-2017-measuring-gender-equality-european-union-2005-2015-report.

European Institute for Gender Equality - EIGE. 2016. *Measurement framework of violence against women in the EU. Interim Report.* Luxembourg: Publications Office of the European Union.

European Institute for Gender Equality - EIGE. 2015. "Gender Equality Index 2015: Measuring gender equality in European Union 2005-2012". *Report.* Luxembourg: Publications Office of the European Union. http://eige.europa.eu/sites/default/files/documents/mh0215616enn.pdf.

European Union Agency for Fundamental Rights - FRA. 2021. "Crime, Safety and Victims' Rights". *Fundamental Rights Survey*. Luxembourg: Publications Office of the European Union. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2021-crime-safety-victims-rights_en.pdf.

European Union Agency for Fundamental Rights - FRA. 2014. *Violence against women: an EU-wide survey. Main results*. Luxembourg: Publications Office of the European Union. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2014-vaw-survey-main-results-apr14_en.pdf.

Farrel, A.K., J.A Simpson, and A.J. Rothman. 2015. "The relationship power inventory: Development and validation". *Personal Relationships*, Volume 22, Issue 3: 387-413.

Felson, R.B. 2014. "Back to Basics: Gender and the Social Psychology of Aggression". In Gartner, R., and B. McCarthy (*Eds*.). *The Oxford Handbook of Gender, Sex, and Crime*, Chapter 5: 77-98. Oxford, UK: Oxford University Press.

Ferrales, G., and S. Maves McElrath. 2014. "Beyond Rape: Reconceptualizing Gender-Based Violence during Warfare". In Gartner, R., and B. McCarthy (*Eds*.). *The Oxford Handbook of Gender, Sex, and Crime*, Chapter 34: 671-689. Oxford, UK: Oxford University Press.

Ferrer-Pérez, V.A., E. Bosch-Fiol, A. Sánchez-Prada, and C. Delgado-Álvarez. 2019. "Beliefs and attitudes about intimate partner violence against women in Spain". *Psicothema*, Volume 31, Issue 1: 38-45.

Flood, M., and B. Pease. 2009. "Factors Influencing Attitudes to Violence against Women". *Trauma, Violence, & Abuse*, Volume 10, Issue 2: 125-142.

Graham-Kevan, N., and J. Archer. 2005. "Investigating Three Explanations of Women's Relationship Aggression". *Psychology of Women Quarterly*, Volume 29, Issue 3: 270-277.

Heberle, R. 2014. "Sexual Violence". In Gartner, R., and B. McCarthy (*Eds*.). *The Oxford Handbook of Gender, Sex, and Crime*, Chapter 4: 59-76. Oxford, UK: Oxford University Press.

Hickman, S.E., and C.L. Muehlenhard. 1997. "College Women's Fears and Precautionary Behaviors Relating to Acquaintance Rape and Stranger Rape". *Psychology of Women Quarterly*, Volume 21, Issue 4: 527-547.

Johnson, M.P. 2005. "Domestic Violence: It's Not about Gender: Or Is It?". *Journal of Marriage and Family*, Volume 67, Issue 5: 1126-1130.

Johnson, M.P. 1995. "Patriarchal Terrorism and Common Couple Violence: Two Forms of Violence against Women". *Journal of Marriage and Family*, Volume 57, Issue 2: 283-294.

Johnson, M.P., and J.M. Leone. 2005. "The Differential Effects of Intimate Terrorism and Situational Couple Violence: Findings From the National Violence Against Women Survey". *Journal of Family Issues*, Volume 26, Issue 3: 322-349.

Kimmel, M.S. 2002. "'Gender Symmetry' in Domestic Violence: A Substantive and Methodological Research Review". *Violence against Women*, Volume 8, Issue 11: 1332-1363.

Lauritsen, J.L., J.G Owens, M. Planty, M.R. Rand, and J.L. Truman. 2012. "Methods for Counting High-Frequency Repeat Victimizations in the National Crime Victimization Survey". *Technical Series Report*, NCJ 237308. Washington, DC, U.S.: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics - BJS.

Makepeace, J.M. 1986. "Gender Differences in Courtship Violence Victimization". *Family Relations*, Volume 35, N. 3: 383-388.

Sagrestano, L.M., C.L. Heavey, and A. Christensen. 1999. "Perceived Power and Physical Violence in Marital Conflict". *Journal of Social Issues*, Volume 55, Issue 1: 65-79.

Simon, T.R., M. Anderson, M.P. Thompson, A.E. Crosby, G. Shelley, and J.J. Sacks. 2001. "Attitudinal acceptance of intimate partner violence among U.S. Adults". *Violence and Victims*, Volume 16, Issue 2: 115-126.

Stets, J.E., and M.A. Pirog-Good. 1989. "Patterns of physical and sexual abuse for men and women in dating relationships: A descriptive analysis". *Journal of Family Violence*, Volume 4, Issue 1: 63-76.

Straus, M.A. 2012. "Blaming the Messenger for the Bad News about Partner Violence by Women: The Methodological, Theoretical, and Value Basis of the Purported Invalidity of the Conflict Tactics Scales". *Behavioral Sciences and the Law*, Volume 30, Issue 5: 538-556.

Straus, M.A. 1979. "Measuring Intrafamily Conflict and Violence: The Conflict Tactics (CT) Scales". *Journal of Marriage and Family*, Volume 41, Issue 1: 75-88.

Tjaden, P., and N. Thoennes. 2000. "Full Report of the Prevalence, Incidence, and Consequences of Violence Against Women. Findings from the National Violence Against Women Survey". *Research Report,* NCJ 183781. Washington, DC, U.S.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice Statistics - NIJ, Centers for Disease Control and Prevention - CDC.

United Nations, General Assembly. 2017. "Work of the Statistical Commission pertaining to the 2030 Agenda for Sustainable Development" *Resolution adopted by the General Assembly on 6 July 2017*, Seventy-First Session, Items 13 and 117. New York, NY, U.S.: United Nations. https://ggim.un.org/documents/a_res_71_313.pdf.

United Nations, General Assembly. 2006. "In-depth study on all forms of violence against women". *Report of the Secretary-General*, Sixty-First Session, Item 60. New York, NY, U.S.: United Nations. https://documents.un.org/doc/undoc/gen/n06/419/74/pdf/n0641974.pdf?token=xQC3NIWWzc1d8THjlt&fe=true.

United Nations, General Assembly. 1993. "Declaration on the Elimination of Violence against Women". *General Assembly Resolution*, 48/104, 20 December 1993. New York, NY, U.S.: United Nations. https://www.ohchr.org/en/instruments-mechanisms/instruments/declaration-elimination-violence-against-women.

United Nations Committee on the Elimination of Discrimination Against Women - CEDAW. 2017. "General recommendation N. 35 on gender-based violence against women, updating general recommendation N. 19". *CEDAW/C/GC/35*. Geneva, Switzerland: United Nations High Commissioner for Human Rights - OHCHR. https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-recommendation-no-35-2017-gender-based.

United Nations Committee on the Elimination of Discrimination Against Women - CEDAW. 1992. "General recommendation N. 19: Violence against women". *Convention on the Elimination of All Forms of Discrimination against Women*, 11th Session. Geneva, Switzerland: United Nations High Commissioner for Human Rights - OHCHR. https://www.refworld.org/legal/resolution/cedaw/1992/en/96542.

United Nations Committee on the Elimination of Discrimination Against Women - CEDAW. 1989. "General Recommendation N. 12: Violence against women". *Convention on the Elimination of All Forms of Discrimination against Women*, 8th Session. Geneva, Switzerland: United Nations High Commissioner for Human Rights - OHCHR. https://www.refworld.org/docid/52d927444.html.

United Nations, Department of Economic and Social Affairs, Statistics Division. 2014. "Guidelines for Producing Statistics on Violence against Women - Statistical Surveys". *Economic & Social Affairs*. New York, NY, U.S.: United Nations Publication. https://unstats.un.org/unsd/gender/docs/guidelines_statistics_vaw.pdf.

United Nations, Department of Economic and Social Affairs, Sustainable Development. 2015. *Transforming our World: The 2030 Agenda for Sustainable Development*. New York, NY, U.S.: United Nations. https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf.

United Nations Economic Commission for Europe - UNECE. 2015. *Indicators of Gender Equality.* Geneva, Switzerland: UNECE. https://unece.org/sites/default/files/2022-02/ECE_CES_37_WEB.pdf.

United Nations Entity for Gender Equality and the Empowerment of Women - UN Women. 1995. "Beijing Declaration and Platform for Action". *Fourth World Conference on Women*. Beijing, China, September 1995. https://www.un.org/womenwatch/daw/beijing/pdf/BDPfA%20E.pdf.

United Nations, Statistical Commission, Statistics Division, and Instituto Nacional de Estadística y Geografía de México. 2010. "Report on the Meeting of the Friends of the Chair of the United Nations Statistical Commission on Statistical Indicators on Violence against Women". *Meeting of the Friends of the Chair of the United Nations Statistical Commission on Statistical Indicators on Violence against Women*, Aguascalientes, Mexico, 9 - 11 December 2009. https://unstats.un.org/unsd/demographic/meetings/vaw/docs/finalreport.pdf.

Waltermaurer, E. 2012. "Public Justification of Intimate Partner Violence: A Review of the Literature". *Trauma, Violence, & Abuse*, Volume 13, Issue 3: 167-175.

Williams, S.L., D.K. McKelvey, and I. Frieze. 2014. "Intimate-Partner Violence". In Gartner, R., and B. McCarthy (*Eds.*). *The Oxford Handbook of Gender, Sex, and Crime*, Chapter 19: 362-378. Oxford, UK: Oxford University Press.

Wolfe, D.A., C.C. Crooks, D. Chiodo, and P. Jaffe. 2009. "Child maltreatment, bullying, gender-based harassment, and adolescent dating violence: Making the connections". *Psychology of Women Quarterly*, Volume 33, Issue 1: 21-24.

World Health Organization - WHO, and Pan American Health Organization. 2012. "Understanding and addressing violence against women. Femicide". *Information Note*. Geneva, Switzerland: WHO. https://www.who.int/publications/i/item/WHO-RHR-12.38.

Yllo, K. 1984. "The Status of Women, Marital Equality, and Violence Against Wives: A Contextual Analysis". *Journal of Family Issues*, Volume 5, Issue 3: 307-320.

2. QUESTIONNAIRE DESIGN[1]

Surveys on gender-based violence cover topics that are essential for policies, address sensitive issues and are affected by the respondent's availability to participate and to disclose their experience.

The questionnaire is one of the most important tool to predict the successfulness of the survey, or at least one of the main factors to achieve the goals of the survey. Therefore a particular attention has to be given to the questionnaire structure and sequence, to the wording, the clarity, and the completeness of the content and the possibility of data analysis, taking also into account the burden for respondents.

Since the aim of the survey is to estimate the gender-based violence, the main challenge is to produce the most accurate estimates, without risking underestimation.

Proposals and suggestions here described take into account the provisions of the Istanbul Convention on Gender-Based Violence, the recommendations of the United Nations Guidelines for Producing Statistics on Violence against Women (United Nations, 2013), the best practices developed in some European countries[2], the experience of the International Violence Against Women Survey (Johnson, Ollus, and Nevala, 2008) and of the Eu-wide survey (Fra, 2014).

2.1 The general structure of the questionnaire

The structure of the questionnaire requires a great attention. As general rule, the first section of the interview should include a suitable introduction where the survey goals are outlined. Providing an introduction of this kind helps the respondent to feel free when answering the questions. This step opens communication between the interviewer and the respondent and increases the respondent's willingness to participate in the interview. This strategy will minimise possible psychological harm or embarrassment of a respondent asked to discuss a traumatic experience.

Following the introduction, questions should be grouped based on themes. To keep the respondent not confused, questions should be ordered avoiding sudden changes in themes. When themes do change, are advisable a few sentences alerting the respondent to a change on the topic – the transition statements. This enables the respondent to prepare for and think about the new topic.

Having different sections for different topics helps the concentration and the disclosure. In this perspective it is strongly discouraged having together, in the same section, questions referring to different kind of perpetrators.

It is advised, also, to use a funnelling technique to fix the order of questions: each thematic section should begin first by asking general questions and next increasingly detailed and sensitive questions. At the end of the questionnaire, questions should move to simple or mundane topics. This technique takes into account the level of interest and attention of respondents. Research suggests that attention and interest of the respondent

1 This chapter has been drawn up by Maria Giuseppina Muratore.
2 In Estonia, France, Italy, Norway, Spain, Sweden and United Kingdom.

start with a high level and increases until it reaches a maximum point. It remains at this maximum for some time before it drops rapidly due to respondent fatigue (Pitrone, 1986). Thus, questions that are the easiest to be answered should be placed at the beginning of the interview while the most difficult and sensitive ones should be in the middle. Demographic and social characteristics of the respondent (gender, age, education, employment, etc.) can be asked at the end of the interview.

2.2 The screening questions approach

The main objective of surveys on gender-based violence and surveys on violence against women is to produce a precise estimate of the victims. To this end, it is necessary the use of screening questions that should be asked after some introductory questions of a more neutral character.

The screeners are a list of questions aimed expressly at determining whether the respondent experienced violence or not in the given reference period and, possibly, how many times it happened. Specific screeners can be used to identify different forms of violence the respondent might have suffered: physical and sexual violence but also psychological or economic violence.

Experiences in conducting surveys on violence against women suggest to list different acts of violence in the questionnaire, rather than attempt to develop and use a general definition of physical or sexual violence. The definitions in fact changes across countries, depend on many factors as cultural backgrounds, traditions and subjective perception (United Nations Statistics Division, 2010).

The survey should collect the number of victims and, for each victim, the number of acts of violence suffered by the victim. Furthermore, since different acts can be part of the same episode of violence - for instance, in the same episode a victim can be beaten with a fist and threatened with a knife (in this example two different acts) - it is possible to use additional specific questions in order to distinguish if several violent acts refer to a single episode or they represent different violent episodes.

Further questions can be also useful to identify when cases of repeated or multiple victimisation have the same perpetrator or different perpetrators, especially when different perpetrators can be referred to the same generic category: for example two episodes of violence perpetrated by a relative (or friend) but it wasn't the same person in the two cases (for instance, in the example, two different relatives or friends for two different episodes).

Only if the screening questions reveal that the respondent suffered violence, additional questions have to be asked in a dedicated section, gathering more detailed information.

2.2.1 The wording

Victimisation surveys should avoid using legal language, as most of the respondents are unfamiliar with the precise meaning of legal terms. Further, in violence against women surveys, for example, some respondents may not feel that what they have experienced was a crime, and they might not report an incident if the emphasis is on breaking the law. Rather than using such words, it is preferred to describe an event in terms of its elements: the use of force, the threat, presence of a weapon, physical contact, physical injuries, etc.

It is not advisable to mention the word violence. Research suggests that asking about situations, instead of using terms as "violence", "rape", "assault", increases disclosure of victimisation events. Stated differently, questions should not ask about any physical or sexual violence directly. Instead, they should describe episodes, examples, or incidents that respondents can recognise as happened in their life. In these scenarios, the details or prompts used should include a variety of situations, locations and relationships to the perpetrator. This approach helps victims in remembering events and decreases possible underestimation of the phenomena of interest.

Moreover the screening questions serve to prompt the respondent's memory of violent events. The wording of screening questions can be brief or very detailed. Questions can be a list describing different situations or a list of examples of victimisation stories. In general, questions describing situations are preferred. Each question should refer to one form of violence only, starting with the less serious forms to the most severe.

The language used for the screening questions should be chosen carefully, especially if the survey is carried out in areas characterised by a variety of languages and ethnicities. In these cases, some wordings could inhibit the recall in some people while stress it in others. Sensitivities about what is acceptable to ask in an interview can also be different across cultures: talking about sex can still be a taboo, many persons avoiding to talk about it, or being very embarrassed.

In order to identify violent situations, it is possible to include in the question words that refer to personal feelings as hurt and fear. For example it is possible to add to the description of the incident the fact that it happened in a way "that hurts you or that could hurt you". This can be useful when asking about physical violence and threats and might help to avoid the collection of minor events. However mentioning the feelings of scare or fright in the text of questions introduce an element of subjectivity in the answers. The reference to this subjective aspect can introduce interpretation and consequently can represent a different stimulus for different respondents with consequent bias risk. On the contrary, the worlds used to formulate questions should refer to objective behaviours that describe the violence occurred (being hit with an object, slapped, etc.).

Description of violent situations can be especially awkward for the rape. It is possible to ask if it occurred to be forced to have a "sexual intercourse"[3]. This expression allows a question relatively easy to be asked by the interviewer, but it is not so clear what sexual intercourse include and different answers as well as different understandings of it are possible.

The alternative is using more detailed questions, asking separately about each different kind of penetration (vaginal, anal and penetration with objects or fingers) and oral sex[4]. This is not easy to be asked, but allows an effective comparability of data between countries and avoid any objection by media or politicians, about what data are about. Furthermore, these questions seem more suitable for collecting data about sexual violence against men.

It is useful to recall some general rules that always apply when formulating questions for a survey. Two fundamental components must be considered. First, it must be taken into account the population of interest. When interviewing the general population, education levels, age and cultural experiences vary. Second, it must be considered the topic of the

3 This is how the question about rape is asked in the EU wide survey (Fra, 2014) in the Italian Women's Safety Survey (Istat, 2005 and 2014), in the International Violence against Women Survey (Heuni, 2005).

4 This is the choice of the US National Violence against Women Survey, carried out on a sample of men and women in 1995-96, of the "Violence and rape in Norway" Survey in 2013.

question: each question must be clear. Ensuring clarity means clarity to all readers in the target population so that nobody is forced to interpret any question. Interpretation by respondents should be avoided because it involves the use of the personal reference system, jeopardising the data gathered.

Also in national survey, different languages may be used, where there are more than one national language or where there is emphasis on including linguistic minorities. After a first translation, the wording of questions may need to be considered again, also in the original language, in an effort to find formulations that would carry the same meaning in different languages.

To avoid common pitfalls, several approaches are recommended. These include:

- use simple terms;
- use precise terms, avoiding ambiguous words or those with multiple meanings;
- use examples to clarify the correct interpretation for complex questions;
- avoid technical terms;
- avoid depreciative words;
- avoid colloquialisms;
- avoid long questions that may confuse the reader;
- avoid double-barrelled questions;
- avoid constructing questions containing too many specifications, since the question could appear rather burdensome and difficult to understand;
- avoid using negative sentence.

2.2.2 How many sections of screening

It is discouraged to have questions referring to different kind of perpetrators in the same section. On the contrary, following the general principle of grouping questions based on themes, screenings should be organised based on the relationship between the victim and the offender.

It is recommended to use three different screening sections: 1) screening questions on violence perpetrated by persons other than intimate partner; 2) screening questions on the current partner; 3) screening questions on former partners.

In addition to screening questions that identify physical and sexual violence, screeners related to violence by current and former partners should include also questions on psychological and economic violence.

It is useful to underline that, without such distinguished screeners, there is the real possibility of underestimating some forms of violence, as for example the sexual violence perpetrated by an intimate partner, because some respondents do not see violence by an intimate partner as a personal or violent victimisation. Experience shows that such screening sections allow to remember much better than a generalised screening section.

The approach of multiple screening sections is adopted successfully in the Us National Crime Victimisation Survey (Ncvs). In the survey there a list of screeners related to the intimate partner violence and a separate list of screeners related to violence by a non-partner author. The same kind of approach was followed also in the International Violence against Women Survey (Ivaws). In the first version of the questionnaire, there were five lists of screeners about physical and sexual violence: one screening for the current partner, one for former partners, one for male relatives, one for other male perpetrators the respondent knew, one for male strangers. However, after pilot tests, the Ivaws team decided to reduce

the number of screening lists due to the excessive length of the questionnaire and the respondent's burden. The number of screening sections suggested was reduced to two: the screening questions for violence by partners and the screening questions for violence by non-partners. The UNECE Task Force and the Friend of the Chair Group, (United Nation Statistical Commission and United Nation Statistics Division) in 2009 recommended three screening sections in order to collect data of good quality to measure violence against women: screeners for non-partner perpetrators, screeners for previous partners and screeners for violence perpetrated by the current partner.

The sequence of such distinguished sections has also to be organised carefully. The position of questions in the questionnaire can influence responses and experience offers some lessons about it.

The complex structure of the questionnaire for the "Women safety" survey, carried out in Italy in 2006 by Istat, needed a test to evaluate the best placement of the sections (the screening questions and the in-depth questions) regarding violence by intimate partner and violence by non-partner. The pre-test of about 200 interviews showed that placing questions regarding non-partner violence before the corresponding questions regarding current or previous intimate partner worked better. It appeared that asking women about violence suffered by their partners at the beginning of the questionnaire upset the women, shut down the communication and in many cases ended the interview. In contrast, asking first questions about non-intimate partner violence followed by the section on intimate partner violence turned out in better relations with the interviewed and the completion of the questionnaire.

Focussing the attention on each perpetrator separately helps the disclosure, especially if starting with the perpetrator least close to the respondent (non-partner) and ending with the closest perpetrator (current intimate partner). Following this approach, the first group of questions make respondents think that violence exists. Moreover the fact of being asked about this kind of experience suggests that what they suffered is something that happens to many people. Therefore, they are helped to be legitimated to recognise it as something of wrong and to recall it without shame. As the confidence between interviewer and respondent increases, the ground become ready for asking questions that are more sensitive about perpetrators much closer to the victim, as the current partner is.

2.2.3 Time frame and reference period

Victimisation surveys generally are possible by asking the respondent to consider events from the past; hence, the questions are retrospective. A problem with retrospective surveys is ensuring that the respondent correctly recall that an event occurred, and correctly recall when this happened. Errors in memory increase as the reference period becomes longer.

Generally the most used prevalence indicators in surveys on gender-based violence refer to the life course period and the previous last five years before the interview; estimates referring to the last 12 months are, in fact, less robust. Furthermore, precise estimations for a specific year are less meaningful because social phenomena such as GBV change over a longer time span and to compare the data with register statistics it is less important than in case of victimisation surveys where data about robberies, thefts and other crimes against properties are collected. GBV is generally repeated and it is characterised by series of events that victims are not able to recognise precisely in a short time span and generally overlaps over the calendar years.

Asking about experienced violence occurred lifetime and successively about possible experienced violence in the last 5 or 3 years and about the last 12 months is a way to improve the recall. In fact, if the respondent reports an event occurred lifetime, the interviewer can ask if the event occurred during the last 5 years and, if yes, it can be asked about a shorter period such as the last 12 months or the last calendar year.

In surveys on gender-based violence, multiple time frames are usually used instead of the lifetime reference period. Often there are questions concerning violence in adulthood and other questions with reference to possible violence suffered as a child. Questions about violence occurred in adulthood refers to possible experience "since the age of 15" or 16 or 18. When asking about violence perpetrated by an intimate partner the reference time can be "during the relationship" while it is recommended to distinguish violence suffered as a child clarifying in the question if the event happened "before the age of 15" (or 16 or 18).

Violence suffered and/or witnessed in childhood

Having witnessed violence between parents, having suffered psychological and physical violence by parents, as well as sexual violence by any kind of perpetrator, are all considered important risk factors for being victim of violence once adult. Therefore the survey should collect information about possible experience of this kind.
As already mentioned, in surveys on gender-based violence multiple time frames are usually used instead of the lifetime reference period and it is recommended to distinguish violence suffered as a child clarifying in the question if the event happened "before the age of 15" (or 16 or 18).
The questions about childhood can be asked in the first part of the questionnaire since might be considered relatively easier to talk about something unpleasant related to the past, compared to a more recent event. This approach is adopted by the Norwegian Centre for Violence and Traumatic Stress Studies (Nkvts) in the survey "Violence and rape in Norway" carried out in 2013. However, since violence suffered in childhood is not a core topic in a GBV survey, it is also possible to address this topic after the questions related to violence experienced in adulthood.

2.2.4 Relationship to the perpetrator

Relationship to the perpetrator is a key element of violence against women. Consequently, this crucial variable has to be included in all GBV surveys. The classification of relationships to the perpetrator should capture the gender of the aggressor and the power relationship that exists between aggressor and victim.

The classification presented in the "Guidelines for Producing Statistics on Violence against Women" (United Nations Department of Economic and Social Affairs, 2014) represents a solid basis for further deliberations. It would be useful adding sub-categories of relationship by disaggregating the main categories (relatives, for example, should be further broken down based on the degree of relationship). The definition of intimate partner should be adapted to the national context (UNSD, 2010).

The current intimate partner can be defined as someone who is married with the respondent, who lives together, or who is engaged at the time of the interview. The previous intimate partner is applicable to persons involved in former marriages, co-habitations or engagements.

The non-partner perpetrators are persons who are not linked now nor in the past to the respondent by any intimate relationship.

Additional information can be gathered on the non-partner perpetrator. The list of possible kind of authors can include main categories (relatives, friends, known person, unknown, etc.) but also some information about the possible relation of authority (mentioning categories as army or police officer/priest/doctor). The UNODC International Crime Classification for Statistical Purposes (ICCS) of March 2015 considers for the relationship between victim and perpetrator: current spouse or intimate partner (cohabitating or non-cohabitating partner or boyfriend/girlfriend), like current spouse or cohabitating partner, current non-cohabitating partner, like boyfriend/girlfriend but not married; relative (like cohabitating blood relative, non-cohabitating blood relative) like child, parent, other household member or relative, other perpetrator known to the victim, like friend, acquaintance, colleague/business or work relationship, authority/care relationship (like doctor/nurse/teacher/police/public official, clergy, etc.); perpetrator unknown to the victim[5]. About the sex of the perpetrator, it is possible that the victim is not aware of it or prefers not answering due to the sensitiveness of the topic.

As already mentioned, the choice of using separate sections for different perpetrators (non-partner, current partner and former partner) seems to be strategic because it allows the respondent to focus more precisely on separate violent events and histories linked to different perpetrators. Moreover, it allows to address the subject of domestic violence in a more gradual way, in an advanced phase of the interview, when a collaborative and trusting relationship is assumed to be established with the interviewer.

2.2.5 Frequency and intensity of violence

Frequency of violence refer to how often the respondent suffered it and gives information on the intensity of the experience. Counting the number of episodes is not always straightforward, especially in case of some forms of violence. For example, gathering information on the frequency of physical violence has a different weight depending of the recollection period and the type of perpetrator. In case of physical violence experience in the last 12 months, ideally the frequency should be defined counting each individual violent act that the victim suffered in that period. However, results of available surveys show that counting individual events presents considerable difficulties and therefore questions should avoid to request the exact number of events. In any case it is recommended to gather enough information to make always possible computing the total number of occurrences. It is possible, for example, to ask if the violence occurred once a week, twice a month, and so forth. In case of violence experienced in the lifetime, frequency should refer to a different count and the approximation will be less precise: violence happened once, few times or many times in the life span.

2.3 Follow up questions for victims

When screening questions reveal that the respondent experienced violence, additional questions have to be asked to gather details about these victimisations.

The questionnaire can have as many in-depth questions or sections about the incident as the number of crimes revealed by the screeners. However, very often survivors of gender-based violence suffered multiple victimisations of the same type. When this is

5 https://www.unodc.org/documents/data-and-analysis/statistics/crime/ICCS/ICCS_English_2016_web.pdf.

the case, researchers have to decide whether to gather information about one incident or about all similar events. This decision will greatly influence the design of the rest of the questionnaire. In case of multiple victimisation, there are several options to choose which incident will be further investigated through additional questions. It is possible to collect detailed information on:

- the incident the victim considers as the most serious;
- the most recent event;
- all episodes of victimisation occurred;
- a subset of episodes of victimisation (*e.g.* last six episodes occurred during a fixed reference time).

Each of these options affects the quality of the data, the cost of the data collection and the interview's length.

For example, the choice of gathering data on the most serious event will bias the data toward more serious crimes and therefore threaten a representative description of victimisation. Moreover, the severity of the incident is not an objective criterion and it can be affected by the respondents' characteristics (sex, age, education, etc.), by the remembering effect and by cultural differences.

The option to gather information on all the episodes of victimisation in a fixed period - for instance the last six episodes - will offer a more precise picture of the qualitative aspects of victimisation in a precise period, but it will provide anyway only a partial information.

Furthermore, it has to be noted that gender-based violence is characterised by a particular kind of victimisations: repeated events. It is a form of continuous violence, due to its nature, and it is difficult to be seen as a single event. These kind of repeated victimisations tend to have similar characteristics and when there is a series of victimisations, the respondents are not able to offer detailed information about each event or the exact number of incidents.

A common way to deal with this issue is recording the number of times the respondent feels as victimised, and gathering further information only on the most recent incident. This approach is widely adopted in surveys on violence against women[6] and it is a way to select the incident randomly. This choice allows a real and representative picture of the existing violence taking into account all kinds of incidents, from least serious to most serious.

When collecting only the most recent victimisation there is a clear disadvantage: other experienced victimisations might be represented not accurately. A solution can be to collect information regarding a longer series of violent events, adding questions that refer to the overall history of violence: for instance, asking if the respondent has ever been injured, if there are long term consequences due to violence, the incurred costs, the presence of children witnessing violence, etc.

Severity of violence

As the severity of a given violent act can be perceived differently by victims depending on the perpetrator, it is not recommended to assign a severity level to individual acts of violence. It would be preferable to assess the topic of severity regardless the list of acts suffered, developing a scale based on the consequences the victim suffered (Un, The Friends of the Chair group). For example, if the victim is not able to lead her daily life as a result of

6 For example, by Statistics Canada, the Australian Bureau of Statistics, Statistics Finland, the US Department of Justice and Statistics Sweden and Istat.

violence, the event should be considered as “severe”, even if there is not physical damage.

There is also the possibility to consider an episode serious if it involved the feeling of fear. However, national cultural contexts determine different social views of “fear” and the subjective assessment affect the comparability of data. Therefore, consideration should be given to both physical and emotional consequences when determining severity.

It is advisable to assess the severity collecting more than one variable/question related to this aspect: presence of injuries sustained, need of medical care, impossibility of carrying out the normal everyday activities and/or the impossibility of working, permanent physical damages, use of medicines, drugs or alcohol subsequent to the episode, the victim being in fear for her life, victim’s perception of seriousness, etc.

Dynamics of violence

Since it is important to know the dynamic of violent episodes, the questionnaire should collect data about what happened as the presence of other people involved, the place where the episode occurred, possible triggering events or circumstances, the victim’s reactions, perpetrator being under the effect of alcohol or drugs.

Victims’ disclosure and the seeking for help

The survivors’ attitude to disclosure can be analysed asking if the respondent ever talked with someone about the suffered violence (for example family members or friends), if sought for help contacting the police forces and/or specialised services. In case of reporting to the police it is possible to ask the satisfaction of the victim with the work of the police. In the opposite case, it is possible to ask the reason why the survivor did not report.

In case of violence perpetrated by an intimate partner, additional questions may investigate if the violence occurred when the relationship with the partner was ending, if it ended as a consequence of the violence. If the couple never split, or split only temporary, it is also possible to ask the reasons for not leaving the partner or the reasons for being together again.

REFERENCES

European Union Agency for Fundamental Rights - FRA. 2014. *Violence against women: an EU-wide survey. Main results*. Luxembourg: Publications Office of the European Union. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2014-vaw-survey-main-results-apr14_en.pdf.

Johnson, H., N. Ollus, and S. Nevala. 2008. *Violence against Women. An International Perspective*. New York, NY, U.S.: Springer.

United Nations, Department of Economic and Social Affairs, Statistics Division. 2014. "Guidelines for Producing Statistics on Violence against Women - Statistical Surveys". *Economic & Social Affairs*. New York, NY, U.S.: United Nations Publication. https://unstats.un.org/unsd/gender/docs/guidelines_statistics_vaw.pdf.

United Nations, Statistical Commission, Statistics Division, and Instituto Nacional de Estadística y Geografía de México. 2010. "Report on the Meeting of the Friends of the Chair of the United Nations Statistical Commission on Statistical Indicators on Violence against Women". *Meeting of the Friends of the Chair of the United Nations Statistical Commission on Statistical Indicators on Violence against Women*, Aguascalientes, Mexico, 9 - 11 December 2009. https://unstats.un.org/unsd/demographic/meetings/vaw/docs/finalreport.pdf.

3. QUALITATIVE METHODOLOGIES FOR QUESTIONNAIRE ASSESSMENT[1]

The topic of gender-based violence is one of the priority areas at the international as well as the national level. The definition of gender-based violence has been discussed and developed by several institutions (Council of Europe, United Nation, European Commission)[2], and presented in several policy documents: gender-based violence is understood as violence directed against a person because of that person's gender. According to the Istanbul Convention, countries should conduct population-based surveys at regular intervals to assess the prevalence and trends of GBV. However, the challenge for collecting the data on GBV has remained, while it is not possible to directly use those definitions in the questionnaire in order to collect harmonised data. Furthermore, the prevalence and disclosure rate might be connected with the extent to which violence is tolerated in the wider community (Waltermauer, 2012).

The most recent EU-wide opinion survey[3] showed that there are significant differences between Member States concerning the level of gender-based violence that is accepted and justified at the level of the population at large.

Survey data are only as meaningful as the answers the survey respondents provide. Therefore, in the questionnaires on GBV and VAW the questions usually focus on measuring behaviours and specific acts, and their effects on the person's physical, sexual and emotional wellbeing.

Nevertheless, it is important to make sure that the respondents understand the questions. Therefore, the suitability of the instrument needs to be evaluated, with a focus on the ease in understanding the questionnaire by the respondents as well as its management by the interviewers.

> "Good questionnaires impose low response burden and remain both respondent and interviewer-friendly. They ask relevant questions and permit data to be collected efficiently and with a minimum number of errors, while facilitating the coding and capture of data and minimising the amount of editing and imputation that is required" (Statistics Canada, 2003).

The questionnaire is both a tool for measuring and for communicating, and for that reason it should be pre-tested. Pre-testing allows one to test operational concepts and definitions as well as specific questions or the questionnaire as whole; it is particularly important in cases where no pre-existing questionnaire or topic-specific survey exists and new survey tools have been developed or if a pre-existing questionnaire has been adapted for a new culture/language[4].

1 Paragraphs 3.1 and 3.2 have been drawn up by Roberta Barletta; paragraphs 3.3 and 3.4 have been drawn up by Elena De Palma.

2 See Council of Europe Convention on preventing and combating violence against women and domestic violence (2011) Council of Europe Treaty Series-No. 210. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168008482e); UN Declaration on the Elimination of Violence against Women Proclaimed by General Assembly resolution 48/104 of 20 December 1993, https://www.un.org/en/genocideprevention/documents/atrocity-crimes/Doc.21_declaration%20elimination%20vaw.pdf).

3 See Special Eurobarometer 449 - November 2016, Gender-based violence. Report .

4 See United Nations Department of Economic and Social Affairs, Statistics Division (2014). Guidelines for producing Statistics on Violence against Women. https://unstats.un.org/unsd/gender/docs/Guidelines_Statistics_VAW.pdf.

Pre-testing plays an essential role in identifying and potentially reducing the measurement error that damages statistical estimates at the population level and thus endangers comparability across populations in multinational, multiregional and multicultural surveys.

3.1 Tools and methods for testing the questionnaire: overview

A wide range of tools and methods can be used to test and evaluate the questionnaire. They include focus groups and in-depth interviews, preferential witness meetings, expert reviews, non-participating observation, cognitive methods such as think-aloud interviews and paraphrasing, behaviour coding. Findings from these tasks can be used to elaborate and refine the instrument until a final questionnaire is created. The suitability and intensity of their use depend on various factors and circumstances.

The pre-testing process can refer to different stages of survey planning and design of the survey's tools (Bradburn and Sudman, 1991; Istat, 1989).

The researchers can plan tests at the initial stage of the questionnaire development process. At this stage, concepts, definitions and knowledge about the topics of interest are evaluated. These tests include:

- *One-on-one in-depth interviews*: they represent the most common data collection method in qualitative research and are a familiar and flexible way of asking people about their opinions and experiences. Qualitative interviews are generally described as either being semi-structured or in-depth. The former are based on a series of open-ended questions about the issues the researcher thinks are relevant to the topic. The latter may only include one or two topics, but in much greater detail. The use of individual interviews may be more appropriate when discussing sensitive issues or topics that require self-disclosure.
- *Focus groups (FG)*: they are small gatherings of 6-12 people who meet with a trained moderator to talk about ideas and materials. A focus group is a type of in-depth interview that takes place in a group and the participants influence each other through their answers to the ideas and contributions during the discussion. The moderator stimulates discussion with comments or subjects. The information gathered can provide important clues about human attitudes and values as they relate to the topic. Focus groups are a useful method for quickly obtaining a wide variety of information on the topic and for getting new ideas, suggestions and recommendations. It allows a variety of different opinions to be heard in a short period of time. However, this method does not allow one to research the question-and-answer process in detail and therefore, for a detailed evaluation of the questions, other methods might be more useful. A focus group could be used in combination with other methods, for example, with a focus group that might be more efficient in identifying definitional problems, while cognitive interviews could be used for evaluating specific wording of the question[5].
- *Expert review*: this method includes individually based expert reviews to get expert opinions in the design stage of the questionnaire, in the pre-testing stage or even on the operational stage. It can be useful for evaluating draft questionnaires or sub-sets of questions. The reviewers should be experienced survey methodologists with knowledge of the theoretical or practical aspects of questionnaire design, fieldwork issues and data

5 Focus group technique is discussed more in depth in paragraph 3.2.

processing. The number of expert reviewers tends to be small, from two or three to over twenty experts, depending on various factors, such as the complexity of the topic under investigation, the level of expertise among the question designers and the time and resources available. Experts can be consulted independently or brought together in the form of an expert panel[6].

Later, when a draft questionnaire already exists, laboratory methods (pre-field) can be used on sub-sets of questions. Generally, they are mainly qualitative tests:

- *Pre-testing of the questionnaire*: they can be conducted as an informal test that helps identify poor wording or ordering of the question, errors in questionnaire layout or instructions, and problems caused by the respondent's inability or unwillingness to answer the questions. Pre-testing can also be used to suggest additional response categories that can be pre-coded on the questionnaire, and to provide a preliminary indication of the interview length and refusal problems.

At this stage, the goal is not to draw a representative sample of the target population, but to assure that the instrument is tested in various difficult conditions to stress its reliability.

Interviewers for the pre-test should be particularly able to help in identifying problematic areas of the questionnaire or of the interview. It is recommended to provide interviewers with a detailed list of the aspects of the questionnaire to be tested, and the feedback of the interviewer should be recorded and analysed. Debriefing sessions with interviewers can take place following a pre-test in order to get their input during the (re)design process. It is also advised that the researchers responsible for the survey conduct some interviews themselves, to test the suitability of the instrument. Pre-testing can be iterative until the final version of the questionnaire is generated:

- *Cognitive interviewing*: cognitive methods are used to examine the respondents' thought processes as they answer the survey questions and to ascertain whether or not they understand what the questions mean and are able to provide accurate responses (Forsyth and Lessler, 1991)[7].

When an advanced version of the questionnaire is available, field methods can be used. These tests are carried out mainly in conditions that resemble the real conditions of investigation. Generally, they are either qualitative or quantitative tests. Pilot surveys are included:

- *Behaviour coding*: it provides a systematic and objective means of examining the effectiveness of the questionnaire by analysing the interviewer-respondent interaction. It can be a field test or a laboratory test. It requires a medium-sized sample (30 interviews are considered sufficient for detecting problems), well-trained coders and consistent use of the coding scheme (Oksenberg, Cannell, and Kalton, 1991; Groves *et al.*, 2009).
- *Interviewers' debriefing*: they are discussion groups with interviewers with the aim at obtaining useful feedback on the performance of the questionnaire. Interviewers' debriefing can be planned daily, soon after the data collection session, or after each interview. It is recommended to take notes during the meeting or to record it.
- *Respondents' debriefing*: it is an interview with the respondent, soon after the survey interview, to get useful feedback on the issues of the questionnaire. These kinds of interviews are very useful when used together with quantitative tests such as behaviour coding, and quantitative data such as non-response rates.

6 The expert review method is discussed more in depth in paragraph 3.3.
7 Cognitive methods are discussed in depth in paragraph 3.4.

- *Pilot survey*: it reproduces all the survey phases on a small number of survey units or respondents. It is not necessary that sample units or respondents are representative of the overall population. Rather, it is important that they represent the various ways respondents will be contacted, so they can reveal any difficulties that may emerge during the survey itself. While a pre-test focuses more on the questionnaire alone, a pilot study deals with the entire survey.

For comparing alternative sequences or alternative wording of the questions, researchers can also conduct experiments or experimental tests consisting of quantitative and qualitative analysis of the results obtained through different versions of the questionnaire:

- *Split sample test*: (the alternative test): it refers to controlled experimental testing of questionnaire variants or data collection modes to determine which one is "better" or for measuring differences between them (Statistics Canada, 1998). Split sample experiments may be conducted within a field or pilot test, or they may be embedded within production data collection for an ongoing periodic or recurring survey. When it is used as a pre-testing method, it is necessary to previously define the standard by which the different versions of the questionnaire or questions can be judged. The sample size for each alternative should be designed to ensure that it has sufficient statistical power to detect real differences on the aspects under assessment. It is important to use a large sample and that the participants have been randomly assigned the test versions. This type of pre-test is expensive, nevertheless, it provides important information for comparing survey questions, assessing the extent and direction of impact concerning changes in question order, wording or layout.

For pre-testing a questionnaire on GBV or VAW, split sample tests may be conducted using, for example, a different sequence of the violence screening (partner and non-partner), different wording of the questions about sexual violence or different options for counting the number of episodes that occurred.

3.2 Focus groups

Focus groups have a long history. Originally called focused interviews, they originated in 1941 when the sociologist Robert Merton was invited by Paul Lazarsfeld to assist him in the evaluation of audience response to radio shows.

Since Merton's pioneering work (Merton and Kendall, 1946), focus groups have become an important research tool for applied social scientists such as those who work in programmes evaluation, marketing, public policy, advertising and communication. Focus groups may be useful at virtually any point in a research process, but they are particularly useful for exploratory research, where rather little is known about the phenomenon of interest, in order to acquire more knowledge about the topic of interest and to clarify the goals of the study. As a result, focus groups tend to be used very early in a research project and are often followed by other types of research that provide more quantifiable data from larger groups of respondents. Focus groups also have been proven useful following the analysis of large-scale, quantitative surveys. In this type of survey, the focus group facilitates the interpretation of quantitative results and adds depth to the responses obtained in the more structured survey.

Among the more common uses of focus groups are the following:

- obtaining general background information about a topic of interest;

- generating research hypotheses that can be submitted to further research and testing using more quantitative methods;
- stimulating new ideas and creative concepts;
- learning how respondents talk about the phenomenon of interest. this, in turn, may facilitate the design of questionnaires, survey instruments, or other research tools that might be employed in more quantitative research;
- pre-testing an available preliminary questionnaire;
- interpreting previously obtained quantitative results.

The Focus Group has some advantages compared to other qualitative methods: it is comparatively easier to lead or conduct, it allows one to explore topics and to generate hypotheses, it generates an opportunity to collect data through group interaction, which concentrates on the topic of the researcher's interest, it gives speed in the supply of the results and it has low cost compared to other methods.

On the other hand, it takes effort to assemble the groups, it demands carefully trained interviewers in order to conduct the discussion in an atmosphere that facilitates dialogue, it is not based on a natural atmosphere and it is not possible to know whether or not the interaction in a group takes into account individual behaviour. Furthermore, the researcher has less control over the data that are generated and the data analysis can be difficult to carry out (Kruger, 1994; Morgan, 1988).

When designing or pre-testing a questionnaire, focus group can be useful to (Groves *et al.*, 2004):

- explore and define concepts;
- detect sensitive topics and domains to check;
- identify what potential respondents know about the subject, how they structure the subject and what they think about it;
- develop survey objectives or data requirements;
- learn from the respondents what issues of the topic are of relevance for the respondents' point of view in order to get in good communication with the respondents in a future survey;
- find out if people understand the terms and how they define them;
- evaluate alternative wording of questions and formats;
- determine the feasibility of conducting the survey;
- understand respondent burden.

Statistics Canada researchers used the focus group method in developing the family violence module introduced in the General Social Survey in 1999. The participants were recruited from the general population and they took part in a two-step procedure: a telephone interview followed a few days later by participation in a focus group. Focus groups were organised with the general population, only with male respondents, only with the elderly and only with people of ethnic minority backgrounds. The moderators of the focus groups also monitored the telephone interviews (with the full knowledge of the respondents) to better understand the experiences of the respondents.

The main findings from the focus group showed that respondents were very sensitive to the interviewers' tone of voice and if the interviewer took the time to ask more sensitive questions. While some respondents indicated that the questions were about painful experiences, others welcomed the opportunity to provide information about their situations because they wanted to tell their story. However, almost all respondents accepted a national

crime survey that collected information on spousal violence and senior abuse and felt that Statistics Canada as a federal government agency should be gathering these data.

Focus group, preferential witness meetings and in-depth interviews have been used by the researchers of the Italian Institute of Statistics while planning new surveys on sensitive topics such as the Violence against women Survey (2005) the survey on Discrimination based on gender, ethnicity and sexual orientation (2009), the survey on Life conditions of foreign citizens (2011-2012). For the Violence against women survey, focus groups have been carried out also after the pilot study.

Focus groups for planning the Italian Violence against Women Survey (2005)

Focus Groups in the survey design phase with the aim of:

– going more in depth and learning more about the phenomenon of violence against women;
– planning a robust methodological structure, from the questionnaire to the choice of the interviewing mode, from the recruitment of the interviewers to the planning of the training programme and all through the data collection phase.

Seven Focus Groups were conducted with:

– workers in shelters for women victims of violence (four groups);
– women victims of domestic violence (one group);
– women from the community (one group);
– interviewers who have already had experience in victimisation surveys (one group).

Soon after the pilot study with the aim of:

– developing and improving the questionnaire and the strategies for contacting and interviewing the respondents.

Three Focus Groups were conducted with:

– interviewers who carried out the pilot survey (one group);
– workers in shelters for women victims of violence (two groups).

Each focus group was conducted by two moderators (at least one psychologist) and was recorded on video or audio and then transcribed verbatim. According to the group, different questions were asked, with the aim of going more in depth about:

– questionnaire contents, structure and wording;
– procedural methods.

Some results about questionnaire contents:

- What comprises domestic violence (especially psychological violence)
- different expressions of violence (economic, psychological, physical and sexual violence)
- how women deal with violence, what is their perception of it
- early precursors of dv, the violence cycle
- risk factors
- consequences of violence
- the way out of violence
- prejudices and stereotypes about violence against women

Some results about wording, sequencing and procedures

- Choose female interviewers;
- ask gradually about violence:

– ask questions about everyday life, leisure time, social networks, health conditions, before asking questions about violence;

–insert questions on physical and sexual violence by the partner after questions about the partner's background and psychological violence.
- not to name it 'VIOLENCE';
- use screening procedures which help to focus the attention on both physical and sexual violence, on the reference period and on all the possible perpetrators (partner and non-partners);
- choose separated screening on violence by current and former partner;
- use clear, simple wording. Avoid jargon.

Characteristics of focus group

Basically, focus groups are interviews, but of 6-12 people at the same time in the same group.

Elements characterising FG are:

- the interaction of a small group of people;
- the presence of one or more moderators;
- the focus on a specific topic that researchers want to analyse in depth.

Table 3.1 - Types of Focus Group

Group composition	– participants don't know each other and don't know the moderators – participants know each other and know the moderators – homogeneous group – heterogeneous group
Structuring degree	– self-managed groups (unstructured) – semi-structured groups (with interview guide) – structured groups (with interview schedule)
Number of moderators	– 1 or 2: whenever FG is conducted by two moderators, one should have specific skills in the focus group process and the other should have expertise in the content area of the study or could have the role of non-participant or participant observer who takes notes and observes the participants' non-verbal communication
Role of the moderator	– negligible (only provides the topic for the discussion and the rules of the interactions) – restricted (facilitates the flow of discussion, promotes interchange and modulates conflict) – directive (with a well-structured interview guide…)
Length	– from 60 to 120 minutes maximum
Strategies to encourage participation	– incentives (gadgets, an honorarium, transportation or light refreshments) – awareness raising strategy – Recruitment of participants interested in the topic
Data collection strategies	– Notes by a non-participant or participant observer – Audio recording – Video recording with or without a one-way mirror
Strategies of analysis	– Content analysis – Matrix based on the study objectives and on the questions asked

Source: Corrao, 2000

Elements that could change are:

- homogeneous vs heterogeneous groups;
- with or without incentives;
- with structured or non-structured interview.

Variants of focus groups include:

- two-way focus group: one focus group watches another focus group and discusses the observed interactions and conclusion;
- dual moderator focus group: one moderator ensures the session progresses smoothly, while another ensures that all the topics are covered;

- duelling moderator focus group (fencing-moderator): two moderators deliberately take opposite sides on the issue under discussion;
- respondent moderator focus group: one and only one of the respondents is asked to act as the moderator temporarily;
- client participant focus groups: one or more client representatives participate in the discussion, either covertly or overtly;
- mini focus groups: groups are composed of four or five members rather than 6 to 12;
- teleconference focus groups: telephone network is used;
- online focus groups - computers connected via the internet are used.

Planning a focus group

This paragraph presents, in greater detail, characteristics to be considered during the planning phase of the Focus Group.

Number of FGs: a minimum of four to five. The marketing researchers, for example, vary the number of groups according to whether or not meetings are producing new ideas. If the moderator clearly can predict what will be said in the next group, then the research is deemed as concluded. This usually happens after the third or fourth session. More in general, a FG can be iterative until the researchers reach their objectives, for example, until the final version of the questionnaire is generated.

Number of participants: eight to twelve. Smaller groups may be dominated by one or two members, while larger groups are difficult to manage and inhibit participation by all members of the group.

Sample selection: who will participate in the study depends on the purpose of the research. People are usually segmented by demographic factors such as geographical location, age, size of the family, status, gender, etc. The choice between homogeneous or heterogeneous groups somehow depends on the need to maintain a reasonable homogeneity within the group in order to encourage discussion. The most general advice is that each participant should have something to say on the topic and feel comfortable speaking with the others even if this does not mean that participants should have the same perspective on the topic.

Length of FG: one to two hours.

Number of moderators: one or two.

Moderator's role and characteristics: the moderator is the key to assuring that a group discussion goes smoothly. The focus group moderator should be well-trained in group dynamics and interview skills. Depending on the intent of the research, the moderator will direct the discussion a great deal or very slightly. It is important to recognise that the amount of direction provided by the interviewer does influence the types and quality of the data obtained from the group:

> More structured groups answer to the researcher's questions; less structured groups help to reveal the perspective of the group participants" (Morgan, 1988).

Facilities: although the FG can be conducted in a variety of settings, ranging from the home to offices and even by conference call, it is most common for focus group sessions to be held in facilities designed especially for focus group interviewing. Such facilities provide one-way mirrors and viewing rooms where observers may unobtrusively observe the interview in progress.

Focus group facilities include equipment for audio or video taping of the interview and perhaps a small transmitter for the moderator to wear (a "bug in the ear") so that observers may have input into the interviews.

Developing the Interview: essential to the FG is a careful development of an interview guide reflecting the research questions. It is important to begin with broad, open-ended questions and with low emotional intensity issues and then move to high emotional intensity issues. It is also important to have probes ready to prompt the focus group participants for further explanation or depth on the topics. The guide can represent merely a map or it can be a semi-structured guide or a structured guide, it depends on the purposes of the study and on the researchers' objectives.

Steps in the design and use of the focus group[1]

Preparing for the Session

1. Identify the major objective of the meeting;
2. plan the session (see below);
3. call potential members to invite them to the meeting. Send them a follow-up invitation with a proposed agenda, session time and list of questions the group will discuss;
4. plan to provide a copy of the report from the session to each member and let him or her know you will do this;
5. about three days before the session, call each member to remind him or her to attend.

Developing Questions

1. Carefully develop five to six questions - the session should last about 1.5 hours - in this time, one can ask at most five or six questions;
2. always first ask yourself what problem or need will be addressed by the information gathered during the session, *e.g.* examine if a new service or idea will work, further understand how a programme is failing, etc.;
3. focus groups are basically multiple interviews. Therefore, many of the same guidelines for conducting focus groups are similar to conducting interviews.

Planning the Session

1. **Scheduling** - Plan meetings to be 1-1.5 hours long;
2. **Setting and Refreshments** - Hold sessions in a conference room, or other setting with adequate airflow and lighting. Configure chairs so that all members can see each other. Also, provide nametags for members. Provide refreshments;
3. **Ground Rules** - It is critical that all members participate as much as possible, yet the session should move along while generating useful information. Because the session is often a one-time occurrence, it is useful to have a few, short ground rules that sustain participation, yet do so with focus. Consider the following three ground rules: a) keep focused, b) maintain momentum and c) get closure on questions;

1 See Carter McNamara. Basics of conducting focus groups. Available at: https://managementhelp.org/businessresearch/focus-groups.htm#anchor913016

4. **Agenda** - Consider the following agenda: welcome, review of agenda, review of goal of the meeting, review of ground rules, introductions, questions and answers, wrap up;
5. **Membership** - Focus groups are usually conducted with 6-12 members who can have some similar nature, *e.g.* similar age group, status in a programme, etc. Select members who are likely to be participative and reflective. Attempt to select members who do not know each other;
6. **Plan to record the session with either an audio or a video recorder**. Do not count on your memory. If this is not practical, involve a co-facilitator who is there to take notes.

Facilitating the Session

1. The major goal of facilitation is collecting useful information to meet the goal of the meeting;
2. introduce yourself and the co-facilitator, if used;
3. explain the means to record the session;
4. carry out the agenda (see "agenda" above);
5. carefully word each question before that question is addressed by the group. Allow the group a few minutes for each member to carefully record their answers, then, facilitate discussion around the answers to each question, one at a time;
6. after each question is answered, carefully reflect back a summary of what you heard (the note taker may do this);
7. ensure even participation. If one or two people are dominating the meeting, then call on others. Consider using a round- table approach, including going in one direction around the table, giving each person a minute to answer the question. If the domination persists, note it to the group and ask for ideas about how the participation can be increased;
8. closing the session - Tell members that they will receive a copy of the report generated from their answers, thank them for coming, and adjourn the meeting.

Immediately after the Session

1. Make any notations on your written notes, *e.g.* to clarify any scratching, ensure pages are numbered, fill out any notes that don't make sense;
2. write down any observations made during the session. For example, where did the session occur and when, what was the nature of participation in the group? Were there any surprises during the session? Did the tape recorder break?

Analysis of the data

Different kinds of analyses can be carried out using the collected information - from simple narrative descriptions based on the notes taken during the discussion of the focus group, to more complex analysis of recorded transcripts. The type of analysis depends, in addition to the time and resources available, on the research goals as well as on the initial hypotheses and on how the results are going to be used.

For exploratory research, a narrative description - taken from the transcript of the interview - enriched with the observations of the moderator and external observers (if any)

is often sufficient. For example, if the goal of a focus group is to get practical suggestions or ideas to be more profoundly studied at a later point, a minimum level of analysis can be sufficient such as comparing impressions among participants. If, however, important decisions will be based on the results, a more rigorous analysis is needed and all the focus groups need to be transcribed verbatim.

Data analysis can take place following two different approaches:

– strictly qualitative and ethnographic;
– "systematic": by coding the recorded transcripts through content analysis techniques (Morgan, 1997).

Actually, one can choose to use both methods since one does not exclude the other, but rather they complement one another.

For a systematic analysis of the data collected during the focus group, it is necessary to:

– build a reading/analysis grid to be applied to the entirety of the material (grid construction is facilitated by a semi-structured focus group);
– identify the range of opinions that emerged.

Content analysis techniques can be also applied to the full transcript of the discussion. Computer-assisted content analysis techniques are now available and can facilitate the task.

In general, the report should be partly composed of the report of the contents that emerged (issues), partly of quotations, and partly of the data interpretation (see also Stewart and Shamdasani, 1990).

However, the results of the focus groups can neither be extended nor projected on the entire population. The total number of participants does not involve a sufficient number of people for this purpose, nor does the sociological sampling strategy of the groups confer statistical projective value to the results.

Conclusion

The Focus Group has some advantages compared to other qualitative methods: it is comparatively easier to lead or conduct, it allows one to explore topics and to generate hypotheses, it generates an opportunity to collect data on the topic of the researcher's interest through group interaction, it gives speed in the supply of the results and it has low cost compared to other methods.

On the other hand, it takes effort to assemble the groups, it demands carefully trained interviewers in order to conduct the discussion in an atmosphere that facilitates dialogue, it is not based on a natural atmosphere and it is not possible to know whether or not the interaction in a group takes into account and reflects individual behaviour. Furthermore, the researcher has less control over the data that are generated, the data analysis can be difficult to carry out (Kruger, 1994; Morgan, 1988) and it is never possible to generalise or make statistical inferences from the results of the focus groups. It is because of this that focus groups are often used together with quantitative tools. Where the quantitative methods provide quantity and correlations, the qualitative methods, and especially the focus groups, facilitate the understanding of the motivations and the deepening of aspects deemed relevant.

In social and evaluative research, the use of focus group is recommended or in the first phase of the research, above all for new and unknown contexts, or in conclusion, as an in-depth study or verification of the results obtained. For some scholars and researchers this

is a technique to be used either way together with quantitative tools. In practice, the focus group is often used also in association with other qualitative tools such as the individual interview because, in these contexts, the subjects say partly different things.

3.3 Expert reviews

Expert reviews are frequently used as a method of evaluating draft questionnaires or subsets of questions, either alone or in combination with other pre-testing methods. This is a relatively quick and inexpensive method (Presser and Blair, 1994; Biemer and Lyberg, 2003), whereby experts identify any sources of non-sample error and suggest ways to minimise/reduce potential errors. In their review, experts have to ensure that the collected information meets the analytic objectives of the survey and check the design and the questions to detect problems of interpretation, breakdowns in the question-answer process, and other measurement errors in survey reports that may undermine data quality (Olson, 2010). Indeed, especially when the questionnaire is translated from another language and pre-testing methods involving respondents (*i.e.* cognitive interviews) are unavailable, expert reviewers must assess whether the questions may be misleading or open to misinterpretation by certain segments of the population and whether they match the designers' intentions across different cultures. Nevertheless, without specific information on how respondents interpret and answer the questions, it is very difficult for the experts to anticipate all the variables (respondents' backgrounds and life circumstances) that may affect data validity.

Expert reviews are usually conducted at the initial stage of the questionnaire development process, before the survey is either pretested or fielded, but they can also take place at a later stage. Experts provide the questionnaire designer with feedback on any pitfalls in the questions and recommend modifications. It is therefore recommended to adopt this method at a stage when suggested changes can still be enacted and an appropriate well-focused discussion with experts can be organised (Campanelli, 2008). Once the designer is satisfied that all the initial major problems raised by expert reviews have been solved, the revised questionnaire may undergo another review process or be submitted to another pre-test method.

It is recommended that reviewers be experienced survey methodologists with knowledge of the theoretical and practical aspects of questionnaire design, fieldwork issues, and data processing in addition to 'experts' on subject matters. Usually, few expert reviewers are engaged, between 2/3 and 20, depending on factors such as complexity of the topic under investigation, level of expertise of question designers, time and resources. Experts can be consulted independently or brought together as an expert panel. The latter is preferable, inasmuch as it allows differences in problem detection and recommendations to be debated.

As Grove stressed, all survey questions must meet three distinct standards: "content (*e.g.* are the questions asking about the right things?), cognitive (*e.g.* do respondents understand the questions consistently; do they have the information required to answer them; are they willing and able to formulate an answer to the questions?) and usability (*e.g.* can respondents and interviewers, if they are used, complete the questionnaire easily and as they were intended to?)" (Grove *et al.*, 2009: 259). Expert reviews assess whether these standards have been met, focusing specifically on: terms and wording; structure; order;

skip patterns/flow of the questionnaire; response alternatives; instructions to interviewers on questionnaire administration; confusing layout; typographical errors (Brancato *et al.*, 2006; Grove *et al.*, 2009; Biemer and Lyberg, 2003).

In their appraisals, the experts can either rely exclusively on their own judgment, making informal assessments that typically yield open-ended comments about the survey items to be evaluated, or be guided by formal appraisal systems that provide a detailed set of potential problem codes (Willis *et al.*, 1999).

Even though an informal review can highlight questionnaire problems effectively, there is still a high risk of missing significant issues. Therefore, it is common practice to adopt a structured questionnaire review approach. This type of review provides more comprehensive data, as each question is examined against a set of criteria coded to indicate the type of problem raised by each question, and helps the reviewer identify potential problems in the wording or structure of the questions leading to difficulties with question administration, interpretation, or cognitive processing (Biemer and Lyberg, 2003).

When expert panels are organised, the discussion is chaired by a moderator and for each question the debate may either follow a structured pattern using a standardised coding scheme, to be filled in before the meeting, or adopt a less structured, free-form approach without a standardised coding list (Brancato *et al.*, 2006). However, even in the latter case, the moderator steers the discussion towards topics like wording, terms, skip instructions, instructions, layout, comprehension etc. It is recommended that the questionnaire designer also participates in the group discussion but with an open mind, avoiding any aggressive or defensive behaviour. It is common practice to audio or video record the discussion to help with the compilation of a detailed report.

The draft questionnaire and other information must be sent to the expert allowing sufficient time for him/her to respond, if consulted independently, or prepare for the expert panel. In particular, the questionnaire must be accompanied by a short note that clearly states the key objectives of the survey and draws attention to the questionnaire design problems and issues on which advice is sought. Other relevant information includes any constraints related to the scope and design of the questionnaire (*e.g.* mode of administration, length, questions inserted for comparability with other surveys, etc.) and the coding scheme, if relevant (Campanelli, 2008; Blake, 2015).

Some expert review appraisal systems

Several standardised coding schemes are currently used. Back in 1992, Lessler *et al.*, (1992) developed a coding scheme for expert appraisal called Questionnaire Appraisal Coding System. This was an exhaustive guide, revised by Lessler and Forsyth (1996) to help the reviewer identify potential issues within a question, in accordance with the 4 stages of the question-answering process: comprehension of the question, retrieval of the information, judgment of the information retrieved and response processes; which, however, did not indicate the reason for the problem and how the question should be rephrased. This coding scheme was deemed by the Questionnaire Laboratory at Statistics Netherlands too detailed and extensive to be effective and efficient in assessing question errors. Hence, they developed the "Condensed Expert Questionnaire Appraisal Coding System", a concise coding scheme that takes into consideration only the most common issues, and requires experts to quote the problematic parts of the questions, in addition to coding the problems and suggesting improvements (Snijkers, 2002).

Table 3.2 Condensed Expert Questionnaire Appraisal Coding System

Problems in the questionnaire with regard to:		
Comprehension of the question	Information processing	Reporting
– Difficult wording => – Unclear wording => – Difficult syntax => – Long question with list of items – Double-barrelled questions – Double-negative questions – Question/answer mismatch – Reference set (perspective) change => – Response task => – ….	– Retrieval task => – Long period of recall – Much information needed to answer the question – Proxy reporting – Judgement task => – Difficult task (complex calculation, estimation) => – Social desirability – ….	– Difficult wording in answering categories => – Unclear wording => – Boundary problems => – Overlapping categories => – Missing categories => – ….

=> Indicates a description of the problem in the question, and suggestions for improvement.

Akkerboom and Dehue (1997) revised this coding scheme removing the classification of the potential errors within the four stages of the question answering process and considering eight main macro areas of problems.

Table 3.3 - An eclectic classification of measurement error risks to assess questionnaires

Problem	**Description**
Applicability /Suitability	*Question is not realistic enough:*
	– Non-existent or inaccessible data – Hypothetical or fictitious data – Data referring to someone else
Question meaning	*Question has unclear/ambiguous/unintended meaning:*
	– Misleading or unclear instructions – Reference set (frame) not sufficiently specified – Ambiguity or vagueness in question meaning – Unintended, though univocal, question meaning
Key concept meaning	*Key concept:*
	– Has unclear or ambiguous meaning – Is unknown or unnoticed
Cognitive difficulty	*High cognitive burden to respondent:*
	– Difficult recall/recognition – Difficult deduction (estimation, guess) – Difficult judgement – Complicated answer format
Technical difficulty	*Complicated question conditions / presentations:*
	– Too many key concepts or clauses – Double negation / other syntax complexity – Implicit assumptions – Unclear presentation (lay-out intonation)
Logical flaw	*Formulation or routing logically incorrect:*
	– Question-answer (Q/A) discrepancy – Incorrect/incomplete Q/A structure – Conflict with previous answers – Conflict with language rules
Motivation/Affection	*Undesirable questions:*
	– Goal of the question unclear or insufficient – Too intrusive or personal question
Social Norms	*Unbalanced, directive or non-neutral question:*
	– Non-neutral concept or directive task – Unbalanced or non-neutral answer – Risk of social desirability

Source: Akkerboom & Dehue 1997

As stated by the authors the scheme “should serve as a list of questionnaire design risk factors, which may be used by designers of questionnaire prototypes for a practical discussion on hypotheses to be (pre)tested, and for classifying pre-test findings and discussing possible actions to amend problems” (Akkerboom and Dehue, 1997).

Willis and Lesser produced a coding scheme called "Question Appraisal System (QAS)" that elaborated on previous question appraisal systems and consists of eight major coding categories, each containing several subcategories, for a total of 27 possible codes. The QAS guides experts through a systematic appraisal of survey questions helping them "to spot potential problems in the wording or structure of the questions that may lead to difficulties in question administration, miscommunication, or other failings" (Willis and Lessler, 1999). The expert is asked to apply the coding scheme to each question in order to decide whether there are any controversial aspects and, if any, explain the reasons; they may also record other comments on any question that fall outside the existing coding categories. The below table is an updated version by Willis.

Table 3.4 - The Question Appraisal System (QAS-2009)

Step 1 Reading: Determine whether it is difficult for the interviewers to read the question uniformly to all respondents		
1a. WHAT TO READ: Interviewer may have difficulty determining what parts of the question should be read.	Yes	No
1b. MISSING INFORMATION: Information the interviewer needs to administer the question is not contained in the question.	Yes	No
1c. HOW TO READ: Question is not fully scripted and therefore difficult to read.	Yes	No
Step 2 Instructions: Look for problems with any introductions, instructions, or explanations from the respondent's point of view		
2a. CONFLICTING OR INACCURATE INSTRUCTIONS, introductions, or explanations	Yes	No
2b. COMPLICATED INSTRUCTIONS, introductions, or explanations.	Yes	No
Step 3 Clarity: Identify problems related to communicating the intent or meaning of the question to the respondent		
3a. WORDING: Question is lengthy, awkward, ungrammatical, or contains complicated syntax.	Yes	No
3b. TECHNICAL TERM(S) are undefined, unclear, or complex.	Yes	No
3c. VAGUE: There are multiple ways to interpret the question or to decide what is to be included or excluded.	Yes	No
3d. REFERENCE PERIODS are missing, not well specified, or in conflict.	Yes	No
Step 4 Assumptions: Determine whether there are problems with assumptions made or the underlying logic		
4a. INAPPROPRIATE ASSUMPTIONS are made about the respondent or about his/her living situation.	Yes	No
4b. ASSUMES CONSTANT BEHAVIOR or experience for situations that vary.	Yes	No
4c. DOUBLE-BARRELED: Contains more than one implicit question.	Yes	No
Step 5 Knowledge/memory: Check whether respondents are likely to not know or have trouble remembering information		
5a. KNOWLEDGE may not exist: Respondent is unlikely to know the answer to a factual question.	Yes	No
5b. ATTITUDE may not exist: Respondent is unlikely to have formed the attitude being asked about.	Yes	No
5c. RECALL failure: Respondent may not remember the information asked for.	Yes	No
5d. COMPUTATION problem: The question requires a difficult mental calculation.	Yes	No
Step 6 Sensitivity/Bias: Assess questions for sensitive nature or wording, and for bias		
6a. SENSITIVE CONTENT (general): The question asks about a topic that is embarrassing, very private, or that involves illegal behaviour.	Yes	No
6b. SENSITIVE WORDING (specific): Given that the general topic is sensitive, the wording should be improved to minimise sensitivity.	Yes	No
6c. SOCIALLY ACCEPTABLE response is implied by the question.	Yes	No
Step 7 Response categories: Assess the adequacy of the range of responses to be recorded		
7a. OPEN-ENDED QUESTION that is inappropriate or difficult.	Yes	No
7b. MISMATCH between question and response categories.	Yes	No
7c. TECHNICAL TERM(S) are undefined, unclear, or complex.	Yes	No
7d. VAGUE response categories are subject to multiple interpretations.	Yes	No
7e. OVERLAPPING response categories.	Yes	No
7f. MISSING eligible responses in response categories.	Yes	No
7g. ILLOGICAL ORDER of response categories.	Yes	No
Step 8 Other problems: Look for problems not identified in Steps 1–7		
8. Other problems not previously identified.	Yes	No

Source: Willis 2015

Conclusion

Expert reviews are a quick pre-testing method, cost and time effective, that aims to correct or erase problems in the questionnaire without resorting to respondents. It relies on experts, individually or as part of a panel, who must be selected carefully based on their expertise in the questionnaire subject matter and in survey methodology.

Citing several studies, Olson (2010) explains the reasons why experts might differ as to the number and types of problems that they identify in survey questions as well as in their improvement recommendations. These include, for example, differences among experts about the time spent on the task, expectations about the task, perceptions about "good" or "bad" questions, experience in performing questionnaire evaluations and familiarity with the subject matter, the quality of the information received, regarding the objectives of the survey and the questionnaire administration mode, and finally the degree of review standardisation.

The use of a standardised task simplifies the process and takes away some of the variations due to the above reasons, however, it should also be considered that these variations contribute to a more exhaustive review process. Lastly, panel reviews also provide a useful opportunity to address in an in-depth discussion any discrepancies and inconsistencies either with the problems or with the suggested improvements.

Expert reviews are widely used as a first step in question evaluation inasmuch as not only are they useful to uncover question aspects that may affect accuracy but they also highlight issues to be evaluated by other pre-test method such as cognitive testing.

3.4 Cognitive interviews

Among the pre-testing techniques aimed at improving the survey instrument and consequently the data quality, cognitive interviewing has emerged as one of the prominent pre-testing methods for identifying and correcting problems in survey questions (Beatty and Willis, 2007). Indeed, more than any other question evaluation method, cognitive interviewing offers a detailed report of meanings and processes used by respondents for answering questions, providing a wide range of information to the researcher about sources of response error. In particular, it can identify sources of response error that the interviewer and the survey practitioner may fail to see, which have to be eliminated/reduced to ensure accuracy of the survey instruments and high quality of the data collected. Through cognitive interviewing it is possible to determine whether the wording of a specific question fully conveys its meaning and to quickly identify problems such as redundancy, missing skip instructions and awkward wording.

The origins of cognitive interviewing are rooted at the intersection of survey methodology and cognitive psychology; in particular, the interest in cognitive testing originated from an interdisciplinary seminar on Cognitive Aspects of Survey Methodology (CASM), when cognitive psychologists and survey researchers met to study the cognitive factors that influence survey responses and the ways in which survey questions should be designed to take such factors into account. The CASM movement not only brought attention to the issue of measurement error, but it also established the idea that individual processes, especially respondents' thought processes, must be understood in order to assess the validity and potential sources of error (Schwarz, 2007).

Cognitive interviewing relies on the cognitive model introduced by Tourangeau (1984). This model explains that the "question answering process" is divided into four stages: comprehension, retrieval, judgement/estimation, and finally response. As the cognitive process leading to question answering is a universal process that occurs in the mind of each and every respondent, the latter will go through those four steps in order to formulate his/her answer, regardless of demographic or personal background.

The CASM approach relies on the assumption that, when a survey question fails, this is due to problems occurring at one (or more) of these stages that cause the incorrect answers. Thus, the main purpose of a cognitive test is to identify problems that may occur at each stage of the answering process.

Stage 1 Comprehension: the respondent has to understand and interpret the question; the cognitive interviewing goal is to assess whether the respondent understands the question as intended by the researcher who designed it and uncover problems due to unknown terms, ambiguous concepts, long and complex questions.

Stage 2 Retrieval: the respondent searches his/her memory for relevant information to answer; the cognitive interviewing allows to establish whether the information needed can actually be recalled, what and which strategy has been used to retrieve information and uncover problems such as recall difficulty, no prior knowledge or experience, perceived irrelevance of the topic.

Stage 3 Judgement/Estimation: the respondent evaluates or estimates while deciding on an answer; the cognitive interviewing evaluates how much mental effort it takes the respondent to answer the question, whether the question wording could lead to more "socially desirable" answers. It can identify questions that could cause embarrassment or defensive behaviors, and uncover problems such as biased or sensitive questions and difficulty in judgement.

Stage 4 Response: the respondent provides information in the requested format; the cognitive interviewing verifies whether the respondent correctly interprets the answer categories provided, whether such categories match answers spontaneously generated by the interviewee, and uncover problems such as incomplete response options, response options not in accord with understanding or judgement of question and unwillingness to answer.

To recap, the cognitive interview improves data collection instruments by providing evidence on whether respondents understand questions in the intended way, are able and willing to provide the requested information and respond using the answer options provided.

3.4.1 Cognitive interviewing techniques

Beatty defines cognitive interviewing as "the practice of administering a survey questionnaire while collecting additional verbal information about the survey responses; this additional information is used to evaluate the quality of the response or to help determine whether the question is generating the sort of information that its author intends" (Beatty, 2004). This definition describes the most common application of this pre-testing technique: in-depth interviews with a small sample of respondents in which respondents first answer a draft survey question and then provide textual information on the cognitive process behind their answer. In particular, respondents are asked to describe how they came upon their answer and their rationale, and this information is then used to understand whether the questions will actually produce the information they are meant to provide.

This pre-testing method actually comprises a set of techniques, and although there has been a debate over interviewing practices[8], a full consensus on best practices or a set of standards is still lacking. Its implementation modalities vary from organisation to organisation, and even among researchers in the same organisation, depending on several factors such as aims of the cognitive test, developmental stage of the questionnaire, and time and resource constraints. The most commonly used techniques, here presented, are think-aloud and verbal probing, recently often combined, while other techniques are less frequent based on the specific intent of the cognitive test.

In the first technique, called "think aloud", respondents are encouraged to verbalise their thoughts as much as possible while answering questions. In particular, concurrent think aloud means that the respondent is asked to articulate all his/her thoughts immediately after the question is asked until he/she comes up with the answer; while in the retrospective think aloud the respondent explains how he/she came upon their answer either immediately after answering each survey question or after completing a session of the questionnaire or even the entire interview.

On one hand, disadvantages of the concurrent think aloud technique are the high burden on the respondent, the disruption of the normal flow of the interview and the fact that expressing one's thoughts aloud might alter the cognitive process under study. On the other hand, while the retrospective think aloud technique is less disruptive to the flow of the interview, in-depth information on the various steps of the answering process might be lost due to the time gap between the response process and the verbal report (Brancato *et al.*, 2006).

While in the think aloud approach the role of the interviewer is, generally, extremely low key, the second technique, usually called "verbal probes", requires the interviewer to play a more proactive role, asking specific follow-up questions (probes) aimed at collecting specific information related to the question interpretation or to the answer. Probes can be classified in various ways depending on: when they are asked (concurrent vs retrospective); how they are administered (proactive vs. reactive); what is their nature (standardised vs. spontaneous or emergent); what is their aim (general vs specific) and, for which phase of the cognitive answer process they are used (Comprehension, Retrieval, Judgement, Response)[9].

Both techniques share the same goal that is to generate verbal information normally missing in a survey interview in order to assess how the questions measure up to their objectives, and both have pros and cons (Beatty and Willis, 2007). According to Willis (2005) the advantages and disadvantages of the two techniques are the following.

Think aloud technique:

Pros

- freedom from interviewer-imposed bias: given that the interviewer's role is restricted to reading the survey question, except for occasionally asking what the subject is thinking, he or she has little chance to influence the subject's responses;
- minimal interviewer training requirements: again, since the interviewer basically reads survey questions, and then listens to the respondent, little training or special expertise is usually necessary;

8 See Beatty (2004); Conrad and Blair (2004, 2009); Willis (2005, 2009); Beatty and Willis (2007), DeMaio and Rothgeb (1996); DeMaio *et al.*, (1998); Miller (2014).

9 See the box "Example of probes" page 52.

– open-ended format: as the respondent's verbalisation is only minimally guided, he or she may provide unexpected information. Therefore, think-aloud interviewing is especially valuable when the subject is outgoing, articulate, and has had significant experience with the topics covered by the survey questions.

Cons

– need for respondent training: since thinking-aloud is unusual for most people, the technique typically requires a significant amount of preliminary practice, so that the respondents may familiarise themselves with the technique;
– respondent resistance: notwithstanding the training many individuals are not at ease when thinking aloud and they tend to simply answer the survey questions, without further elaboration;
– risk for the respondent to stray from the task: Under the think-aloud technique, the respondent controls the nature and the direction of the discussion. Therefore, it is easy for the respondent to wander off-track, and to spend a significant amount of time on one question, so that the interviewer must struggle to "bring the subject back on track." In general, the think-aloud technique results in relatively few survey questions being tested within a definite amount of time, compared to the other approach;
– bias in respondent information processing: by its nature, thinking-aloud requires subjects to think. As such, subjects may invest a considerable amount of mental effort into processing the survey questions, which is not needed when they simply answer the questions. This extra effort may contaminate the cognitive processes put in place to answer the question.

Verbal Probing technique:

Pros

– control of the interview: the use of targeted probing to guide the respondent steers the interaction in a way that is controlled mainly by the interviewer, who can focus on particular areas that appear to be relevant as potential sources of response error while avoiding irrelevant and non-productive discussion;
– no need for participant's training. It is fairly easy to induce the interviewee to answer probe questions, as they are often administered in the same way as the survey question. In fact, subjects will sometimes begin to expect probes, and to offer their own spontaneous thoughts and criticism, so that the interview comes to resemble a think-aloud.

Cons

– artificiality. The validity of verbal probing techniques is occasionally questioned, because the interjection of probes by interviewers may produce a situation that is different from the usual survey interview, in which the interviewer simply administers questions and the respondent answers them. However, note that the verbal probing technique is certainly no more unrealistic than the alternative of thinking-aloud;
– potential for Bias. A related criticism is that the use of probes may lead the respondent to particular types of responses. This is of course possible, but can be controlled through the careful selection of "non-leading" probing techniques that minimise bias. In other words, probes should be characterised by unbiased phrasing, in the same manner that survey questions are intended to.

It is generally advised to develop a cognitive interviewing technique that combines think-aloud with verbal probing, considering the objectives of the cognitive test as well as the ability of the respondent to think aloud. A hybrid between the two approaches is a practical solution

that seeks to maximise the autonomous flow of thought of respondents on questions without hesitating to probe when it is deemed appropriate in order to obtain the information needed (DeMaio *et al.*, 1998; Beatty, 2004; Willis, 2005; Beatty and Willis, 2007; d'Ardenne, 2015).

3.4.2 Planning a cognitive test

In order to test a questionnaire or some questions by cognitive interviewing, several factors must be taken into account, such as: the development stage of the questionnaire/questions, time and financial resources, expert or non-expert interviewers, type of sample to be enrolled, whether other pre-test methods and more than one round of cognitive interviews are foreseen, location of the cognitive interviews, paperwork required, and, last but not least, some ethical issues.

The time needed to implement a cognitive test depends on a number of other factors, including: ease of recruitment of the respondents, number of testing rounds and interviews per round, number of interviewers and questions to be tested, and finally the actual deadline.

In order to carry out a cognitive test, the below main steps must be followed thoroughly:

- development of the interview protocol;
- sampling of respondents;
- recruitment of respondents;
- selection of interviewers;
- training of interviewers;
- conducting the cognitive interviews;
- data management, analysis and interpretation;
- writing the cognitive test report.

Developing interview protocols

The cognitive interview protocol is the guideline that all interviewers are required to strictly adhere to in order to reduce the interviewer's effect. The protocol focuses on the following main information: a) aims of the testing and how to introduce it to the respondent, b) how test questions should be administered, and c) which cognitive interviewing techniques should be used and how.

The protocol is closely linked to the type of cognitive interview technique, which may vary along a continuum from "think aloud" to a highly standardised "verbal probing", and therefore it should be adjusted accordingly. Nevertheless, irrespective of the type of approach, the protocol usually includes a standardised introduction to respondents to explain the test objectives and how the test will be performed, and how confidentiality and privacy will be protected.

If the "think aloud" approach is applied, the interview protocol may simply consist of a list of questions to be tested, along with the recommendation to the interviewer to keep reminding the respondents that they should express their thinking processes and thoughts aloud when answering questions.

As mentioned before, the "think aloud" method can be concurrent (while the respondent is providing the answer) or retrospective (elicited by the interviewer after the answer is provided using precise pre-scripted or spontaneous probes), in any case this must specified in the testing protocol and explained to the respondent. The interview protocol should also instruct the interviewers on how to introduce the think aloud approach. d'Ardenne and Collins (2015) provide specific indications in this regards:

- explain to the respondent what thinking aloud means by saying for example: "Please say "out loud" anything that comes into your head whilst you are answering […]" or "When you answer the questions I would like you to tell me what you are thinking […]" or "Please tell me what is going through your mind as you are considering each of your answers […]";
- give a practical demonstration with a simple example such as: "Try to visualise the place where you live, and think about how many windows there are. As you count up the windows, tell me what you are seeing and thinking about." as suggested by Willis (1994);
- let the respondent practice the think aloud technique before the interview takes place. Not all respondents are at ease with this approach, so the interview needs to offer positive feedback and praise to encourage reticent participants to continue talking during the practice.

Keeping in mind that not all respondents feel comfortable thinking aloud and may, therefore, not provide sufficient information, or may not verbalise all of their thinking processes, some probes can help elicit more information. It is up to the interviewer to decide during the interview which types of probes to use.

When the "verbal probing" approach is adopted, it can be more or less standardised depending on the type of probes. In a highly standardised one the probes are pre-scripted, meaning that they have already been decided by the researcher and written in the protocol with the exact wording to be used and in the exact position where they have to be asked. In this case, the interviewer simply has to strictly follow the testing protocol to obtain a structured interview. One advantage of using pre-scripted probes is that all the testing aims are addressed and the same areas are explored consistently for every participant.

Nevertheless, it is quite common that especially very experienced interviewers rely also, or exclusively, on spontaneous probes. In this case, the interviewers are allowed to generate their own spontaneous probes to explore any issues raised during the interview (even if these have not been anticipated in the protocol), to clarify what the respondents said, or to deeply investigate inconsistencies in their answers. To facilitate the interviewer's job, a list of suggested probes can also be added to the protocol. This is a less standardised testing protocol.

Generally, in the cognitive interview protocol the probing questions are highlighted with different colours, typeface or placed in a box labelled "probe". The definition and formulation of the probing questions is critical to enhance the effectiveness of the cognitive test in capturing the point of view of the respondents. Therefore, the researcher has to conduct an in-depth questions' evaluation in order to identify any possible difficulties that can be addressed during the test, and which the probes can help shed light on. Effective probes must be: unbiased, non-leading, open-ended and straightforward (*e.g.* not long-winded, double-barrelled or with multiple clauses). Furthermore, the probes should be limited in number and relevant to the purpose of the testing. Unnecessary probes will put too much strain on the respondent who may get tired quickly jeopardising the accuracy of the information provided during the entire interview. While it is important to cover all the objectives of the testing through the probes, a good balance should be found between accuracy of the information and length of the interview.

Depending on the development stage of the questionnaire and the test objectives, either approach (think-aloud and verbal probing) or a combination of the two may be used during the entire interview or in some parts thereof.

Once the cognitive interview protocol has been developed, it is important to test its effectiveness through mock interviews with colleagues or other people. This will bring up any methodological issues, including whether the probes are all necessary, appropriate, well-phrased, and whether instructions are clear and exhaustive. The mock interviews will also help to assess how easy it is to administer the testing protocol and the time required. If the testing protocol is too long and complex, tiredness on the part of both the respondent and the interviewer may affect the interview outcome. In that case, it is preferable to cut some questions or unnecessary probes, or to develop more testing protocols. The latter option is generally applied when there are numerous questions under test and each participant is only asked a limited number thereof; obviously, in order to get meaningful results, a much larger sample of respondents is required.

Sampling of the respondents

Examples of probes

General probes:

How did you go about answering that question? / How did you arrive at the answer?
What went on in your mind when you were asked the question?
Was that easy or difficult to answer? Why?
Why do you say that?
Can you tell me a bit more about that?

Comprehension probes:

What does the term "----" mean to you?
What did you understand by "-------"?
In your words, what is "------"?
What, to you, is "--------"?

*Paraphrasing probes: **

What would you say that question was asking of you?
How would you say that question to yourself?
Can you repeat the question in your own words?
Could you tell me in your own words what this question is asking?
*Obtain information about the common language usage of the target population but they may not work in all situations because respondents may feel embarrassed if they really do not understand (or cannot articulate), what a question is asking.

Retrieval probes:

How did you remember that?
What brought that to your mind?
What time period were you thinking about when you answered this question? (From when to when?)
When did you last do "-----"?

*Judgement probes: **

How sure of your answer are you?
What did you think about when deciding how to answer this question?
How accurate would you say your answer is? Why?
Did you try to count each time you [did ----], or did you make an estimate?

**May not work in all situations*

Sensitivity probes:

How did you feel about answering this question?
Did you find this question embarrassing to answer? Why?
Do you find this question too personal/intrusive? Why?
Do you think other people would find this question sensitive? Why?

Response probes:

How easy or difficult was it select an answer from the options provided? Why?
Why did you choose that particular answer rather than one of the others?
Are there any categories missing from the options provided or do they cover everything? What is missing?
Was it easy or difficult to choose that particular answer? Why?

Remind and encourage probes:

"Can you tell (me) more what are you thinking?"
"Keep talking…"
"You are doing really well"
Examples of Spontaneous Probes
Probes based on Observation

– I noticed you were spending some time with that question - can you tell me what you were thinking about?
– I noticed you hesitated before you answered - what were you thinking about?
– You answered that very quickly - why was that?

Regarding the sample, two main issues need to be addressed: sample characteristics and size.

Cognitive Testing is a qualitative method where in-depth interviews are conducted paying explicit attention to the respondent's thinking process in answering survey questions. Since the interviewer is interested in the process behind the answer rather than in the reply itself, a purposive sample of the respondents is used. It is important to clearly define the sample selection criteria, which must ensure diversity within the sample so that differences in interpretation or problems can be explored and that all possible interpretations of and answers to the questions by the survey's target population are covered (Collins and Gray, 2015). Moreover, it is important to select respondents with the most diverse socio-demographic characteristics and living in different geographical contexts to increase the scope of the experiences and thus of the issues that may arise (Miller, 2002).

There is not a general consensus on the ideal number of interviews; the right number of participants in cognitive interviewing "may be hard to specify beforehand as some problems may be easily identified, while others are harder to root out, perhaps because they are relatively subtle or affect only respondents with certain characteristics or experiences. Similarly, some response tasks may pose difficulties for most respondents, while others affect only a small proportion of respondents, but perhaps cause serious measurement error when they occur" (Blair *et al.*, 2006: 4041). The size of a purposive sample for a cognitive test tends to be, at any rate, small (10-30) because it is also influenced by the time available for the test, budget and other resource constraints and whether other pre-testing methods are foreseen. Obviously, the larger the sample the higher the probability to observe a given problem or reveal more and diversified problems.

According to Willson and Miller:

"The ultimate number of interviews is based not on a particular numerical goal, but on the ability to construct a theory that explains why respondents answer questions the way they do and, in the end, the construct (or set of constructs) that a particular question captures" (Willson and Miller, 2014).

Therefore, cognitive interviewing is expected to continue until no new problems or new patterns of interpretation are discovered, and this may require more than one round of testing. Willis (2005) suggests planning at least two rounds since "the nature of the CT rounds tends to change over the course of the development of a questionnaire. Early in the process, findings relate not only to particular question wordings but to more global issues, such as the appropriateness of the survey measurement of major concepts that is attempting to cover, while later rounds focus mainly on polishing specific question wordings". Of course, any changes to the questionnaire or to individual questions resulting from a testing round need to undergo a further cognitive test.

Recruitment of respondents

The recruitment process is often the most time-consuming part of a cognitive test. It is important to address two main issues: who is responsible for the recruitment and which method will be used.

Based on time and resources, and other factors such as sample size and individual characteristics, a decision on who will be in charge of the recruitment process needs to be made. Some institutions have a specific Unit or Centre in charge of all cognitive tests and its members are also responsible for recruiting potential respondents; others rely on researchers or cognitive interviewers to find the participants; still others tend to use *ad hoc* external agencies.

There are a number of different methods to recruit people that can be used, alone or in combination, based on the types and number of participants needed for the sample and the resources available (Collins and Gray, 2015). The most common methods are:

Direct recruitment approaches prospective participants face-to-face, by telephone, SMS, or email. It is possible to recruit people directly by going door to door or setting up stands in strategic locations where the types of persons sought to be interviewed are easily found.

Snowballing or chain sampling uses participants who have already been recruited and/or interviewed. These may help to identify other people they know who might fit the sampling criteria. This technique is useful when a quite specific, or even hidden, subset of the population must be engaged and it is supplementary to the main recruitment method generally. A research project information leaflet needs to be prepared for distribution to the participants, with contact details.

Using advertisements includes putting up posters or leaflets on notice boards in places where prospective interviewees can be easily found, distributing leaflets or flyers, advertising in newspapers and newsletters. Advertisements can also be virtual, for example on websites, online forums, or social media.

Recruiting via organisations or groups is helpful to recruit a specific target of people that are likely to be members or users of, or affiliated with particular groups, organisations, or services. This may save time in the recruitment process but the individuals selected in this way may not reflect the variety of persons of this group.

Pre-existing sampling frames can also be used. This may include previous research participants who have given consent to be re-contacted for future studies and it can be

a very efficient recruitment method because you potentially already know at least some information about the people you are going to approach.

One factor can that influence the time needed to recruit participants and facilitate the availability of participants is the financial incentive. Indeed, if the budget allows it, it is common practice to provide a small financial incentive (cash, voucher, etc.) as a token of appreciation and acknowledgement of the effort and time involved in taking part in the cognitive interview. Participants are requested to sign a receipt at the time of the cognitive interview. A reimbursement for travel or other expenses should also be considered, if appropriate.

Selection of interviewers

When interviewers are selected and trained it is of the utmost importance to keep in mind the role that they will assumers during the cognitive interview. This role may vary from being a simple "data collector" to acting more as a "detective"/analyst (Willis, 1999; Beatty and Willis, 2007). The first type of role is based on the assumption that the minimisation of interviewer interaction will result in lower bias and higher data accuracy and validity. Therefore, when a think-aloud approach is chosen, the interviewer will remain as silent as possible; and when verbal probing is required, he will use standardised and structured probes so that every interviewer asks the same set of probes in the same way. Thus, the interviewer behaves as if he/she were administering a survey question, inasmuch as "the interviewer's job is not to look for contradictions and gaps in respondent explications, but to administer the interview protocol, without deviation, as accurately and systematically as possible" (Willson and Miller, 2014). In this case, the interviewers do not need to be particularly high-skilled but should possess, or be trained to develop, the skills and techniques suitable for conducting interviews by following a rather structured protocol, which may include some set probes. Detailed training, which includes the description of the cognitive objectives and also covers the fundamentals of the research methods, would be nice-to-have but are not mandatory.

Instead, the interviewer acts as an "investigator" when the cognitive interview has "an exploratory character being more attuned toward generating ideas about potential problems than determining their extent in the population" (Beatty and Willis, 2007). If so, the interviewer plays a more active role, relying at least partially on the so-called "emerging probes" to explore and analyse the content of each individual interview at a deeper level. According to Willson and Miller (2014). During the interview, he/she must assess the information provided by the respondents, examine the emerging information to pick up any gaps, contradictions or incongruences in the respondent's narrative and then ask probes that can clarify those. This entails the need for more experienced interviewers whose training ranges from the study of questionnaire design techniques, to the types of errors and their possible effects on the accuracy and validity of the questionnaire, to the knowledge of the goals associated with each of the questions under test (Willis 2005; Beatty and Willis, 2007). Generally, good cognitive interviewers have specific prior experience, and are familiar with the questionnaire topics, design and survey measurement problems (bias, context effect, acquiescence, social desirability, etc.).

Conducting a cognitive interview requires multifaceted skills and qualities on the part of the interviewer. Technical ability is essential but interpersonal skills are also important: such as putting the interviewee at ease, being a good listener, flexible, spontaneous, non-judgemental and able to manage any emotional reaction by the respondent.

To recap, according to Willis (2005) the ideal cognitive interviewer is, at the same time, a technician who can administer the test protocol, a detective who pinpoints problems in the questions, and an engineer who works toward developing feasible solutions to the emerging issues.

Training of interviewers

It is essential that the cognitive interviewer receives specific training on the project at hand, even if he/she has already conducted several cognitive tests.

Some National Statistical Offices (NSOs) have an internal unit entrusted with all survey pre-testing and, therefore, already have experts in cognitive interviewing; nevertheless, even expert interviewers need time to learn about the specifics aims of the test, get familiar with the testing protocol and practice the administration of the specific protocol. In other NSOs researchers or field interviewers may be occasionally involved in performing cognitive tests, thus these are definitely in need of training, given the peculiarities of cognitive interviewing. Indeed, the main difference between qualitative and cognitive interviews is that while the qualitative interviewer wants to know the details of the respondents' answers and of their life-experiences, the cognitive interviewer is not directly interested in the answers to the survey questions but rather in the cognitive processes that led the respondents to the answers and what difficulties or ambiguities they had to overcome. In particular, Willis (2005) recommends that the training focus on some fundamental differences between the two types of interviews. Therefore, the interviewers should always keep in mind that: a) the main objective of the cognitive interview is to identify any problems with the questions and not to "run" a question at all costs, for example by reformulating it if unclear or providing explanations if requested; b) the cognitive interview must necessarily proceed at a slower pace as the respondents need time to think and explain their thought process; c) they are often asked to work with less sophisticated questionnaires as regards formatting and filters, so flexibility is often a requisite.

In preparing the training course, some aspects are especially relevant: what the interviewers need to know in order to correctly play the role envisaged by the chosen technique, how much time is needed - but also available - for the training, which documents have to be distributed (*e.g.* project background and instructions, test questionnaire and interview test protocol, sampling and recruitment information…) and which training supplies are necessary (*e.g.* a video recorder for the mock interview could be very useful).

During the training, it is important to go over all the methodological and practical aspects related to conducting a cognitive interview in detail as the interviewer needs to become familiar with the principles and objectives of the cognitive approach and with the questionnaire and each question. Practising the testing protocol is critical and this may be done in many ways, by watching experienced interviewers conduct an interview, carrying out mock cognitive interviews with peers and other people with similar characteristics of the sample, interviewing a more experienced interviewer, who will replicate some of the difficult situations that may arise during the real interviews, and by watching videos showing "good" and "bad" cognitive interviews.

Furthermore, it is crucial to furnish detailed information on how to organise and analyse the information and how the interview results should be reported.

Conducting cognitive interviews

Before starting the cognitive interview phase, some logistical aspects[10] must be sorted, such as location where the interview will take place, equipment needed, number of interviews per day, number of cognitive interviewers and how confidentiality and privacy will be fully respected.

Interview location. The interview may take place in a cognitive laboratory, if available at the institute, or in any quiet and private room, such as a conference room or an empty office. It may even be carried out at the participant's home or at locations (*i.e.* offices of service agencies, churches, libraries) that are comfortable as long as privacy can be maintained. As the location set-up is also important, the interviewer needs to check beforehand that the venue is appropriate, especially when it is the participant's home or any other public place. It is necessary that both the interviewer and the respondent are at ease and that the interview be conducted without external interruption or distraction.

Equipment. It is common practice to record cognitive interviews and sometimes videotape them. Video recording is not always feasible on account of unsuitable location, unavailability of equipment, or lack of consent from the respondent; nevertheless, its use makes it possible to study the behaviour of the respondent, too (as well as the performance of the interviewer for further improvement). Moreover, depending on the type of survey instrument being tested, *e.g.* a self-compilation questionnaire, video recording may be a requisite. In this case, it is important that this information is clearly highlighted in the recruitment materials and/or during the face-to-face recruitment to avoid any last minute refusal from the respondent.

The recording (audio or video) frees the interviewer from the pressure of having to take detailed notes, so he/she can focus on listening to the participant and asking the appropriate probes, and the recording provides a full account of the interview, which is essential for further in-depth analyses of questions' performance.

Number of interviews per day. Considering that a cognitive interview requires a lot of attention and concentration on the part of the interviewer, a limited number of interviews should be conducted in one day.

This number may vary depending on the length and complexity of the testing protocol, the interviewer's expertise as well as the respondent's own characteristics (*e.g.* persons with intellectual difficulties).

The length of each interview depends on several factors (Willis, 2005; Gray, 2015): the complexity and number of questions under test, approach adopted (think aloud verbal probing or their combination), amount of probing as well as respondent's ability to think aloud or the overall speed at which the targeted survey questions or the probe questions are answered. Furthermore, the number of skip patterns also affects the number of questions the respondent has to answer and be probed on and the length of the interview. There is a general consensus that around 1 hour is a reasonable time for a cognitive test as longer interviews would put too much strain on both interviewer and respondent. Another aspect to keep in mind is that the interviewer may have to travel to the location of the interview, which is time-consuming.

Willis (2005) recommends a maximum of three interviews, especially when they adopt a less structured testing protocol, which gives the interviewer greater autonomy and the possibility of interaction with the interviewee.

10 See Willis 2005 for more information.

Number of cognitive interviewers. Based on resources, time available and the expected number of interviews, the number of the interviewers may vary; of course, the more interviewers are involved the faster the project moves forward. More interviewers mean a higher chance of identifying problems and ensure a richer discussion on how to solve them; nevertheless, this requires additional effort to ensure consistency in the way the interviews are performed, especially when spontaneous probes are allowed. In this regard, training and interviewer briefing are even more essential.

Respect of privacy and confidentiality. This issue is particularly relevant especially when private and sensitive information is to be gathered. Details that could potentially identify participants must not be included in any report or any other type of research outputs, and the data collected have to be used only for the purpose stated. Names and contact details, the tape interviews and other documentation must be stored under lock and key, labelled with subject ID numbers rather than personal identifiers.

When these issues have been clearly defined and taken care of, then it is possible to start with the cognitive interviews as planned. As already mentioned, in order to obtain high-quality information on question performance, the interviewer must be confident and at ease with all the steps of the interview and necessary paperwork. Even if a training course has been attended to gain a deeper knowledge of the specific cognitive interview technique, it is important that he/she revisit the research objectives, practice the administration of the questions under test and of any probes as well as the introductory statements. In brief, the interviewer must master the interview protocol guidelines. The interviewer needs to be familiar also with all the documents to be read or handed out to the participant.

Before making their way to the venue of the interview, he/she must be sure to have everything required such as cognitive interview protocol, consent forms, participant incentives and receipts for incentives, audio or video recorder, ID card and/or institutional badge, computer/tablet/mobile device to be used by the respondent. Furthermore, to avoid any problems during the interview, it is recommended to check and familiarise oneself with the equipment (recorder or video recorder) and have spare batteries. It is also, polite to call or send in advance a text message to the participant as a reminder of the appointment and to avoid any misunderstandings about the day, place and time of the interview.

In any interview, there are three phases: introduction, actual cognitive interview and its end, for each of which crucial aspects have to be considered as follow.

Introduction to the interview. Before starting the actual interview, the prospective respondents need to receive detailed information to be able to make an informed decision on whether to participate, even if they know something from the recruitment contacts. Main information is about: who is conducting the research, what is the topic of the questions, how long the interview will take, whether the interview will be recorded and how the collected information will be used and by whom. Furthermore, they need to understand that their participation is on a voluntary basis, and that they can skip questions they do not want to answer and even withdraw at any time during the interview. The prospective respondents need to know how privacy and confidentiality will be maintained during all phases of the cognitive test and for all material collected. It is also important to let them know whether they will get a token for their participation and/or reimbursement for travel expenditures. At this point, the respondents are requested to sign consent forms: a) for participation in the interview; b) for recording; and c) a specific consent form if the interview is to be videotaped; the last two have to be distinct because some respondents may not consent to both.

After consent forms have been signed, the interviewer can proceed to explaining in detail how the interview will be carried out, stressing that the purpose of the interview is to test the questions and not the respondent, and that what is important is how the respondent arrived at an answer to each survey question and any problems encountered in this process. When the think aloud approach is adopted, this must be clearly explained and the respondent be given sufficient time to practice and feel comfortable with it[11].

Of course, it is important to thank the person for his/her participation and to encourage the participant to express any questions or doubts so they can be cleared and he/she feels at ease and confident during the interview. This introductory phase also works as an "ice breaker".

The cognitive interview. When the respondent is ready, the cognitive interview can start with easy background questions, usually the social-demographic ones; not only are these questions useful for a deeper analysis of the interview results, they also serve as a "warm-up", getting the participants to talk about themselves.

The interviewer has to find the right pace and ensure that all topics are covered and that the respondent provides useful information without feeling embarrassed, anxious or frustrated. Particular attention should be put on the respondent's behaviour as it may betray his/her discomfort, tiredness or reluctance to answer some questions, thus affecting the results of the test. The interviewer has to acknowledge any respondent's difficulty, reassure him/her, and if appropriate, reaffirm the purpose of the interview. The quality of the information obtained depends on the respondent's level of attention and motivation, therefore when those dwindle and signs of fatigue appear, it may be better to cut the interview short. If only few questions are left, the respondent may be asked whether he/she is willing to complete the questionnaire. Occasionally, the interviewer may be unable to go through the entire questionnaire because the respondent is too talkative, needs time to answer, asks for clarifications, or because the protocol is too complex. The interviewer must also be able to cope with any emotional outburst on the respondent part by keeping calm and professional, giving the respondent time to recover from the distress or offering to move on. If several interviewers report similar incidents during debriefings, this means that there is a flaw in the interview protocol that needs to be addressed.

The interview must follow the testing protocol. Before moving on to the next question, the interviewer must be sure that he has collected all the information needed to correctly assess the question's performance. It is always better to test fewer questions and collect in-depth information on the question's performance than to complete the testing protocol with information that is not useful for the analysis.

In carrying out the cognitive test the interviewer usually keeps some basic rules in mind:

- leave enough time for the respondent to think and answer, don't put words into the respondent's mouth and never use leading probes;
- ask instead of assuming or pretending to have understood what the respondent said and when summarising check with the respondent to avoid errors or misunderstandings;
- bring the discourse back on track if the respondent strays to non-relevant issues and kindly prompt the excessively talkative respondent to move on when enough information on one question has been provided;
- play questions back to the respondent when he/she asks for clarification and do not offer explanation on the question but ask what it means to him/her;
- accept changes to answers already given to the test questions and investigate the reasons;
- listen to the respondent carefully, taking only a few notes that may help you probe to clarify points, especially if the interview is recorded;

11 See the cognitive interviewing protocol above.

- do not assume that the respondent had no problems with a question just because no other interviewee had problems with that question and allow the respondent to make mistakes as they may indicate problems;
- reassure the respondent that not understanding a question is perfectly ok because it is the question that is under evaluation and not he/she or the answer provided;
- remind the respondent that the overall purpose of the interview is to evaluate the questions rather than to obtain information about him/her (this is particularly useful to defuse resistance when sensitive questions are tested).

Ending the interview

Occasionally, the interviewer should discreetly check the time and let the respondent know when the interview is nearing its end. When the interview comes to an end, it is important to allow time for the interviewee to ask any questions or make comments and thank again the respondent for his/her time and contribution. It may also be important to re-affirm interview confidentiality, and, especially if sensitive issues have been addressed, give the respondent time to "come out" of the interview mode.

Usually, the participant receives an information leaflet explaining the purpose of the CT, how their data will be used, and the contact details of a reference person for any post-interview queries. In some cases, the interviewer may also provide a list of support organisations, groups or institutions the respondent can contact about issues raised during the interview that he/she may wish to follow up on.

Data Management, Analysis and Interpretation

Generally, each cognitive interview generates several types of raw data (also called primary data) that are essential for the subsequent analysis:

- socio-demographic and background information about each respondent;
- recording of interview and/or transcripts of recording;
- completed test questionnaire;
- interviewer's observations and notes (reordered in the interview protocol);
- interviewer's notes written after each interview.

If the interview is either audio or video recorded, the recording provides the most complete account, capturing everything the participant said and did during thinking aloud and probing. A decision should be made whether or not to transcribe the recording, taking into account that it is a really time-consuming activity which, however, can make the analysis easier.

If the interview is not recorded, the interviewer needs to put in extra effort by taking very detailed notes during the interview. The notes must be exhaustive regarding the respondents' answering process, any inconsistencies or lack of clarity in the answers, changes after probes, and also difficulties with answer categories or terms, confusion about the meaning of the question and so on. Since this is a very cumbersome task, it is advisable to record the interview. If that is not possible, have a checklist handy (*e.g.* skipping of questions, changing answers after probes, problems in selecting answer categories, etc.) or entrust an observer with taking notes and compiling a grid about the respondent's behaviour.

It is generally recommended to draw up a draft summary - right after each interview and before the next one - taking into account any notes in the testing protocol and other material used by the observer.

Irrespective of the approach used, verbal probing or think aloud, the main output of cognitive interviews is still verbal texts that need to be analysed to determine whether or not respondents have issues with a particular question. The analysis consists essentially in extracting information from these verbal reports about question performance:

- always provide the original answer to each test question given by the participants. Sometimes the respondents may change their initial answer offering a revised response; both have to be reported in the summary explaining how and why the original answer has been changed;
- describe both positive and negative findings on how the questions performed;
- specify the type of difficulties the respondent may have experienced in answering the question (*e.g.* the question/answer categories had to be repeated, clarifications were asked, etc.);
- report any confusion the respondents may have shown regarding the survey questions, contradictory answers to the various survey questions and inconsistencies in the respondents' narrative explaining how and why that answer has been given;
- always provide a clear distinction between participant-driven findings (think aloud, general probes) and interviewer-driven findings (specific probes);
- use examples and quotes that can better illustrate and substantiate the summary;
- make a clear distinction between quote and summary of what has been said;
- describe the findings using as much detail as possible. If any questions were not tested, this should be stated in the notes explaining why.

The next stage of data management consists in merging all interview summaries into a single-data set ready for analysis. d'Ardenne and Collins (2015) suggest using the "Framework method" that is a matrix-based approach to handle qualitative data, whereby the content of each interview is entered into a grid where each row represents a different interviewee and each column an area of investigation. Data within the grid can then be read horizontally as a complete case summary for each participant, or vertically for a systematic review of the data collected per area of investigation. Framework matrices can be set up using a non-specialist software (such as an Excel spreadsheet) or a specialist, qualitative data analysis software (such as NVivo developed by the NatCen Social Research).

A different specialist software is Q-Notes, developed by the Collaborating Centre for Question Evaluation and Design Research at the USA National Centre for Health Statistics, to help researchers summarise interview notes, develop thematic schemes, and draw conclusions about question performances (Mezetin and Massey, 2014). Q-notes allows for easy and accessible data entry by centralising the process using a uniform format to record information about all interviews and enables researchers to conduct the various tiers of analysis in a systematic, transparent fashion.

Despite wide recognition of cognitive interviewing as an established and respected pre-testing method, there is little consensus on how to handle verbal reports and carry out a systematic and transparent analysis in order to produce valuable and reliable findings (Blair and Brick, 2010); Collins, 2007 and 2014; Ridolfo and Schoua-Glusberg, 2011). Willis (2015) wrote a book specifically on analysis strategies for the Cognitive Interview and presented five different analysis models. Indeed, depending on the chosen approach the analysis can be conducted with or without a coding scheme. In particular, he called "text summary" (aggregated comments of the cognitive interviewer/s) the approach that does not use a coding scheme, which focuses on the description of dominant themes, conclusions, and problems highlighted by a set of aggregated interviewer notes.

Other models use coding schemes to reduce data in a more punctual manner than the mere description of basic findings. This coded approach can be either top-down (deductive codes), which entails the development and application of potential coding categories prior to data analysis, or bottom-up (inductive codes), which builds codes from data. In the top-down approach, analysts use two main variants according to their theoretical viewpoint: coding, which refers to the Tourangeau four-stage model, and question feature coding, which centres on the types of problems raised by the questions (*e.g.* Question Appraisal System[12]). These can also be combined. The bottom-up approach builds codes from the raw data that are mainly related to themes and patterns of interpretation. These codes are built from either quotes, or notes, or a combination of both; however, rather than achieving data reduction only through the production of a text-based summary, the analyst instead develops codes to represent key findings and applies these to each interview.

The analysis process involves two phases: description and interpretation The first one focuses on understanding how the survey question has been interpreted, responded to, and the rationale, therefore outputs of the descriptive analysis are: the identification and classification of "potential" errors, a map of the circumstances in which errors occur, and a description of who was affected. The second one investigates how "potential" problems and patterns in the problem observed have emerged, therefore main outputs of the interpretation phase are: reasons why a problem occurs and its implications for the actual survey (Collins, 2015).

Having gathered enough evidence on how a question performed, it is then possible to indicate what actions can be taken. Cognitive interview results may suggest that it would be better, in some case, for example to simplify wordings (avoid double-barrelled, long preface or eliminate examples or specification, change terms, etc.), add examples or leave open to adapt to specific countries, change answer categories or time reference, review order of questions, sections, answer categories, drop the question altogether or add questions to catch the right population, and finally change the instructions for the respondents or add instructions for interviewers. As already mentioned, if changes are made the new questionnaire should be tested again.

The analysis is an iterative process for which sufficient time must be allotted when planning the cognitive test, as materials may need to be reviewed several times for inconsistencies or missing information. In synthesis, "The general process for analysing cognitive interviewing data involves synthesis and reduction – beginning with a large amount of textual data and ending with conclusions that are meaningful and serve the ultimate purpose of the study" (Miller *et al.,* 2014) by following five incremental and iterative steps: conducting interviews, producing summaries, comparing across respondents, comparing across subgroups of respondents, and reaching conclusions.

Writing the Cognitive test report

This is the last stage in a cognitive interviewing project.

Generally, a cognitive interviewing report consists of five sections: introduction, summary of the findings, description of the methodology, detailed question-by-question review, and appendix (Cheep and Scanlon, 2014**)**. Nevertheless, these reports may vary widely as to the information provided and the level of details. Indeed, cognitive interviewing practitioners have often been criticised for neglecting key pieces of information in a report

12 See paragraph 3.3 Expert reviews in this chapter for more information.

undermining the test's credibility and validity (Miller, 2011). "How to write" a report is an important skill that only recently has been brought to the practitioners' attention. In this respect, Boeije and Willis (2013) have developed the Cognitive Interviewing Reporting Framework (CIRF) proposing to include in the report the following key aspects:

Table 6 - Cognitive Interviewing Reporting Framework (CIRF): Checklist of Major Elements to Include in a Cognitive Interviewing Report

CIRF Elements	Description
Research objectives	Clearly define the research objectives Provide a review of the relevant background literature
Research design	Describe the features of the overall research design
Ethics	Describe any review by a body responsible for ethical or administrative review of human subject research, such as an Institutional Review Board (IRB)
Participant selection	Description of the participants and of the recruitment sources and procedures. The last information helps to establish the credibility of the study by demonstrating that appropriate processes and sources were used.
Data collection	Include details relevant to interviewing procedures: number of interviews and interviewers, location of the interviews, whether they have been recorded and/or notes have been taken; interview approach and type of probes if any, number of CT rounds, whether the testing protocol was adjusted along the way.
Data analysis	Provide key details of the data analysis process such as, how it was carried out using what information, who has been involved in the process, comparison carried out during the analysis, whether software was used and whether recommendations for question modification were made.
Findings	Include a systematic summary of results. The results should be presented in sufficient detail for each item evaluated, relying as far as possible to quotes to illustrate points and evidence, to make them transparent to readers. If certain items were skipped or otherwise untested, this should be explained, rather than implying that items were evaluated but no issues observed. Indicate where there was variation in results, indicating participant subgroup membership or other important distinctions.
Conclusions, implications, and discussion	Discuss the overall impact of the cognitive test by reporting the extent to which the research objectives have been addressed and by discussing findings and solutions in light of previous evidence.
Strengths and limitations	Include and discuss strengths and weaknesses of the test design and/or the cognitive interviewing and provide appropriate caveats. Discuss potential for generalisation to the wider population from which the participant groups were drawn, and the potential for applicability to other contexts.
Report format	Use a structured format for ease of reading. One common approach consists of sequential sections, *e.g.*: Executive Summary Introduction Design and Methods Findings Conclusions and Implications, with limitations Appendix materials, including Interview Guide, probes…
Source: Boeije & Willis 2013	

It is of the essence to make the reports widely accessible to the public. This is why the U.S. National Centre for Health Statistics has developed an online database called Q-Bank where statistical agencies and research institutes may upload the results of cognitive test (Mezetin and Massey, 2014). These are available to the public using a simple query search: enter survey questions, topics, and several other keys to recall the original evaluation report.

3.4.3 Conclusion

Cognitive interviewing is a pre-test method that aims to find and correct any problems associated with a particular question by extracting information about respondents' cognitive processes. This brings to the open any problems in the question's formulation that would not surface in traditional interviewing.

Nevertheless, like other pre-testing methods, this has also drawbacks (Willis, 2005; Brancato *et al.*, 2006; d'Ardenne and Collins, 2015). For example, it is not possible to determine the frequency of the questions-related issues in the general population as the sample used is small, purposive and non-random; it is not possible to determine how long it will take to administer the questionnaire in the field as this is contingent on the

time needed to think aloud and answer probes, and because in some cases only some questions of the survey questionnaire are under test. Moreover, the test modalities affect how participants answer, for example, probing after each question or set of questions may change the way in which participants approach subsequent questions; or the ability to articulate their thoughts may vary between respondents. Finally, if the interviewers do not follow the testing protocol, *e.g.* they do not ask the survey question or the probes as worded, the interviews will be biased.

Notwithstanding these limitations, cognitive interviewing has proved to be an effective tool to pre-test questionnaires, picking up different types of problems, providing the rationale behind these problems and revealing any discrepancy between the intended aim of the question and its interpretation.

This method is also widely used to test cross-national surveys. Indeed, both individual life circumstances – *e.g.* socio-economic status, level of education, age – and country-specific socio-cultural factors determine how respondents interpret and answer survey questions. Therefore, through a comparative analysis of cognitive interviews, it is possible to identify not only patterns of interpretation and error but also "the interpretative value of a question and the way in which that question may or may not be consistent with its intent – across particular groups and in different contexts" Miller, 2014). Finally, cognitive interviewing is used in cross-national researches because it helps "uncover translation issues that might go unnoticed in other translation assessment processes such as expert reviews" (Schoua-Glusberg and Villar, 2014).

To summarise, the merits of cognitive interviewing largely outweigh its drawbacks, hence it is widely used to test not only questionnaires but any type of document related to a survey *i.e.* information leaflets, survey advance letters or consent forms. However, it is a complex method that requires a high level of professionalism in all its aspects. The testing protocol needs to be thought through in detail taking into account the testing objectives, the respondents, sample has to reflect the target population for which the questions are designed as well as those individuals, within this population, who will most likely experience or process questions in a variety of ways, the interviewers must receive high level training enabling them to draw valuable information, useful to assess question performance, to analyse them for providing recommendations for any improvement of the questions, and finally, several logistic and organisational aspects must also be considered.

This method can be used as a standalone pre-testing method or in combination with others. Indeed, applying different methods increases the likelihood not only to uncover problems in a questionnaire but also to shed light on the nature of those problems and therefore to strengthen survey data quality.

REFERENCES

Akkerboom, H., F. Dehue. 1997. "The Dutch Model of Data Collection Development for Official Surveys". *International Journal of Public Opinion Research*, Volume 9, Issue 2: 126-145.

Beatty, P. 2004. "The Dynamic of Cognitive Interviewing". In Presser, S., J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (*Eds.*). *Methods for Testing and Evaluating Survey Questionnaires*, Chapter 3: 45-66. Hoboken, NJ, U.S.: John Wiley and Sons.

Beatty, P.C., and G.B. Willis. 2007. "Research Synthesis: The Practice of Cognitive Interviewing". *Public Opinion Quarterly*, Volume 71, Issue 2: 287-311.

Biemer, P.P., and L.E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ, U.S.: John Wiley & Sons.

Blair, J., and P.D. Brick. 2010. "Methods for the Analysis of Cognitive Interviews". In *Proceedings of the Survey Research Methods Section – SRMS*, 2010 Joint Statistical Meetings - JSM: 3739-3748. Alexandria, VA, U.S.: American Statistical Association - ASA. http://www.asasrms.org/Proceedings/y2010/Files/307865_59514.pdf.

Blair, J., F. Conrad, A. Castellano Ackermann, and G. Claxton. 2006. "The Effect of Sample Size on Cognitive Interview Findings". *Proceedings of the Survey Research Methods Section – SRMS*, 2006 Joint Statistical Meetings - JSM: 4041-4046. Alexandria, VA, U.S.: American Statistical Association - ASA. http://www.asasrms.org/Proceedings/y2006/Files/JSM2006-000127.pdf.

Blake, M. 2015. "Other Pretesting Methods". In D. Collins (*Ed.*). *Cognitive Interviewing Practice*, Chapter 2: 28-56. Thousand Oaks, CA, U.S.: Sage.

Boeije, H., and G. Willis. 2013. "The Cognitive Interviewing Reporting Framework (CIRF): towards the harmonisation of cognitive testing reports". *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, Volume 9, Issue 3: 87-95.

Bradburn, N.M., and S. Sudman. 2004. "The Current Status of Questionnaire Research". In Biemer, P.P, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (*Eds.*). *Measurement Error in Surveys*, Chapter 2: 27-40. Hoboken, NJ, U.S.: John Wiley & Sons, *Series in Probability and Statistics*.

Brancato, G. *et al.,* 2006. "Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System". *European Commission Grant Agreement 200410300002*. Luxembourg: Eurostat.

Campanelli, P. 2008. "Testing Survey Questions". In de Leeuw, E.D., J. Hox, and D. Dillman (*Eds.*). *International Handbook of Survey Methodology*. Abingdon-on-Thames, UK: Routledge.

Cheep, V., and P. Scanlon. 2014. "Conveying Results". In Miller, K., V. Chepp, S. Willson, and J.-L. Padilla (*Eds.*). *Cognitive Interviewing Methodology*, Chapter 6: 69-89. Hoboken, NJ, U.S.: John Wiley & Sons, *Series in Survey Methodology*.

Collins, D. 2007. "Analysing and interpreting cognitive interview data: A qualitative approach". In *Proceedings of QUEST 2007*: 64-73. Statistics Canada, Ottawa, Ontario, Canada, 24th - 26th April 2007.

Collins, D., and M. Gray. 2015. "Sampling and recruitment." In D. Collins (*Ed.*). *Cognitive Interviewing Practice*, Chapter 4: 80-100. Thousand Oaks, CA, U.S.: Sage.

Conrad, F.G., and J. Blair. 2009. "Sources of Error in Cognitive Interviews". *Public Opinion Quarterly*, Volume 73, Issue 1: 32-55.

Conrad, F.G., and J. Blair. 2004. "Data Quality in Cognitive Interviews: The Case of Verbal Reports". In Presser, S., J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (*Eds.*). *Methods for Testing and Evaluating Survey Questionnaires*, Chapter 4: 67-87. Hoboken, NJ, U.S.: John Wiley and Sons.

Corrao, S. 2000. *Il focus group*. Milano, Italia: Franco Angeli, *Laboratorio Sociologico*.

d'Ardenne, J., and D. Collins. 2015. "Data Management". In D. Collins (*Ed.*). *Cognitive Interviewing Practice*, Chapter 7: 142-161. Thousand Oaks, CA, U.S.: Sage.

DeMaio, T.J., and J.M. Rothgeb. 1996. "Cognitive interviewing techniques: In the lab and in the field". In Schwarz, N., and S. Sudman (*Eds.*). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*: 177-195. Hoboken, NJ, U.S.: Jossey-Bass/Wiley.

DeMaio, T.J., J.M. Rothgeb, and J. Hess. 1998. "Improving survey quality through pretesting". In *Proceedings of the Survey Research Methods Section – SRMS*, 1998 Joint Statistical Meetings - JSM: 50-58. Alexandria, VA, U.S.: American Statistical Association - ASA. http://www.asasrms.org/Proceedings/y1998f.html.

Forsyth, B.H., and J.T. Lessler. 1991. "Cognitive Laboratory Methods: A Taxonomy". In Biemer, P.P, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (*Eds.*). *Measurement Error in Surveys*, Chapter 20: 393-418. Hoboken, NJ, U.S.: John Wiley & Sons, *Series in Probability and Statistics*.

Gray, M. 2015. "Conducting Cognitive Interviews". In D. Collins (*Ed.*). *Cognitive Interviewing Practice*, Chapter 6: 126-141. Thousand Oaks, CA, U.S.: Sage.

Groves, R.M, F.J. Fowler Jr., M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau (*Eds.*). 2009. *Survey Methodology, 2nd Edition*. Hoboken, NJ, U.S.: John Wiley & Sons.

Istituto Nazionale di Statistica - Istat. 1989. "Manuale di tecniche di indagine", Fascicoli da 1 a 6. *Note e Relazioni*, N. 1/1989. Roma, Italia: Istat.

Krueger, R.A. 1994. *Focus Groups: A Practical Guide for Applied Research, Second Edition*. Thousand Oaks, CA, U.S.: Sage Publications, Inc.

Lessler, J.T., and B.H. Forsyth. 1996. "A coding system for appraising questionnaires". In Schwarz, N., and S. Sudman (*Eds.*). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*: 259-291. Hoboken, NJ, U.S.: Jossey-Bass/Wiley.

Forsyth, B.H., J.T. Lessler, and M.L. Hubbard. 1992. "Cognitive Evaluation of the Questionnaire". In Turner, C.F., J.T. Lessler, and J. C. Gfroerer (*Eds.*). *Survey Measurement of Drug Use: Methodological Studies*. Washington, DC, U.S.: US Government Printing Office.

Merton, R.K., and P.L. Kendall. 1946. "The Focused Interview". *American Journal of Sociology*, Volume 51, Issue 6: 541-557.

Mezetin, J., and M. Massey. 2014. "Analysis Software for Cognitive Interviewing Studies: Q-Notes and Q-Bank". In Miller, K., V. Chepp, S. Willson, and J.-L. Padilla (*Eds.*). *Cognitive Interviewing Methodology*, Chapter 8: 107-132. Hoboken, NJ, U.S.: John Wiley & Sons, *Series in Survey Methodology*.

Miller, K. 2014. "Introduction". In Miller, K., V. Chepp, S. Willson, and J.-L. Padilla (*Eds.*). *Cognitive Interviewing Methodology*, Chapter 1: 1-6. Hoboken, NJ, U.S.: John Wiley & Sons, *Series in Survey Methodology*.

Miller, K. 2011. "Cognitive Interviewing". In Madans, J., K. Miller, A. Maitland, and G. Willis (*Eds.*). *Question Evaluation Methods. Contributing to the Science of Data Quality*, Chapter 5: 49-75. Hoboken, NJ, U.S.: John Wiley & Sons.

Miller, K. 2002. "The Role of Social Location in Question Response: Rural Poor Experience Answering General Health Questions". Paper presented at the *American Association for Public Opinion Research Conference*, St. Pete Beach, Florida, U.S., May 2002.

Miller, K., S. Willson, V. Chepp, and J.M. Rayn. 2014. "Analysis". In Miller, K., V. Chepp, S. Willson, and J.-L. Padilla (*Eds.*). *Cognitive Interviewing Methodology*, Chapter 4: 35-50. Hoboken, NJ, U.S.: John Wiley & Sons, *Series in Survey Methodology*.

Morgan, D.L. 1997. *Focus Group as Qualitative Research, Second Edition*. Thousand Oaks, CA, U.S.: Sage Publications, Inc.

Morgan, D.L. 1988. *Focus Group as Qualitative Research*. Thousand Oaks, CA, U.S.: Sage Publications, Inc.

Oksenberg, L., C. Cannell, and G. Kalton. 1991. "New Strategies for Pretesting Survey Questions". *Journal of Official Statistics - JOS*, Volume 7, Issue 3: 349-365.

Olson, K. 2010. "An Examination of Questionnaire Evaluation by Expert Reviewers". *Field Methods*, Volume 22, Issue 4: 295-318.

Presser, S., and J. Blair. 1994. "Survey Pretesting: Do Different Methods Produce Different Results?". *Sociological Methodology*, Volume 24: 73-104.

Ridolfo, H., and A. Schoua-Glusberg. 2011. "Analysing Cognitive Interview Data Using the Constant Comparative Method of Analysis to Understand Cross-Cultural Patterns in Survey Data". *Field Methods*, Volume 23, Issue 4: 420-438.

Rothgeb, J., G. Willis, and B. Forsyth. 2001. "Questionnaire Pretesting Methods: Do Different Techniques and Different Organisations Produce Similar Results?". In *Proceedings of the Survey Research Methods Section – SRMS*, 2001 Joint Statistical Meetings - JSM. Alexandria, VA, U.S.: American Statistical Association - ASA. http://www.asasrms.org/Proceedings/y2001/Proceed/00476.pdf.

Schoua-Glusberg, A., and A. Villar. 2014. "Assessing Translated Questions via Cognitive Interviewing". In Miller, K., V. Chepp, S. Willson, and J.-L. Padilla (*Eds.*). *Cognitive Interviewing Methodology*, Chapter 5: 51-68. Hoboken, NJ, U.S.: John Wiley & Sons, *Series in Survey Methodology*.

Schwarz, N. 2007. "Cognitive Aspects of Survey Methodology". *Applied Cognitive Psychology*, Volume 21, Issue 2: 277-287.

Snijkers, G.J.M.E. 2002. "Cognitive Laboratory Experiences: On Pre-testing Computerised Questionnaires and Data Quality". *Ph.D. Thesis*. Utrecht and Heerlen, The Netherlands: University of Utrecht, and Statistics Netherlands - CBS.

Statistics Canada, Methodology Branch. 2003. "Quality Guidelines, Fourth Edition". *Catalogue N. 12-539-XIE*. Ottawa, Ontario, Canada: Statistics Canada. https://www150.statcan.gc.ca/n1/en/pub/12-539-x/12-539-x2003001-eng.pdf?st=FBdw6uTL.

Statistics Canada. 1998. "Quality Guidelines, First Edition". *Surveys and statistical programs – Documentation*. Ottawa, Ontario, Canada: Statistics Canada.

Stewart, D.W., and P.N. Shamdasani. 1990. *Focus Groups. Theory and practice*. Thousand Oaks, CA, U.S.: Sage Publications, Ltd.

Tourangeau, R. 1984. "Cognitive science and survey methods: A cognitive perspective". In Jabine, T.B., M.L. Straf, J.N. Tanur, and R. Tourangeau (*Eds.*). *Cognitive Aspects of Survey Methodology. Building a Bridge Between Disciplines. Report on the Advanced Research Seminar on Cognitive Aspects of Survey Methodology*: 73-100. Washington, DC, U.S.: National Academy Press.

Waltermaurer, E. 2012. "Public Justification of Intimate Partner Violence: A Review of the Literature". *Trauma, Violence, & Abuse*, Volume 13, Issue 3: 167-175.

Willis, G.B. 2015. *Analysis of the Cognitive Interview in Questionnaire Design. Understanding Qualitative Research*. Oxford, UK: Oxford University Press.

Willis, G.B. 2005. *Cognitive Interviewing. A Tool for Improving Questionnaire Design*. Thousand Oaks, CA, U.S.: Sage Publications, Inc.

Willis, G.B. 2004. "Cognitive Interviewing Revisited: A Useful Technique, in Theory?". In Presser, S., J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (*Eds.*). *Methods for Testing and Evaluating Survey Questionnaires*, Chapter 2: 23-43. Hoboken, NJ, U.S.: John Wiley and Sons.

Willis, G.B. 1999. "Cognitive Interviewing: A "How To" Guide". *Guide*. Rockville, MD, U.S.: Research Triangle Institute. https://www.hkr.se/contentassets/9ed7b1b3997e4bf4baa8d4eceed5cd87/gordonwillis.pdf.

Willis, G.B. 1994. “Cognitive Interviewing and Questionnaire Design: A Training Manual”. *Working paper series*, *Issue 7*. Hyattsville, MD, U.S.: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Willis, G.B., and J. Lessler. 1999. “Questionnaire Appraisal System QAS 99”. *Manual*. Rockville, MD, U.S.: Research Triangle Institute.

Willis, G.B., S. Schechter, and K. Whitaker. 1999. “A Comparison of Cognitive Interviewing, Expert Review, and Behavior Coding: What Do They Tell Us?” In *Proceedings of the Survey Research Methods Section – SRMS*, 1999 Joint Statistical Meetings - JSM: 28-37. Alexandria, VA, U.S.: American Statistical Association - ASA. http://www.asasrms.org/Proceedings/y1999f.html.

Willson, S., and K. Miller. 2014. “Data Collection”. In Miller, K., V. Chepp, S. Willson, and J.-L. Padilla (*Eds.*). *Cognitive Interviewing Methodology*, Chapter 3: 15-34. Hoboken, NJ, U.S.: John Wiley & Sons, *Series in Survey Methodology*.

4. SAMPLING ISSUES[1]

4.1 - Sampling issues for a victimisation survey about violence

4.1.1 Target population

When planning a population survey, the first step to be taken is defining the target population, that is, the set of units to be investigated, about which information is sought and estimates are required. This definition has to be made with respect to:

- type of units: the population units can be households, persons or both, meaning that the survey is supposed to produce estimates about individuals as well as about households (not for the survey object of this note);
- territory: people resident in the country, or having the nationality where the survey is carried out;
- time: a reference period of the survey has to be stated, which can be the precise moment of time to which the population frame or list is referenced;
- age, sex or other characteristic restrictions;
- inclusion of persons living in institutions, such as hospitals, prisons, barracks (in theory it is usually better that the latter are included, but in many situations, this is not an easy task to deal with).

The VAW and GBV surveys generally consider a target population of individuals in a specific age range and, for some countries, also a specific gender. Defining the target population is fundamental both for dedicated surveys and for the module on VAW/GBV included in some other surveys.

When a specific survey is planned, the sample design can take into account the peculiarities of the population and of the subject investigated. Otherwise, the specific module will be administered only to the subset of individuals belonging to the target population, selected from the general sample of the larger survey.

4.1.2 Sampling frame and coverage

The sampling frame is the list of the target population units from which the sample is drawn. It is any material or device used to obtain observational access to the finite population of interest (Sarndal, 1992). It can be any list that delimits, identifies and allows access to the element of the target population. The frame should contain all the units belonging to the population, reporting information about address and localisation, together with other variables (age, gender, etc.) useful for the designing of the sample. In some cases, a unique and complete list may not exist or can be built only by means of laborious work.

1 This chapter has been drawn up by Claudia De Vitiis.

The characteristics of the sampling frame are a crucial issue. It is important to consider the type of sampling units listed in the frame, the extent of coverage of the target population, the accuracy and completeness as well as the amount and quality of the auxiliary information. The extent to which a frame includes all the elements of the target population is referred to as coverage.

Loss of coverage can be due both to under-coverage, that is failure to include some units, or over-coverage, that is including some extraneous elements into the sampling frame.

Under-coverage is different from deliberate exclusion of sections of the target population. Usually, survey objectives and practical difficulties determine such deliberate exclusions. For example, in a national survey, certain segments may be excluded because the survey objectives are confined to the other segments of the target population (*e.g.* institutional households, etc.). Such exclusions are not errors of non-coverage, it has to be emphasised in the survey report that the results apply only to the groups of the target population included in the sample.

Over-coverage is when some selected units do not belong to the population. Some units may appear in the frame more than once, giving them a chance of selection into the sample that is larger than intended. But if a person is appearing more than once in the frame, it is not necessarily over-coverage if the person belongs to the target population. It is just the nature of that particular sampling frame, but this needs to be known and taken into account when selecting the sample by using different sampling probabilities for different units.

A good sampling frame should also provide sufficient information on the basis of which the selected units can be contacted and uniquely identified on the territory (precise and up-to-date address). Failure to do so can result in bias in the probabilities of selection and in the sample structure, because some units cannot be contacted and their inclusion probability becomes equal to zero.

Ideally, the sampling frame should be complete, unduplicated and unambiguous. Therefore, initial preparation of the frame may be required to ensure that the ideal conditions are met to the extent that is feasible. A relevant issue is the evaluation of the coverage level of the frame with respect to the target population, that is, to estimate the percentage of excluded population members, which gives rise to under-coverage.

Types of sampling frames

For violence surveys the most complete register including the units defined as target population can be chosen among population registers, electoral lists and census lists.

In some countries there are population registers containing the names and addresses of all the persons in the population, as well as additional auxiliary information, such as age, sex, education, and so forth. In such situations, drawing a representative sample is a particularly straightforward issue.

Another example of a list used for sampling is the telephone directory. This list is not always complete because of unlisted numbers or households without a telephone (and because of the increasing number of mobile phones often not included in directories), it can also include some households more than once (households with more than one number).

When one of these two kinds of lists of individuals is to be used for the sample, a one-stage sample design can be drawn up, in which each final sampling unit is directly selected from the list. Of course, even when a population register is available, a two-stage sampling design can also be defined for data collection implications.

In many countries, however, a complete list of all the persons in the target population may not exist. In this case an area frame can be defined as a geographic frame consisting of an area unit; every population element belongs to an area unit and can be identified after inspection of the area unit (Sarndal, 1992). In these situations, the total number of population units is often unknown. The sets of elements drawn up using an area frame are usually called clusters. The selected clusters can be sub-sampled in a secondary selection phase.

One relevant issue, also for the surveys on violence, is the inclusion or exclusion from the target population of individuals living in institutions (hospitals, prisons, religious institutions, etc.): these types of units are generally not listed in household frames and therefore difficult to contact.

4.1.3 Planning the sampling design

In order to define the sampling strategy, it is essential to precisely define the main parameters of the required estimates. The parameters are defined with regards to the variables to be collected with the questionnaire: for categorical variables, the parameters are absolute or relative frequencies, for continuous variables, the parameters usually are means or totals. In the context of surveys on violence, the main parameters are relative frequencies or proportions, and average numbers of events.

The domain of estimates are the levels for which estimates are to be produced, such as regions or sex and age groups. The domains are considered planned in the sampling design if they can be obtained by means of aggregation of the sampling strata, unplanned if they are not (or cannot be) taken into account when designing the sample.

The phase of planning the sample design has to take into account what has been previously established in terms of objectives and constraints of the survey and it consists of defining:

– the sampling scheme, characterised by:
  - selection stages;
  - stratification criteria: stratification variables, number of strata;
  - unit selection probabilistic method: equal or unequal probabilities;
– total sample size and its allocation among strata.

The use of probability samples is usually recommended. In a probability sample each person in the target population has a known (calculable) non-zero chance of being included in the sample. It does not necessarily require that every person have the same probability of being selected, but simply that each person has a chance of being included. The probability sampling has to be defined at each sampling stage.

Only probability sample designs are based on accepted sampling distribution theory, allowing the estimates derived from the survey data to be legitimately generalised to the population from which the sample is drawn, and also permitting the estimation of measures of precision of the survey estimates.

Selection stages

The first step of a sampling design is to decide the number of sampling stages, which usually is defined by the availability of sampling frames, the type of interview and the possible need for concentrating the fieldwork geographically.

A one-stage sample design can be chosen only if a list of the population units is available, so that each unit can be directly selected from the list. In this case, the final

sampling units (individuals or households) can be drawn from the list by simple random sample or, as is preferable, by a stratified random sample, if the frame contains some variables (such as gender, age, or other) that can be used to stratify the population units. The selection process can be either via a simple random sampling or a systematic sampling within each stratum. This type of design is also possible when the chosen data collection technique does not foresee interviews by the house of respondents. This is the case of web, mail or Cati techniques.

When the survey involves an in-person interview (Papi or Capi data collection techniques), multistage cluster sampling is a common way of designing a sample that requires data collection on the field. The cost of data collection will be greatly reduced using a design that involves geographic clustering.

A two-stage sample design is, moreover, always needed when a list of all population units does not exist and it is necessary to build the selection list involving local authorities (for instance population registers in each municipality). In this situation, a sample of territorial units (census areas, municipalities, postal addresses) - named primary sampling units, PSUs - is selected at the first stage and the list of households or individuals is built for each of them. From these lists, the second stage sample (secondary sampling units, SSUs) is drawn. For the selection of PSUs at the first stage, a probability proportional to size scheme is typically used (Cochran, 1977). The most common measure of a PSU's size is the population.

The choice of using a two or more stage sample design produces, however, an increase in the variability of the estimates, given the same sample size; the higher the intra-cluster correlation coefficient (homogeneity in the target variables within the PSUs) and the number of SSUs for PSU, the higher the consequent increase in variability. This effect is known as cluster effect.

It is true, nonetheless, that when comparing the efficiency of a one-stage and a two-stage sample design, the latter is less efficient given the same sample size but more efficient given the same cost, because it allows stricter control on the data collection phase.

In general, PSUs have to be large enough to reduce clustering effects and small enough to be appropriate for interviewers to travel within.

Stratification

The stratification process involves dividing the population into groups (strata) before the selection of a sample within each of these subsets. The aim is to gain precision by creating strata that are internally homogeneous with respect to the characteristic to be estimated. The stratification is usually defined in such a way that the planned domains for the estimates are obtainable by aggregating strata.

The information needed for the stratification on the selection frame has to be available in advance; therefore, when an individual register is available for the survey it will be possible to define a stratification relative to an individual characteristic, otherwise, only geographic variables, attached to the territorial units, can be used.

For the survey on violence, the population units could be stratified by territorial areas and by age groups, provided that a list of individuals is available.

Selection of individuals in selected households

When a sample of individuals is the objective, but households are the sampling units listed in the frame, the issue arises in regard to which method to use for the selection of the individuals for the final sample. One possibility is retaining all eligible individuals in a selected household in the sample, thus introducing, at any rate, a cluster effect due to the similarity of the individuals belonging to the same family and, possibly, some data collection drawbacks. Otherwise, a method for selecting individuals has to be defined.

The chosen within-household sampling rule must be random and probabilistic, and must avoid the introduction of bias in total coverage of the target population. One method that can be used for selecting individuals within a household is the Kish grip method (Kish,1949), consisting in selecting one individual randomly from the list of eligible household components by means of a file of random numbers associated to the sample households.

Another method for selecting one individual in the sample household is the last-birthday (LB) method: after having listed all household members eligible to be interviewed, the one who last had a birthday (or who will have one next) is chosen for the interview.

What is important when selecting individuals within households is keeping track of how many eligible individuals belong to the household, so it is possible to correctly calculate the inclusion probability.

4.1.4 Sample size

In order to fix the sample size and the allocation among domains and strata, the precision required for the estimates has to be stated. There exists, in fact, an inverse relation between sample size and estimate precision, which depends on the type of sample design as well. So, the larger the sample size, the higher the precision (the lower the sampling error) of the estimates. When determining the sample size, it is also necessary to take into account the type of sample design to be used.

To express the precision of a survey, estimating the coefficient of variation (cv) is the measure commonly used. It is the ratio of the standard error of the estimate and the population value of the parameter. For categorical variables it can be expressed with the formula:

$$cv = \sqrt{deff \frac{(1-p)}{n_0 \cdot p}}$$

where n_0 is the sample size, p is the proportion of the population belonging to the category of interest (for example, the relative frequency of people with a chronic disease), *deff* is the design effect (that is, the relative efficiency of the sample design used as compared to a simple random sample with the same size); *deff* is usually greater than unity for cluster samples.

Solving for n_0, we obtain:

$$n_0 = deff \frac{(1-p)}{cv^2 \cdot p} \qquad (*)$$

In order to determine the required sample size, one can specify the desired level of reliability to be associated to the main survey estimates, and it is evident that the smaller p is, the larger the sample size must be in order to attain a specified level of precision. Therefore, the measurement of rare characteristics or estimates of very fine breakdowns of the population require larger sample sizes.

To give an example of sample size definition, let us consider two cases:

A. category with p=10% and a desired precision level fixed at cv =15% at maximum, which produces the 95% confidence interval (7.1%-12.9%);
B. category with p=5% and a desired precision level fixed at cv=15% at maximum, which produces the 95% confidence interval (3.5%-6.5%).

The design effect used to take into account for the use of a complex multi-stage sample design can be approximately fixed equal to 2. In this situation, according to the (*), the sample size n0 results as equal to 800 for case A) and 1,698 for case B).

It is useful to add that when a one-stage sample scheme is used, the design effect is usually much smaller, as it can be considered approximately equal to 1. In this situation, the sample size required to obtain the same desired precision as in the above example will be cut by half.

The example highlights the crucial role of the choice of the guide variable or the guide parameter to be estimated. Although a survey is usually intended to produce a high number of estimates, it is always possible to state which ones are the most relevant.

It is important to notice that the reliability of an estimate, all things being equal, depends almost exclusively on the sample size, while the influence of the population size is very weak. Thus, estimates of comparable reliability for all required domains (territorial domains or sex or age groups, for instance) can be achieved by allocating the same sample size to all domains (equal allocation). On the other hand, the reliability of national estimates is maximised by allocating a sample to each sub-domain that is proportional to its population (proportional allocation).

Therefore, as reliable national estimates can only be achieved at the expenses of the reliability of estimates for sub-national domains, surveys often define a compromise by allocating samples to sub-national domains that are in between equal and proportional allocation.

When the estimate domains are unplanned domains and their definition is not obtainable in the sample planning phase because the needed information is not available in the frame (for example, sex and age groups when a population register is not available), it is not possible to define the sample size of these domains and it is necessary to evaluate their size in terms of expected size.

In this case, it can be assumed that a random non-stratified selection will produce a distribution of the sampling units with respect to the interest variables (sex and age groups in the example) similar to the distribution in the population. In this way, it is possible to calculate the expected sample size for the estimate domains by applying this distribution to the entire sample size, or by fixing the minimum required sample size for these domains and calculating the overall size on the basis of this minimum. For example, if it is known that the smallest sex and age group represents 5% of the population, the overall sample size has to be fixed as twenty times (20=1/0.05) the minimum sample size for this domain. If this minimum equals 500, the overall sample size has to be 10,000.

The sample size discussed so far is the effective number of respondents to obtain in order to produce estimates with a desired level of precision. As in every survey, non-responses always exists, it can be necessary to calculate the actual number of population units, higher than the effective one, to be selected in order to take into account an expected non-response rate.

4.1.5 Special features for sampling in mixed-mode surveys

For surveys on violence, as for other social surveys, a mix of data collection modes can be adopted[2]. The use of mixed-mode surveys, in fact, is spreading both to contrast declining response rates and to reduce the total cost of surveys (de Leeuw, 2005).

There are two main types of designs of mixed-mode surveys:

– survey design with "concurrent mixed mode" in which one mode is used to reach a part of the sample and a different mode is used to reach another part of the sample;
– survey design with "sequential mixed mode" in which the same mode is used for the entire sample in the first phase of the survey, followed by a different mode on those units who have not responded in the first phase.

Survey designs with concurrent mixed modes are generally used for overcoming coverage problems related to a single mode (*e.g.* in order to tackle the problems of coverage linked to Cati, they can be accompanied by a face-to-face mode). Therefore, the concurrent mixed mode can be used to increase the coverage (*i.e.* cover segments of the population who have different characteristics) although this does not necessarily entail an increase in the response rates.

The survey designs with sequential modes are generally used in order to overcome total response problems together with coverage issues of the single modes. In these cases, one mode is used (generally less expensive) as the main mode for the entire sample, followed by another mode (usually more expensive) for which a higher response rate is expected. One first uses, for example, a web mode, followed by an interview with a data collector (Capi/Papi) to those not responding by a certain date. With this type of design, in addition to increasing response rates, one can also try to keep the surveying costs reasonable.

In the choice of the design type (concurrent or sequential) there is also the influence of the survey's management aspects, which may be more or less complex, depending on the cases. Both designs, concurrent and sequential, may induce a mode effect, in particular as regards the measurement effect (differences in measurement errors due to the different mode used). The use of sequential modes can induce a selection effect deriving from the different representativeness of the two sub-samples. Furthermore, the measurement effect can be confused with the selection effect (de Leeuw, 2005). For surveys that use different modes in a concurrent way, with the assigning of the mode according to the characteristics of contact of the units of the population, it is necessary to evaluate, together with the measurement error, also the selection effect caused by the dissimilar distribution of the identified sub-populations.

In the design of a mixed-mode survey, the definition of the sample design is closely connected to the choice of modes and must take into account mainly the following two aspects:

– the type of selection list, which can be single or separated by mode and may contain elementary units or clusters of units;
– the procedure of the mix of modes, in particular, if it is concurrent or sequential.

Relatively to the selection list, a single frame containing units of the population is generally available, allowing the possibility of contacting these subjects through the different modes. In some cases, a list for each mode may be available, for example, the telephone list for Cati interviews or lists of email addresses for web surveys. In this latter situation,

2 See Chapter 5 "Modes of interviewing in victimisation surveys about violence: review and suggestions".

separate sample designs can be defined that can be related to the populations that do not overlap or those that partially overlap, a circumstance which must be managed by following appropriate sampling and estimation methodologies (multi-frame approaches).

Even if a single frame is available, independent samples can be defined with reference to each of the sub-populations that are reachable with the different surveying modes. In particular, in cases where it has been decided to use the different modes concurrently on different sub-populations, for example, using Cati for the units with a fixed telephone line and using other modes for units without one, it is possible to define independent samples based on different sample schemes.

If, instead, the modes are administered sequentially, they must necessarily be defined by a single sample selected from the general population and therefore, the two sub-samples which respond to the different modes are not independent.

In the less frequent case in which the modes are proposed simultaneously to the respondents, leaving them the choice of which mode to use, the sample will be necessarily unique and the consequent breakdown in sub-samples related to the different modes will be the result of a mechanism dependent on the respondents themselves.

In conclusion, the relationship between the sub-samples relating to different modes is determined by the method of the mix of modes.

As regards the definition of the sample design/s, each of them follows the same criteria used for the mono-mode surveys. The sample design can, therefore, involve one or more stages, with stratification of the units (first and/or second stage) on the basis of the information contained in the archive.

Generally, when the face-to-face interview is planned (Papi or Capi mode) a multi-stage design is used, in which the territorial areas are the units of the first stage, requiring the concentration of the sample on the territory; in the case of sequential designs, this implies that if at the end of the sequence of modes such a contact mode is provided, it will be necessary to define a multi-stage design.

If the direct interview is not foreseen (*i.e.* in cases where the modes in the field are only Cati and/or Cawi), it is possible and preferable to define a design in a single selection stage, generally with stratification, that has known advantages relating to the efficiency of the estimates.

Regarding the definition of the sample size, when for a concurrent mixed-mode survey different sample designs are defined on the sub-populations identified by the different modes, it is necessary to take into consideration the fact that the two sub-samples should be used jointly for the production of the estimates of the survey. Accordingly, the definition of the overall sample size and sizes of the samples to assign to the sub-sample of each mode, must follow the usual criteria of efficiency of the estimates as if it were a single sample. This means, in practice, that it is appropriate that the sample size assigned to each mode is proportional to the weight of the corresponding sub-population in the overall population and that the allocation of the sample in the strata is consistent with this principle. In addition, the strategies to be implemented to obtain the expected sample sizes (over-sampling or substitutions) can be defined and managed independently for the two samples.

For the sequential multi-mode, instead, the sample size is defined following the usual criteria based on the efficiency of the estimates in compliance with the cost constraints. Particular attention should be devoted to the definition of over-sampling necessary to obtain the expected sample size, which is defined on the basis of the expected response rates of the different modes used in sequence.

REFERENCES

Cochran, W.G. 1977. *Sampling Techniques, 3rd Edition*. Hoboken, NJ, U.S.: John Wiley & Sons, *Series in Probability and Statistics*.

de Leeuw, E.D. 2005. "To Mix or Not to Mix Data Collection Modes in Surveys". *Journal of Official Statistics - JOS*, Volume 21, Issue 2: 233-255.

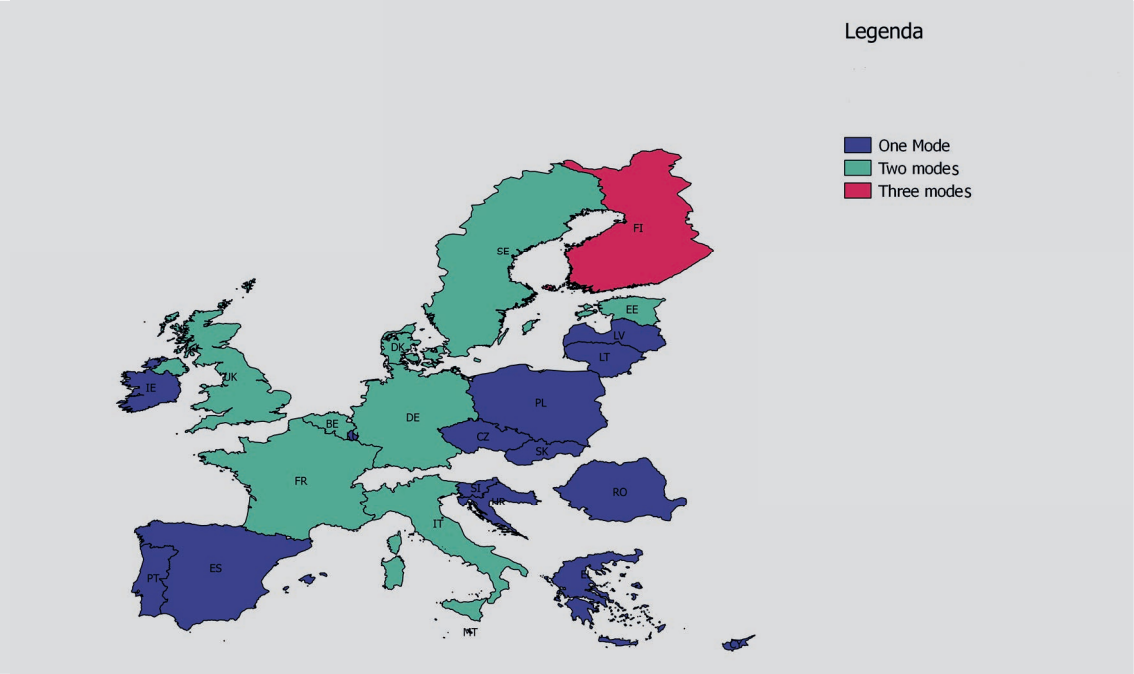
Kish, L. 1949. "A Procedure for Objective Respondent Selection within the Household". *Journal of the American Statistical Association*, Volume 44, Issue 247: 380-387.

Särndal, C.-E., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. Berlin, Germany: Springer Nature, *Springer series in statistics*.

5. MODES OF INTERVIEWING IN VICTIMISATION SURVEYS ABOUT VIOLENCE: REVIEW AND SUGGESTIONS[1]

In 2016, we counted, through a provisional census, around 40 surveys carried out in Europe, once or periodically, from the oldest ones - usually modules inserted in larger surveys - to the most recent ones[2]. According to this review, about 13 surveys (representing almost one-third of the total) used the multi-mode approach for data collection (Figure 5.1).

Figure 5.1 - Geography of data collection in the VAW Survey



Source: Our own Census

That means that more than one mode was used in order to finalise the interview. In detail: in one case, Finland, three modes of data collection were employed (Capi, Cati and Cawi); in seven cases, the multi-mode approach turned out to be a Capi interview, which switched to a Casi mode; in the remaining cases the chosen options were a mix of Capi and Cati or of different kinds of self-administered questionnaires.

A rapid glimpse of these experiences makes it easy to understand how important a full comprehension of the effects of each data collection mode has become. Not only any single interviewing mode has strong and weak points, but more and more surveys are carried out with more than one data collection mode, so that researchers must take into account the intersection of the potential distortion characteristics of any interviewing mode.

1 This chapter has been drawn up by Alberto Violante; the box “The interviewer effect in the Violence against Women survey” has been drawn up by Gabriele Ascari; the subparagraph “An example of a multivariate strategy of analysis for controlling the mode effect“ has been drawn up by Isabella Corazziari.

2 The review was made mainly online and asking for information from our contacts in the professional community engaged in the VAW research.

The aim of this paper is to make a summary of the nature of these kinds of errors. First, we will briefly describe the most frequently used options on the interviewing modes. In a second step, we will proceed in borrowing from the existing literature the most proper definition of a mode effect. Then we will try to describe how mode effects entail problems in the production of prevalence rates in surveys on violence against women. Lastly, we will offer an overview of the ways in which the evaluation of bias derived from the mode effect has been approached, and we will propose an example based upon an analysis of data from the Italian Citizen's Safety Survey carried out in 2015-16.

5.1 Interviewing Modes

A survey is usually carried out with a questionnaire composed of fixed choice questions and a limited number of open questions. Decades ago, the face-to-face interview was considered the superior method of carrying out an interview (de Leeuw, 1992). After the penetration of telephone spread throughout the western world, the choice of which mode to use was no longer so trivial, because the telephone interview was discovered as carrying a number of advantages. Use of the telephone is less expensive and more immediate, favours anonymity and is more suited to the needs of fast marketing surveys or election polls. A number of issues were raised concerning the quality of the interviews, since the telephone interview started to replace the personal interview.

How is it possible to distinguish different interview modes? The discriminant point between different interview modes, beyond the medium channel, is the presence or the lack of an interviewer. This discriminant point entails a sort of script and, according to social psychologists, somehow directs the behaviour of the respondent, who more or less follows the script. In the case of the Cati mode, for example, to answer a phone call received by an unknown person is regarded as a kind of courtesy, the interviewer knows this and tries to hurry up, in order to avoid taking too much of the respondent's time. In the case of the Capi mode, the interviewer is the "guest" and must be received with a feeling of hospitality. In the first case, there is the risk of having a low-quality interview, since the respondent is engaged in other activities and is answering absent-mindedly, in the second case, there is the risk of the interviewee getting embarrassed in front of a stranger.

Table 5.1 - Pros and Cons of different interviewing modes

Interviewing mode	Response rate	Population to be reached	Economic cost	Role of the interviewer	Implementing electronic questionnaire
Online Interview (Web Survey)	Lower	Easy Internet access with a certain degree of ICT literacy	Lowest	None	Yes
Postal Survey	Lower	Any sort of population; written language literacy	Low	None	No
Telephone Survey (Cati)	Intermediate	Population phone-connected	High	Yes	Yes
Paper personal Interview	Usually the highest	Any sort of population	Very high	Yes	No
Computer-assisted personal interview	Usually the highest	Any sort of population	Very high	Yes	Yes
Computer-assisted self-administered interview	Usually the highest	Any sort of population with a certain degree of ICT literacy	Very high	None	-

Source: De Leeuw 2005

The best mode depends on the tasks that the survey has to accomplish. The Handbook on victimisation survey (UNODC, UNECE, 2010), even if noting that more than half of the victimisation surveys have been carried out with face-to-face interviews, gives some criteria for choosing the most useful mode of interviewing:

- the kind of available frame which will be sampled (address or telephone list);
- the available resources;
- the type of population which is sampled.

Van Kesteren *et al*. (2009) suggest that only a mail survey has to be left out from the different interview modes, because experiences led to a very low response rate, to the point that estimates would not be usable. Current literature (which is more and more concentrated on Internet rather than on the mail survey) does not completely agree. Even if the response rate in a survey without any interviewer is usually lower, that does not mean it cannot be improved with acts of reinforcement. Effort in sustaining the propensity to answer has itself a differentiated outcome, but a number of tools (toll-free phone numbers for information, advance letters, etc.) can increase meaningfully the rate of participation in the survey.

5.1.1 Different available modes

We can notice that the choice among different modes of data collection should not be interpreted as the classic division between face-to-face and telephonic interviews, but a broader range of options has been given by the adoption of computerised solutions. If - as we mentioned - the medium is not enough to establish a clear-cut conceptual distinction between different sorts of interviews, it is also true that the mix of the medium and the presence of an interviewer draws diverse kinds of interview modes as commonly intended. In the previous paragraph, we spoke mainly about the interviewer, but the medium channel is also important. The dimensions of the presence of an interviewer (which can be intended also as the degree of proximity of the interviewer - if there is one - to the interviewed) and the presence of a computer can be crossed in order to classify different interviewing modes. Smith and Kim (2015) make a synthesis of this element in Table 5.2.

Table 5.2 - Data collection mode

	No Computer		Computer	
	Interviewer	Self-Administered	Interviewer	Self-Administered
In Person	Papi	Papi + Self-administered questionnaire	Capi	CASI/CSAQ ACASI VCASI
Telephone	Telephone Interview	-	Cati	IVR
Mail	-	Mail Questionnaire	-	Mailed Disk
Internet	-	-	-	Online Questionnaire

Source: Smith and Kim, 2015

The table shows the various uses of the Information and Communication Technologies (ICT) tools. A laptop support can turn out to be a Capi, a Casi, an ACASI and a Vcasi interview. Computerisation of the survey does not mean that any customisation of data collection is prohibited. For example, as we mentioned, there are already experiences in which the same electronic questionnaire has been used as a Capi and a Casi interview. Most of the multi-mode approaches outlined on the map (Figure 5.1) above in fact contains

self-administered questionnaires provided by ICT tools. Contemporary instruments allow a flexible use of the electronic questionnaire, trying also to include the use of visual cards in the self-administered interview (the absence of this was usually considered the weakest point of the self-administered interviews). The crucial choice is now whether or not an interviewer should be present, and at which degree of proximity. The approach of choosing not exclusively one mode of interviewing is increasingly becoming mainstream. If a certain interview mode is the best for reaching one specific kind of population, perhaps by using more than one mode of data collection it is possible to arrive at more than one subgroup in the population.

Far from being a peculiar problem concerning victimisation surveys of violence against women, this is an issue, which has become common in the survey research field. Cost problems and coverage issues more frequently drive National statistical offices and Institutes of research to combine different data collection modes. If this practice is already accepted (although not dominant), the community of Survey researchers and professionals necessarily has to establish a way to gauge risks and opportunities regarding the overall quality of the survey and to quantify the non-sampling errors derived from a multi-mode approach. In order to attempt achieving this task, an International project[3] led by the German and the Dutch National Statistical Offices has been completed. Many contributions have also been given in the context of the scientific debate on the mixed-mode approach to research (Dillman *et al*., 2009).

It could be said that while the use of the multiple mode of data collection is an answer to the general decrease of the response rate and an attempt to reduce the bias in the least expensive way possible, the use of multiple modes itself causes non-sampling errors, due to the mode effect.

Definition of data collection mode and its effect

There are two ways to define the collection mode and its effect on the survey estimate. One way is reducing it strictly to the mode as a source of bias deviating from the "true" estimate (De Vitiis *et al.,* 2017). The other way is looking at the choice of the medium to be used for collecting the data as a momentum embedded in a process of getting closer to interviewed people (Biemer and Lyberg, 2003). The latter position distinguishes three dimensions:

1. the degree of contact with the respondent;
2. the degree of interviewer involvement;
3. the presence or the lack of a computer during the interview.

Of course, if we consider the definition of data collection as a process, also a more nuanced view of the possibility for controlling every consequence in terms of bias has to be assumed. Since a real process depends on the concrete place where it happens, in the same way, an identical method used to collect the information can produce very different bias depending on the issues treated in the questionnaire, the social environment in which the interview takes place and the type of subpopulation interviewed. Korner (2014) notes:

3 The project tried to understand the impact of the mixed mode of data collection especially on the Labour Force Survey and was carried out for the European Statistical System under the coordination of Destatis (the German National Institute of Statistics) and Statistics Netherland (The Dutch National Agency for Statistics).

> "A whole set of interacting factors need to be considered, including the type of social interaction, the type of communication, the technology used, but also features of the reporting situation. Any analysis of the impact of the data collection mode on statistical measurement needs to be aware of these facets of the phenomenon. The communication channel itself can probably be described as an important underlying factor that does however not allow predicting the effect on survey measurement. The complexity of the interplay of these factors at the same time makes it very difficult to generalise findings established for one data collection mode applied in one survey to another survey (apparently) using the same data collection mode". (Korner, 2014: p.5).

This long quotation fits quite well the peculiarities of a victimisation survey about gender-based violence, because questionnaires usually concern the attempt to discover dark figures through declarations of a very peculiar population. In other words, not only does the collection medium have to be considered, but also the entire design of the research is crucial and it should be understood how it influences the collection phase. It is implicit that if we take the position of considering data collection as a process finalised at reaching the selected interview, we must assume that the beginning of this process starts somehow with sampling design and that there is not a "one size fits all" choice. This statement proposes two problems:

- the choice of the data collection mode has to be considered right from the moment of the design of the research;
- there are many principles that could be used for addressing the choice.

The choice of a data collection mode

The use of more than one mode of data collection is usually chosen in order to cut costs and increase the quality of coverage and the response rate. In both cases, the use of more than one mode is intrinsically intermingled with the sampling design. If a larger sample is needed to have smaller standard errors of estimated values, given a certain budget burden, less expensive methods must be chosen. If it is known that there are risks of incomplete coverage because part of the real population could be outside of the frame, a field exploration of the interviewer could be necessary, and a data collection method implying the presence of an interviewer could be necessary.

Much has been written (Groves and Heeringa, 2006) on how to mix the choice of modes. Mainstream opinion is concentrated around the idea that the choice of the data collection mode should be adaptive and responsive. The first adjective is intended as representing the ability to foresee the best option for increasing the response rate according to the socio-demo characteristics of the selected sample, and responsiveness is about 'just in time' learning from the ongoing survey to correct errors if it is not possible to prevent them. This approach has very often built very complex statistical models to predict the best strategy to adopt.

When a survey with a panel component has to be planned, there is the possibility of using one data collection method for the first wave and another data collection method for the subsequent waves. When the survey is planned as an occasional edition and different modes must be concurrent, the problem is identifying the correct sequence of the methods to be used. Differently, economically driven criteria could lead to postponing the most expensive methods (usually Capi or any other mode implying an interviewer exploring an area during a field period) or the least expensive ones. In the first case, the attempt is to try to obtain as many respondents by telephone or mail as possible, to reduce the economic

burden and to complete the missing section of the sample with a face-to-face mode of data collection, which usually has a higher response rate. In the second case, the section of the sample considered as strategic or less reachable is immediately interviewed with the most expensive mode, and the response rate of the other sections of the sample complete the interviewed population.

It is difficult to say which choice is better, because, even if - as we have said above - there is increasingly an attempt to predict the propensity to answer, response expectations have often changed in different directions according to different influences. This is why the capability to tailor the design of research according to the situation is so often quoted (Dillman *et al.,* 2009). Of course, one consequence of this approach is that the old idea of interpenetration[4] (Mahalanobis, 1958) is currently difficult to apply.

In analysing the mode effect in any of the examples made above, there will be a problem of different responding populations. For example, in an Estonian survey about the awareness of gender-based violence, as in other surveys in Nordic countries, It was foreseen the possibility to shift from the Capi to the Casi mode: in other words, from a face-to-face interview mediated by a laptop to a self-administered interview. If this can increase response rate, and - as we will discuss later - stimulate disclosure about the topic, it is clear that a self-selection bias can exist. In this case, it is very difficult to separate the bias exclusively caused by a mode effect from the one caused by sample design or propensity to answer.

In the Netherlands, trying to assess the bias coming from the use of two different interview modes in two subsequent phases of the National Crime Victimisation Survey, they found that, both for Cati and paper surveys, it takes more effort to reach victims of crimes and persons who feel less secure (Cobben *et al.,* 2014).

On the contrary, in France Razafindranovona (2016) found that in self-administered surveys there was a positive self-selection, because within the household, people who were victimised wanted to answer in substitution of the person who had been selected. Problems of self-selection were faced in the Italian Violence against Women Survey as well.

In the last Italian Survey of Violence against Women, Cati and Capi were used as concurrent modes and the sample was extracted from the Municipalities' Register. The individuals extracted were then coupled with a list of phone numbers available from commercial registers. Some of them matched available phone numbers while others did not. Presentation letters to enforce the propensity to answer were mailed to the entire sample. The letters to the selected people, whose name did not match with the marketing registers, included the request for phone numbers, to be contacted later. It created three subpopulations: the one that was contacted at the available phone number, the one comprised of people who sent their own phone numbers, and finally, foreign women who were extracted for a face-to-face interview. As we can see, there were differences, sometimes unexpected, among these populations (Table 4.3).

The interruption rate of the Capi interviews was, for example, not lower than for the Cati interviews. This is not a small matter for the agreement in the existing literature about the fact that Capi should guarantee a higher response rate and willingness to complete the questionnaire (Smith, 1984), but this unexpected result could be attributed to a higher propensity of foreign women to refuse the interview. The truly surprising difference is however

4 Interpenetration is a sample design in which the population inhabiting the same area will face different stimuli by different interviewers (the original idea was about interviewer effect, but the same concept could be transposed to mode effect).

between the Cati interviews themselves. Italian women who were contacted at their mobile phone numbers were clearly a quite distinct population, showing less collaboration. At least part of these respondent women was a specific subpopulation. This point was suggested to us not only by a higher interruption rate, but also by a higher violence prevalence rate, so that it could be hypothesised that there has been a sort of self-selection bias among the Cati interviews, even for what regards the kind of telephone channel chosen.

Table 5.3 - Interruption rate and Refusal rate according to type of sample

TYPE OF SAMPLE	Interruption Rate	Net Refusal rate*
Registered in Telephone registers contacted at mobile phone	134	1953
	0.82	11.89
Registered in Telephone registers contacted at fixed phone	3	18
	0.52	3.09
Missing in Telephone registers contacted at fixed phone	41	115
	0.42	1.19
Missing in Telephone registers contacted at mobile phone	42	459
	0.68	7.43
Face-to-Face Capi Interview	30	366
	0.80	9.6

Source: Italian Violence against Women Survey (2005). Figures refer to the 5th month of fieldwork
*Net refusal rate is a refusal rate calculated only among the persons who began the phone contact.

Summary

- Data Collection modes have to be intended as a process rather than as a technical mechanism
- The interaction between the data collection mode and sampling design has to be carefully considered
- Very often, it will be difficult to correct the error behind the mode effect if the different data collection modes are not randomly assigned. There are clear signs of self-selection in the data collection mode
- Different modes exist, able to mix the precision of the electronic questionnaire with other features
- A certain degree of self-selection is unavoidable

5.1.2 Mode effect related problems

A multi-mode approach to data collection had begun both due to economic reasons and as a solution to the declining response rate, but that does not mean that researchers were concerned only with economic and management issues. Problems arising from a multi-mode approach have been taken into account in an enormous amount of literature.

It has been more than twenty years that Survey research has taken a “cognitive turn” (Tourangeau *et al.*, 2000) studying the psychological, cognitive and emotional issues behind the completion of a questionnaire. Part of this literature concerns how a specific data collection mode influences these processes.

The outcome of this research can be classified in two macro-fields. Research about social desirability and research about ‘satisficing’ (Roberts, 2007).

Social desirability and disclosure

The point, mostly interesting victimisation surveys about violence, is maybe desirability. We should try to understand exactly where this topic is relevant when it concerns the problem of domestic violence. It is well known that respondents, when asked about sensitive issues, could answer driven by an idea of what has been described as a "socially desirable" response. In other words, the respondent will try to encounter what he/she imagines to be the most appropriate answer to represent his/her behaviour as socially acceptable by the dominant social norms. To explore the relevance of this topic, two questions must be clarified. The first is about what a sensitive issue represents and the second involves investigating the best method for asking a sensitive question.

Survey research is increasingly expanding into fields which were unexplored until a few decades ago. Research about unhealthy behaviour in Epidemiology is an example of this kind of research. There is an ethical and practical interest in having knowledge about social issues that take place entirely in the private sphere of individuals, but nonetheless have consequences on society. Domestic violence and intimate partner violence (IPV) is another typical example of that. It is extremely difficult to be aware of the true dimension of this sort of phenomenon, because it cannot be tracked on administrative registers, nor by other visible indicators.

In this case, a representative sample of people has to be asked directly about the phenomenon you are interested in. In this case, the respondent is not going to be asked about his/her structural characteristics, but about things or facts implying:

- his/her behaviour or beliefs regarding legal or social norms;
- his/her behaviour or beliefs regarding institutional duties.

The topic of sensitive questions has recently been considered regarding the Labour Force Survey concerning some aspects of job-searching behaviour. Being active in searching for a job is something to be expected for unemployed people, and the over-estimation of active behaviours in the labour market was under scrutiny. Victimisation surveys finalised at discovering the real prevalence rate of the dark figure of crime (violence) are by all evidence much more involved in this aspect. In the case of IPV or more generally violence against women, being a victim or having been a victim it is not something that is easily revealed. As it is sadly known (Suarez and Gadalla, 2010), being a victim of a sexual crime or of gender violence is a condition that is often turned against the victim herself. The dominant patriarchal discourse tries to instil doubt about the complete innocence of the victims. A number of myths about the reasons of the perpetrator's act are still circulating. These false beliefs assume in the last instance that if you are victim of a rare event, you probably have done something risky other people have instead not done, if you did not provoke it at all. Of course, these risky behaviours concern just the personal sphere, ranging from being too attractive to having a dangerous use of free time, so that there is pressure to avoid this type of social stigmatisation. Furthermore - in the case of IPV - if the victim is still in a relationship with the partner, she could be subjugated and want to avoid legal consequences for her partner.

On the other hand, given the socially constructed roles of men and women in most societies, it could be considered a source of shame or ridicule for a man to reveal having been victim of physical violence from a woman. That is why it is not simple to disclose a state of gender victimisation. At the same time, the declarations of having reported the violence to the police could overestimate the actual behaviour of reporting. A woman could

be positively biased to reveal that she has reported to the police to demonstrate the effective gravity of the violence she suffered, or because she imagines that non-reporting could be considered as anti-institutional.

Additional difficulties arise when considering that some of the factors influencing the constraint on the answers to sensitive questions are culturally related to an individual society and are not universally valid for any single country (Harkness *et al.*, 2010). Looking at the results from the existing European Survey (FRA, 2014) the first thing to appear is a counter-intuitive highest value in Northern Europe. There is debate about the possibility that these data reflect just a higher disclosure (and maybe a greater capacity to recognise violence) of the Nordic women rather than reflecting a worse situation from the point of view of IPV.

In the same way, culturally diverse subpopulations (such as migrant women among native women) could have a very different perception of what is socially expected from them. According to the figures of the World value survey, there is - in some countries - a very high share of women who think men are somehow legitimated to beat their partner in trivial circumstances.

We have grossly defined what a social desirability bias is and why it is so relevant for a victimisation survey about violence. Now we should ask ourselves which data collection modes are the best for overcoming this potential source of error. Again, there is no universal answer to this question, even if the literature agrees on the fact that sensitive issues should be collected with a self-completed questionnaire without any interviewer. There are a number of contemporary modes - as we have already seen - which could meet these requirements, and - for quite a long time - evidences from experimental empirical research (Aquilino, 1994) demonstrate that topics like the use of drugs/alcohol, sexual or deviant behaviours, etc., show higher prevalence rates if collected through the self-administered questionnaire.

That does not mean that choosing a self-administered mode of collection is immediately the best option:

- victimisation surveys have long and complex questionnaires, for which an electronic version is usually considered necessary. The mail survey is therefore automatically excluded, while its actual substitute (Cawi) is dependent upon the state of cable infrastructure and internet literacy of any single country;
- even if supported by an electronic questionnaire in respecting the correct sequence of questions, the questions could be of difficult comprehension by the interviewee (especially those persons with a lower level of education);
- in case of questionnaires dedicated to minorities or other subpopulations, the text should be necessarily translated without loss of meaning, otherwise the advantage of a self-completed questionnaire could - in absence of any interviewer - be more than undermined by the linguistic barriers;
- with a self-administered questionnaire, the interruption rate is usually higher, because the only feeling of duty the respondent has is towards the institution organising the survey;
- furthermore, some respondents could feel emotionally distressed and the fact of being alone could present an obstacle;
- the Internet survey and the mail survey have - according to most of the results - more or less the same performance in terms of response rates, but personal interviews supported by advance letters have higher response rates than self-administered questionnaires do.

The challenge is to find a point in the trade-off between the higher credibility and motivation for the survey guaranteed by a personal interview and the confidentiality given by a self-completed interview.

Satisficing

Cognitive Survey Research (Biemer, Lyberg, 2003; Krosnick, 1991) usually divides the answering process into four phases:

1. Comprehension, where the respondent has to infer the real meaning of the question;
2. Retrieval, where the respondent has to recall the requested event from his/her memory;
3. Judgement, where the respondent evaluates the consistency of what he/she has just remembered with what has been asked;
4. Response, when the question is actually answered.

During these phases, the cognitive effort required varies from question to question, but is always a burden for the respondent. It calls for his/her attention and engagement. It is always possible that respondents deviate from the necessary attention required, and give the best answer they are able to provide quickly, but that is a sort of second best option compared to the real answers. This process, which has been called satisficing[5], has many outcomes. One is the so-called item non-response rate. 'Item non-response rate' is a measure of the attitude for which an item is completely skipped, which occurs more often when it is a paper self-administered questionnaire; the option 'Don't know' is chosen in order to get rid of the question quickly. Concerning this dimension, there are no unique results. While the existing literature (Denscombes, 2009) comparing different channels of the same self-completion questionnaire (on paper and online) does not give evidence of a mode effect on the item non-response rate, other experiments comparing self-administered online surveys with interviewer-administered surveys, give clear results in favour of interviewer administered data collection modes.

Data from the module on sexual harassment in the Italian Citizen's Safety Survey, show that the degree of proximity is important for increasing the embarrassment and the difficulty in answering sensitive questions. Below you can see the effect of the Cati mode compared to the Capi mode, on the item non-response rate of some sensitive questions about experiences of sexual harassment. The effect is expressed in β coefficients of a logistic regression controlling the mode effect for the basic socio-demographic variables, and it is negative and always statistically meaningful.

Table 5.4 - Mode effect on item non-response rate of different sorts of harassments screened in the Italian Citizen's Safety Survey

	Item non-response rate			
	Verbal harassment	Sexual harassment	Forced to watch porno images	Exhibitionism
β Coefficient of Cati Mode (reference category Cati)	-0.27	-0.51	-0.30	-0.58
χ Wald	3.54	14.58	3.84	19.75
Pr>χ	0.05	0.0001	0.04	<.0001

Source:

5 With a merge between the words sufficient and satisfy (to represent the idea that a second best option is enough).

- Problems concerning the cognitive capability of the respondent to put the right effort in answering are interconnected differently with each interviewing mode.
- Problems concerning the sensitivity of the questions are interconnected differently with each interviewing mode.
- The presence of an interviewer is better for solving the first kind of problem.
- Self-administered interviewing modes are better for the second kind of problem.

5.2 Strategies for controlling differences in the interviewing mode

After analysing the mode effect, we shall now discuss how to check for the size and direction of non-sampling errors caused by the interviewing mode. In his review, Korner (2014) divides methods of validation into two approaches: experimental and non-experimental approaches.

Experimental approaches can occur concurrently, or in two steps. They occur concurrently when a sample is split in two, and one of each part is exposed to a distinct data collection (Cobben *et al.*, 2014). Difficulties arise in establishing whether the samples are identical for condition of exposure to the mode effect. If the experiment is planned in two steps, the same sample is going to be interviewed in two steps with different modes. Here, the difficulty is choosing the correct time interval at which to re-interview the same person, who will probably remember the given answer and try to reply in the same way (Schouten, Van der Laan, 2014). A number of difficulties derive from the fact that an experiment is the attempt to reproduce an abstract situation to verify a single hypothesis in the lack of any disturbing noise, but if carrying out a survey is a process of interacting with various factors, it is hard to assert that none of these factors is intervening. What's more, in the case of an experiment with a split sample, it is difficult to have samples large enough to be representative on both modes with reasonable standard errors, since most probably the non-sampling errors caused by the mode effect will be relatively small.

In the use of non-experimental approaches, modelling is prevalent. It can be a regression model or a recalibration of the results, but what the various approaches based on modelling share in common is the use of a third variable (which is supposed to be mode-independent) to produce unbiased estimates to compare with the results affected by an interviewing mode bias. Very often, especially in Germany and the Netherlands (see Buelens Van den Brakel, 2014), a recalibration technique is used to obtain a non-biased result. It implies the construction of a model with the interviewing mode as a dependent variable, to produce weights to reallocate the single respondents according to the Capi or Cati mode, correcting the selection bias. This poses the problem of which variable to use, and if it can really be defined as mode insensitive. They can be:

- variables retrieved by an administer register through record linkage;
- variables taken from the survey itself, but mode insensitive (demographic variables);
- para data (quality indicators collected during the field phase).

These variables can be accompanied by meaningful covariates from the survey itself. The results produce coefficients, which can be used to recalibrate the selection effect of the mode in case there was a no controlled allocation of the interviewing mode. Record Linkage has been used also for identifying the characteristics of non-respondents and, in this way, estimating the mode effect on the non-response rate, or trying to discover the "true" value beyond any social desirability bias (Kreuter, Presser and Tourangeu, 2008).

Another way to use modelling is to build a regression model using the data collection mode as an independent variable to estimate the size of the mode effect.

Along with modelling methods to assess the mode effect, if qualitative methods are considered to be more appropriate, there is always the possibility of re-examining the answer given together with the respondent through cognitive interviewing methods.

Another completely different strategy to check for the mode effect is to adopt different multivariate strategies of data analysis. The task is not to compare the biased estimates with some sort of unbiased estimates to correct the first ones, but the attempt is to find latent dimensions of meaning between the variables, to see if the cases are related to the dimensions according to the interviewing mode used. This strategy of analysis allows you to check if the mode effect has somehow altered the contents of the survey itself (Razafindranovona, 2016).

The interviewer effect in the Violence against Women survey

Among the non-sampling errors that may affect a survey using direct data collection modes is the so-called 'interviewer effect', that is, the effect on the final estimates introduced by the behaviour and the actions of the interviewer. This issue may be particularly felt in surveys dealing with sensitive themes, as the trust with a specific interviewer, or lack thereof, may influence the answers of the respondent.

In their work based on multilevel regression Bocci *et al.* (2005) show that, for the Italian Citizen's Safety Survey, questions regarding sexual crimes were more at risk of increased variance than questions dealing with less sensitive themes (for example, questions on crimes against property). Such results are to be expected and have been confirmed in an analysis of the 2014 data of the VAW survey with the use of similar methodologies that is the application of multilevel models for the computation of the Intraclass Correlation Coefficient (ICC).

For this survey questions about sexual violence generally show a greater presence of interviewer effect as measured by the ICC than, for example, questions on physical violence. As the survey adopted a multi-mode approach, the Cati component and the Capi one were investigated separately, considering the fact that the two pools of interviewers did not overlap. The analysis has shown that the issue of the presence of interviewer variance was more relevant for the Capi component. A number of reasons may be assumed for this: Capi interviewers are usually less monitored than their Cati counterparts; the face-to-face approach may leave more freedom to (erroneous) interpretation of questions by the interviewer, and so on. However, in this case the most relevant explanation may lie in the fact that the Capi mode was used exclusively for foreign women, so a nationality effect could have played a role in the respondents' answers. Unfortunately, discerning such effect from a mode or interviewer effect can be quite complex. A way to explore this issue, which is currently under investigation, is using the common questions and variables in the Citizen's Safety Survey as a comparison term.

Multivariate analysis to assess the impact of the mode effect on survey estimates

The last wave of the Citizen's Safety survey was carried out using a mixed mode of data collection, referring to two different samples of people, but using the same questionnaire.

The reason for using two different modes, the Computer-Assisted Telephone Interviewing (Cati) and the Computer-Assisted Personal Interviewing (Capi), is related to coverage issues. An increasing proportion of households with only mobile phones and the problem of incomplete lists of owners of landline phones due to reserved numbers has increased non-response or measurement error problems. In fact those with a landline

connection differ in many ways from households with only mobile phones; for example they differ in their demographic and social characteristics and consequently in the survey topics, suggesting that a new type of survey combining different data collection modes could be, instead of one mode only.

The main list was the municipality's list of citizens, linked to the database of owners of landline connections to obtain two different frames from which to design and select two subsamples, households owning a landline connection (Cati mode) and households not owning a landline connection (Capi mode), to be combined for the final estimates and outcomes.

The two samples are balanced enough in terms of gender and territorial areas in which the respondents live. Greater differences refer to age, marital status, educational level and work condition. In the Cati sample, there were older respondents (over 50 years) but also more 18-30 aged married respondents and wider households.

In the Capi sample there were more middle-aged respondents (31-50 years old), divorced people, low level professions (workers or those looking for an employment) and lower level of education. All of the above variables are related to the experience of sexual victimisation. After weighting the two samples, the above differences remain, but to a lesser extent.

As follows, a descriptive multivariate analysis has been performed to evaluate if the mixed-mode data collection affects the profile of the sexual victims.

The analysis consists in a Multivariate Correspondence Analysis (MCA) of main violence indicators followed by a discriminant analysis on the quantitative factors obtained by MCA to check if different interviewing techniques, controlling for social and demographic characteristics of respondents, affect victimisation rates.

To perform the MCA analysis, the first step has been to recode main indicators as follows: the screening questions about sexual harassment, some types of stalking behaviours (shadowing, insistently calling on the phone, indecent exposure…), sexual proposals so as to obtain improvements on the job, etc., have been recorded in terms of victims in the 12 months before the interview, distinguishing victims if once or more than once, victims in the last three years but not during the last year, victims before the last three years from the interview.

First analysis: MCA of thematic and social demographic variables

After recoding the variables, a multivariate correspondence analysis has been performed on the whole of the recoded variables, including some of social demographic variables referred to the interviewee (sex, age, area of residence and type of municipality, educational level, professional level, number of members of the interviewee's family), that could affect the risk of victimisation. Sample weights to describe associations among categories of the variables in the population have been used in the analysis.

Two factors have been considered in the MCA analysis, the first one relating to the social-demographic variables, the second one identifying the multivariate victimisation profiles indicating more serious situations (more types of sexual or physical violence suffered by respondents: more recently and repeatedly) with increasing and positive values. The two factors, considering the Benzecri index of goodness of fit, explain 72% of the total variability described in the multivariate contingency table[6].

6 The Benzecri index corrects the sparseness of categories, each considered as a single dummy variable in multivariate correspondence analysis.

In greater detail: the first MCA factor individuates with negative values no victim, males younger than 30 years old, unmarried, with low level of education and modest professions or looking for employment, living in households with more than 4 members, mainly from southern regions and Sicily, and from small municipalities (less than 10,000 inhabitants). With positive values the factor individuates females from central metropolitan areas, central regions, but also from Trentino Alto Adige and Friuli Venezia Giulia, with higher levels of education and more qualified professions. Positive values of this dimension are also related to victimisation experiences mainly before the last three years before the interview, and with sexual victimisation before 16 years of age.

Positive values of the second factors indicate victimisation in the last three years or in the last year before the interview, more than once, representing with increasing values the emergency posed by this type of victims.

The weighted means of the two factors by the mode of data collection (Cati or Capi) show few differences, with a slightly greater capability of the Capi mode to discover victimisation.

Second analysis: Discriminant analysis on the above MCA profiles

The two dimensions summarising the above profiles with quantitative scores, have been used in a discriminant analysis to assess if the two modes of data collection (Cati and Capi) discriminate between different profiles in particular between victims and no victims.

The classic discriminant analysis referred to quantitative variables provides coefficients of a linear combination of the variables versus the grouping (discriminant) one, considered as a response.

Such coefficients are provided under some constraints: to minimise the variance within the new groups (within variance) and maximise the distance between the barycentre of the groups (between variance), and is lastly defined as a problem of eigenvalues and eigenvectors of a matrix based on the within and the between covariance matrix (Anderson, 1984).

The analysis provides the following linear discriminant functions:

Linear discriminant function for Capi variable (1=Capi; 0=Cati)

Variable	0(Cati)	1(Capi)
Constant	-0.00570	-0.00703
Dim1	0.20041	0.04448
Dim2	0.13955	-0.28098

The values of the two linear combinations will be evaluated for each unit, and each unit will be reallocated to the group that, according to the corresponding linear combination, will provide the lower distance with the centre of the group.

The linear discriminant function provides a more balanced reallocation of the units as we obtain 43% of units reallocated in the Cati group and 57% reallocated in the Capi group. Considering both the factors, the first one controlling for the social-demographic characteristics, the function is not able to definitely discriminate among groups, rebalancing samples in favour of the smallest one (Capi).

When performing the discriminant analysis with only the second factor, related to more serious experience of victimisation, the reallocation is less balanced, inflating the Capi

group (61% after reallocation) due to the high negative coefficient for the variable dim2 (the second factor), not balanced as it was in the previous analysis, by the positive values referring to social-demographic characteristics.

Linear discriminant function for Capi variable (1=Capi; 0=Cati)

Variable	0(Cati)	1(Capi)
Constant	-0.00163	-0.00683
Dim2	0.13741	-0.28145

The final consideration is that the two data collection modes perform somewhat differently on sensitive topics, with a little more disclosure when the interviewees are not in front of the interviewer. In other words, from the point of view of self-disclosure, Cati is a better performing mode, even if when controlling for the social-demographic characteristics distributed differently in the two groups, the differences in disclosure are less frequent.

- Different methods, both experimental and non-experimental, can be used to take account of the mode effect.
- Experimental methods usually have problems of resources and difficulty in reproducing independent situations.
- Mode-independent variables can be used to model unbiased results.
- Linkage to administrative data is another path to compare obtained data with unbiased results.
- Different strategies of Multivariate data analysis and also different models as for example multilevel models, can be used to check if item results are biased by interviewing mode.

REFERENCES

Anderson, T.W. 1984. *An Introduction to Multivariate Statistical Analysis, 2nd Edition*. Hoboken, NJ, U.S.: John Wiley & Sons.

Aquilino, W.S. 1994. "Interview Mode Effects in Surveys of Drug and Alcohol Use: A Field Experiment". *The Public Opinion Quarterly*, Volume 58, Issue 2: 210-240.

Biemer, P.P., and L.E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ, U.S.: John Wiley & Sons.

Bocci, L., M.G. Muratore, M. Signore, G. Simeoni, and G. Tagliacozzo. 2005. "Analisi dell'effetto intervistatore attraverso l'utilizzo dei modelli multilevel". In Istituto Nazionale di Statistica – Istat. "Il monitoraggio del processo e la stima dell'errore nelle indagini telefoniche. Applicazione all'indagine sulla sicurezza dei cittadini", Capitolo 6: 73-93. *Metodi e Norme*, N. 25/2005. Roma, Italia: Istat. https://ebiblio.istat.it/digibib/Metodi%20e%20norme/IST0048279Ed2005N25.pdf.

Buelens, B., and J. Van den Brakel. 2014. *Measurement error adjustment in the Dutch Labour Force Survey*.

Cobben, F., B. Schouten, and J. Van der Laan. 2014. "The impact of contact effort on mode-specific selection and measurement bias". *Survey Methods: Insights from the Field*. http://surveyinsights.org/?p=3629.

de Leeuw, E.D. 2005. "To Mix or Not to Mix Data Collection Modes in Surveys". *Journal of Official Statistics – JOS*, Volume 21, Issue 2: 233-255.

de Leeuw, E.D. 1992. "Data Quality in Mail, Telephone and Face to Face Surveys". *Technical Report*. Amsterdam, The Netherlands: T.T. Publikaties.

De Vitiis, C., A. Guandalini, F. Inglese, and M. Terribili. 2017. "Methods for treating bias deriving from different sources of non-sampling errors in mixed mode social surveys". *Internal Document* (unpublished), for the Methodological Advisory Board of the Italian National Institute of Statistics – Istat. Roma, Italia: Istat.

Denscombe, M. 2009. "Item non-response rates: a comparison of online and paper questionnaires". *International Journal of Social Research Methodology,* Volume 12, Issue 4: 281-291.

Dillman, D.A., J.D. Smith, and L.M. Christian. 2009. *Internet, Mail, and Mixed-Mode Survey: The tailored design method*. Hoboken, NJ, U.S.: John Wiley & Sons, Inc.

European Union Agency for Fundamental Rights - FRA. 2014. *Violence against women: an EU-wide survey. Main results*. Luxembourg: Publications Office of the European Union. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2014-vaw-survey-main-results-apr14_en.pdf.

Groves, R.B., and S.G. Heeringa. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs". *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. Volume 169, Issue 3: 439-457.

Harkness, J.A., M. Braun, B. Edwards, T.P. Johnson, L. Lyberg, P.P. Mohler, B.-E. Pennell, and T.W. Smith (*Eds.*). 2010. *Survey methods in multinational, multiregional, and multicultural contexts*. Wiley Chichester. Hoboken, NJ, U.S.: John Wiley & Sons, Inc.

Korner, T. 2014. "Report on the definition, Identification and Analysis of Mode effects". *Deliverable for WP III of the ESSnet on data collection for Social Survey Using Multiple Modes*. Wiesbaden, Germany: Statistiches Bundesamt.

Korner, T., A. Liesch. 2014. "Case study on mode effects in the German Labour Force Survey". *Deliverable for WP III of the ESSnet on data collection for Social Survey Using Multiple Modes*. Wiesbaden: StatistichesBundesamt. Wiesbaden, Germany: Statistiches Bundesamt.

Kreuter, F., S. Presser, and R. Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity". *Public Opinion Quarterly*, Volume 72, Issue 5: 847-865. https://doi.org/10.1093/poq/nfn063.

Krosnick, J.A. 1991. "Response strategies for coping with the cognitive demands of attitude measures in surveys". *Applied Cognitive Psychology*, Volume 5, Issue 3: 213-236.

Mahalanobis, P.C. 1958. "Recent Experiments in Statistical Sampling in the Indian Statistical Institute". *Sankhyã: The Indian Journal of Statistics (1933-1960)*, Volume 20, Issue 3-4: 329-398.

Razafindranovona, T. 2016. "Exploitation de l'enquete experimentale: Vols, violence et sécurité". *Document de travail*, N. M2016/03. Paris, France: Institut National de la Statistique et des Études Économiques - INSEE. https://www.insee.fr/fr/statistiques/2022146.

Roberts, C. 2007. "Mixing modes of data collection in surveys: A methodological review". *NCRM Methods Review Papers*, NCRM/008. Southampton, UK: National Centre for Research Methods – NCRM. https://eprints.ncrm.ac.uk/id/eprint/418/1/MethodsReviewPaperNCRM-008.pdf.

Smith, T.W. 1984. "A Comparison of Telephone and Personal Interviewing". *General Social Survey (GSS) Methodological Report*, N. 78. Chicago, IL, U.S.: National Opinion Research Center, University of Chicago. https://gss.norc.org/Documents/reports/methodological-reports/MR028.pdf.

Smith, T.W., and J. Kim. 2015. "A Review of Survey Data-Collection Modes: With a Focus on Computerisations". *Sociological Theory and Methods*, Volume 30, Issue 2: 185-200. https://gss.norc.org/Documents/reports/methodological-reports/MR126.pdf.

Suarez, E., and T.M. Gadalla. 2010. "Stop Blaming the Victim: A Meta-Analysis on Rape Myths". *Journal of Interpersonal Violence*, Volume 25, Issue 11: 2010-2035.

Tourangeau, R., L.J. Rips, and K. Rasinski. 2000. *The Psychology of Survey response*. Cambridge, UK: Cambridge University Press.

United Nations Office on Drugs and Crime - UNODC, and United Nations Economic Commission for Europe - UNECE. 2010. *Manual on Victimisation Surveys*. Geneva, Switzerland: United Nations.

Van Kesteren, J., M. Aebi, and J. Van Dijk. 2009. "Discussion paper on Interview Modes". *Working paper*, prepared for the *Task Force on the preparation of a Victimisation Survey Module*. Tilburg, The Netherlands, and Lausanne, Switzerland: University of Tilburg and University of Lausanne.

6. SELECTION AND TRAINING OF INTERVIEWERS[1]

6.1 Skills and attributes of interviewers

Interviewers play a crucial role in the data collection phase. Personal characteristics, skills and attitude of the interviewers should be given great consideration, as they affect the willingness of respondents to participate in a survey and to disclose very personal and sensitive information. Interviewers must have the qualities needed to obtain the required information with accuracy and within a reasonable timeframe, which is why they should be selected with great care and receive specialised training and ongoing support.

In recruiting and selecting potential interviewers it is important to consider several factors also based on the socio-cultural context where the survey is carried out. Among these factors, gender, age, education and previous experience in interviewing are key requirements to consider.

Experience to date suggests that women respondents feel most comfortable talking about violence with other women (EIGE, 2017). Utilising female interviewers increases the disclosure of sensitive information, particularly what relates to experiences involving sexual victimisation and violence perpetrated by male partners.

Generally female interviewers improve the disclosure of these events regardless of the gender of the respondent. The Norwegian experience with the survey "Violence and Rape in Norway"[2] suggests that female interviewers performed well also in surveys on violence against men, because they "act as a facilitator" (Barletta, Muratore, Scarnicchia, 2017). Nevertheless, in countries where sexual discrimination is common and a female interviewer might have difficulty in obtaining an interview, it could be wise to provide a couple of interviewers (a man and a woman) to facilitate contact households and decrease refusals.

While it is not possible to recommend an age limit or age range applicable across all settings, many respondents perceive older female interviewers as instilling more warmth and reassurance than younger women. It has been found, indeed, that adult women prefer not to discuss violence with interviewers whom they perceive may not understand or be sympathetic to their experiences. In some contexts, it would be considered even inappropriate for a young woman to pose questions about violence to an older woman. Interviewers who are seen as too young may elicit the distrust of respondents, leading to an outright refusal to participate in the survey or reluctance to disclose personal or sensitive information.

It is also recommended to select interviewers with a higher than primary level of education - as suggested by the World Health Organization (2005) - to be able to manage the sensitivity of the topic, the relationship with the respondents and the complexity of the questionnaire.

A working knowledge of interviewing techniques, knowledge and experience of technology, such as personal computer or telephone interviewing, are essential but can be reinforced by the

1 This chapter has been drawn up by Roberta Barletta.

2 The survey has been conducted in 2013 by the Norwegian Centre for Violence and Traumatic Stress Studies (Nkvts) with a sample of 2,435 women and 2,092 men.

training programme. Nevertheless, it should not be assumed that the regular interviewers used for other surveys are necessarily appropriate for surveys on gender-based violence. Even very experienced interviewers who have demonstrated competence in other surveys may not feel equipped to manage the potential stresses that participation in this kind of survey may bring. Selection of interviewers should ensure they have been carefully briefed on the subject matter and have considered their ability to perform effectively over the duration of the fieldwork.

Beyond the above requirements, the presence of which can be verified considering the curriculum of the potential interviewers, a set of other elements should be assessed with personal interviews. A possible summary of elements to be evaluated in the interview with candidates is the following:

- personal motivations in choosing the interviewer's work;
- communication skills;
- ability to engage with people of different backgrounds in an empathetic and non-judgemental manner;
- ability to build an appropriate relationship with the respondent, to establish and maintain an appropriate level of professionalism while expressing warmth and reassurance;
- prosody, tone and cadence of the voice;
- emotional maturity, awareness of one's emotions related to problematic situations and difficulties that may arise when interviewing. Interviewers in possession of personal strategies for self-care and stress reduction will be better able to perform their duties for the duration of the interviewing period;
- presence of judgements, stereotypes or prejudices towards persons who have experienced violence;
- availability to involve own self in an active training - including role-playing, simulations, discussions and group cooperation - documented, for example, by similar previous experiences.

It is recommended to always recruit more fieldworkers/trainees than needed, since some may not be adequate and others may drop out.

From the Manual on Sensitive training for the International Violence against Women Survey (Johnson, Ollus and Nevala, 2007)

Interviewers should possess the following characteristics:

- a level of knowledge and comfort discussing issues related to violence against women and men;
- a non-judgemental and empathetic attitude toward victims of domestic violence and sexual assault;
- good interviewing skills, preferably with interviewing experience;
- a pleasant personality and ability to relate to people in all walks of life;
- the absence of highly biased views related to the subject matter or characteristics of respondents (age, social status, ethnicity, religion);
- willingness to adhere to instruction and time schedules;
- willingness to report problems and discuss them in an objective manner;
- an acceptable level of literacy and ability to understand the questionnaire and issues being surveyed;
- knowledge of and facility in local languages and dialects;
- availability for evening and week-end work;
- all interviewers must be female.

The nature of the job should be made clear at the time of hiring. Some women may be attracted to the job because of their interest and experience in working with victims of violence, but it must be stressed that interviewers should not consider themselves counsellors. Women with counselling experience may bring some good traits to the job, but must be firmly discouraged from acting in a counselling role. This may prove to be frustrating for some interviewers who may drop out of the project during the training.
Specific considerations in selecting the interviewers must also be considered. These considerations include the ethnic composition and age range of interviewers.

The training is essential not only to provide all knowledge and skills needed to conduct the interviews in a scientific and ethical way, but also to complete the selection process. Based on the learning outcomes and the personal skills showed during the training by the potential interviewees, it is possible to take the most appropriate final decision.

When the NSOs already have interviewers, the researchers' team needs to select them through curriculum vitae and individual interviews, choosing the persons who better meet the selection criteria as described above (personal characteristics, skills and attitudes).

The selection of the field supervisors is also extremely important. In some cases, field supervisors are selected early in the planning phase before the interviewers have been selected. However, in many surveys using face-to-face interviews, field supervisors are selected among the participants in the interviewer/fieldworker training, either during or after the training (Jansen, 2010). It cannot be assumed, just as for the interviewers, that field supervisors who perform well on other surveys are suitable or willing to work on a survey on violence because of the additional stressors that such a survey entails. Field supervisors should also be carefully selected based of their maturity, their motivation to work on the topic and their work experience on surveys with similar sensitive topics. Like interviewers, field supervisors must go through extensive training before working on a survey on violence, including specific training on their tasks and responsibilities as supervisors.

According to the United Nations Guidelines for producing Statistics on Violence against women (UNSD, 2014), in face-to-face interviewing situations, field supervisors must be female, since they must travel with interviewers periodically to oversee their work while in telephone interviewing situations, it is possible to use male field supervisors, provided they do not have any contact with respondents and provided they have the training and sensitivity required to train interviewers and to support them through regular debriefings.

Due to the complexity of the aspects to be evaluated, when possible, it could be useful to select interviewers and field supervisors through a specialised company. However, even in this case, the research team has to define accurately personal, motivational and psycho-attitudinal requirements of the interviewers and field supervisors in order to meet the research goals.

In the Italian VAW survey, the researchers' staff for the recruitment of interviewers have undertaken the following phases:

1. Pre-selection of curricula;
2. Individual interviews;
3. Computer testing;
4. Theoretical and practical training;
5. Further selection at the end of the training;
6. Analysis of performance through both quality indicators and observation of field behaviours.

The Job Interview schedule

- Can you talk about your training and your work experience? (How many years of work and how many as an interviewer?)
- If the candidate has experience as an interviewer: What do you like and what do you dislike about this job?
- If the candidate has no experience as an interviewer: What attracts you to this job? What are your fears and worries, and what are your expectations?
- Do you know what the survey is about?
- What do you think the difficulties with the women interviewed might be?
- What are the pros and cons of a telephone (or face-to-face) survey on these issues?
- Reading of some questions of the questionnaire for assessing the voice stress, intonation and inflection, and the attendant reactions to the questions: What effect do these questions have on you? What effect do you think they can produce on the interviewee?
- What do you think is the attitude to keep when a woman gives you an affirmative answer to one or more of these questions? What do you think a woman can expect from you after an affirmative answer to one of these questions?

– Reading some stereotypes about the issue of violence against women: What do you think about it? Do you agree or disagree? Why?
– A brief description of the training activity, then: Do you think you can play actively in front of the group during training and during the debriefing and discussion groups that will follow?
– Willingness to work hard during the training phase, availability to work shifts, availability for the entire duration of the data collection.

6.2 Training of interviewers

The quality of the survey data largely depends on the interviewers' understanding of and commitment to the objectives of the survey as well as on their comprehension of the tools and methodology. Appropriate behaviour and an adequate level of professionalism on their part can prevent respondents from giving erroneous information. Furthermore, non-response mainly occurs when interviewers have either not fully understood the survey goals and concepts, asked the questions inconsistently, or recorded answers inadequately, which, in turn, leads to a general lack of uniformity in the way the survey is implemented.

Therefore, since only a well-trained interviewer can provide quality survey data, it is deemed important to plan and provide appropriate training to the entire team of interviewers.

In order to deliver high-quality training it is essential to collect national and international references on the subject (articles, books, survey reports, etc.), plan the training in all its aspects (aims and objectives, timing and phases, participants, location, tools and equipment) prepare training material: questionnaire guide, interviewer's manual, role-playing, hand-outs, newspapers, books and movies to be used during training and develop tools (questionnaires, tests) for end-of-training assessment of results.

It is recommended to design the training assuming that interviewers have little pre-existing knowledge of the topic. The goal of the training is to ensure that the interviewer's work is qualified, focussed on the interviewee, methodologically correct, rigorously conforming to the interview technique and contents of the research and fully integrated within the whole survey process. All interviewers and field supervisors need to understand all aspects of the interviewing and data collection processes, including the safety of respondents and interviewers and other ethical issues.

With reference to research on violence, all team members need specialised training and support over and above that normally provided to research staff. The objective of the course is to enable the interviewers to perform their duties across a variety of scenarios and possible outcomes. These include any dangers that respondents may face when responding to questions concerning their experiences of violence, as well as ways to guarantee the safety and emotional well-being of the respondents and to protect the confidentiality of the information collected.

This training should include an overall orientation to the concepts of gender, and gender discrimination/inequality and must provide a mechanism for fieldworkers to confront and overcome their own biases, fears and stereotypes regarding, for example, abused women. In addition, training should include an opportunity for research staff to come to terms with their own experiences with abuse.

Interviewers who experienced themselves some kind of abuse can be more empathetic with victims, but also more emotional. Being involved in the study may awaken images, emotions, internal confusion and conflict. These reactions may affect the interviewers' ability to work and may have a negative impact on their health and on their relationships. Even when a researcher or fieldworker herself has not experienced violence, listening to stories of violence and abuse may be draining and even overwhelming. Experience has shown that unless this reality is confronted directly, research projects can experience high rates of attrition among staff.

There are a number of ways to address the emotional needs of researchers and fieldworkers. During the training process, it is important that the subject of violence is openly discussed, and that research team members are given the option of withdrawing from the project without prejudice. During the fieldwork, regular debriefing meetings should be scheduled to enable the research team to discuss what they are hearing, their feelings about the situation, and how it is affecting them. These meetings should be aimed at reducing the stress of the fieldwork, and avert any negative consequences. Interviewers should have the opportunity to discuss this in private with study leaders or a counsellor if they so wish.

6.2.1 Initial training

Considering the sensitivity of the issues investigated, it is therefore recommended that all interviewers attend an extensive and in-depth training course. By the end of the training, each interviewer must have gained full understanding of the following:

- purpose, goals and background of the research project, in order to be able not only to respond to participants' questions but also to inspire confidence and legitimacy to the survey and to further motivate participation;
- overall structure of the questionnaire, including skip patterns, filter questions and specific components and therefore be able to properly 'navigate' through the questionnaire;
- purpose of each question and possible responses;

- concepts and terminology, such as the definition of physical, sexual, economic and psychological violence;
- methodology behind the survey, including sampling and respondent selection as it is common for respondents to want to know how their household or they themselves were selected;
- interviewing techniques to be used, including the correct ways to read the survey questions and to record respondents' answers, and the proper use of the technical tool (such as software).

Moreover, the interviewers must be able to suitably manage the relationship with the respondents. In particular, it is deemed important for the interviewers to learn how to:

- contact the sample members and persuade them to participate in the survey;
- effectively develop and maintain rapport, by creating a climate during the interview process that encourages participation and disclosure, and respond to questions;
- be welcoming, reassuring, responsive, open, patient and tolerant, whilst appearing assertive, empathic and interested;
- manage difficult situations while remaining calm and professional;
- decode non-verbal signals from the respondent (pauses, silences, changes in tone, etc.) in order to ascertain if the interviewee is tired, annoyed, unable to speak at that time or not willing to answer any questions at all;
- avoid leading questions, use non-judgemental language and attitude, and engage in active listening.

Even if questions are worded in a non-leading manner in the questionnaire, interviewers may still ask them in a leading way, for example, by straying from the language of the questionnaire when actually asking the questions, or showing their prejudices and biases when reacting to respondents' questions via facial expressions, body language and tone of voice. The interviewer should not exhibit any prejudices in the way he or she reacts to the respondents' answers. To avoid this, interviewers must be carefully trained to remain neutral and true to the original wording of the questions and the interview protocol. Interviewers must be made aware of potential sources of gender biases that may arise during an interview and trained to avoid such biases. Ideally, questionnaires and interviewers should be jointly tested during pilot/trial interviews.

Due to the extreme sensitivity of the survey topic, all interviewers need also to go through "specialised training" to acquire a good understanding of specific issues, such as:

- violence against women and men and its impact on victims;
- how violence against women and men is defined by law or customs in the country and how it is measured in the survey;
- societal myths concerning women's/men's experiences of sexual violence, intimate partner violence and other forms of violence, how these may affect respondents and their willingness to report their experiences to a survey interviewer;
- ethical requirements of the surveys, including strategies for addressing the confidentiality and safety of and support for the respondents;
- specific skills needed to conduct this type of interview, including how to create a climate that facilitates disclosure of sensitive information;
- the ability to detect when respondents are at risk of being overheard and how to reschedule interviews accordingly;
- how to guarantee the respondents' safety during the interview (this will vary depending on the mode of interviewing);
- how to ensure their own safety when interviewing face-to-face;

– how to identify and respond appropriately to emotional trauma by referring respondents to resources in the local community and by avoiding emotional involvement or counselling;
– how to recognise their own emotional reactions to stressful situations (such as reliving their own traumatic experiences or hearing traumatic stories day after day) and how to develop the skills required to manage and reduce stress.

The main challenge for the interviewer concerns their ability to ask very private questions about experiences of violence in a respectful manner, to accurately assess the feelings or reactions of the respondents in a variety of situations and to respond appropriately.

Respondents may react in many different ways: some may view the survey as an opportunity and may be open to disclosing their experiences; others may be fearful that a violent partner might learn of their participation in the survey, feel disturbed by the content of the interview, be traumatised by recent experiences of violence or feel embarrassed or stigmatised when disclosing their experiences. Accordingly, the training course must go into a multitude of possible reactions by the respondents, teaching the interviewers to respond in an empathetic and supportive manner.

Interviewers must be helped to understand their role in relation to a person who reports experiencing violence. In some cases respondents may show signs of strong emotional distress and plead for immediate assistance to which interviewers need to react in a warm, empathetic but neutral manner. They should be open to assisting them if asked and may provide information on potential sources of support, referring to a pre-prepared list of agencies in the local community that can provide assistance, but they should not tell them what to do. Interviewers should not take on a role as counsellor and any counselling activity that may be offered in the context of the study should be entirely separate from the data collection.

Another important issue to be addressed in the training concerns the safety of the respondents during the interview. Indeed, situations may arise in which the respondent's safety is at risk or it is not possible to guarantee the necessary privacy to respond to sensitive questions. As each mode of interviewing presents different challenges, measures to tackle these types of situations must be tailored to the specific mode of survey administration. For example, in face-to-face interviews, interviewers should be trained to terminate the interview or to change the subject to a "dummy" questionnaire if interrupted. In telephone interviews, interviewers should be trained to detect when a respondent is at risk of being overheard, either because he/she hears other people in the background, or the respondent speaks to somebody else or there are pauses in his/her replies. Before beginning modules of questions on sensitive topics, it is good practice to confirm with respondents that they have the necessary privacy to continue and, if necessary, to reschedule the interview by finding a more convenient time or a safer place.

In preparing the training, it is also important to consider the impact that the topic under investigation may have on the interviewers. Regardless of the mode of interviewing, interviewers will be engaged in emotionally draining work: they will hear numerous personal disclosures of violence and in the case of face-to-face interviews they may also occasionally observe the effects of violence on the respondents. Thus, the well-being of the interviewers must be protected and the training should instruct them to recognise signs of their emotional distress, and manage or minimise it (Claramunt, 1999).

It is advisable to include a counsellor or a psychologist in the project team of the research in order to train interviewers to recognise their own signs of stress overload, and teach them self-care techniques that help minimise the negative effects of stress over the

short and the long term. Furthermore, field supervisors need to learn how to recognise signs of emotional distress in the interviewers and prevent their burn-out by adopting various measures such as offering interviewers regular debriefings, or giving them time between interviews to participate in less taxing administrative tasks.

Careful selection and thorough training of the interviewers, together with the offer of emotional support throughout fieldwork will help interviewers do their work effectively. Nevertheless, whenever an interviewer feels unable to continue, the field supervisor should be supportive and allow him/her to leave the interviewing team.

The interviewers' personal safety is another aspect to take into consideration when the survey mode is face-to-face. Indeed, the interviewers' personal safety may be threatened if they are required to enter unsafe neighbourhoods or travel alone or if they are in a household with an aggressive household member. The safety of female interviewers may be threatened in locations where it is not customary for a woman to work outside the home or to travel with a man who is not her husband. Proper training must be provided so that the interviewer knows how to act in an unsafe situation, and the field supervisor can confidently assess each situation and deal with it (for example, providing male escorts, emergency telephone numbers and other security mechanisms).

The survey entails a wide range of ethical issues that need to be clearly elaborated during the planning phase and throughout the training and consistently monitored during data collection. Interviewers must be trained not to jeopardise or compromise the confidentiality of the data collected. They must be given clear instructions not to comment on any aspect of the survey work and be made acutely aware of the consequences of a breach of confidentiality. They are also responsible for the protection and safekeeping of questionnaires (for a Papi survey) until they are submitted to field supervisors.

6.2.2 Continuous training

Considering the sensitivity of the survey topic, the interviewer training should be extended beyond the actual training course and into the field implementation phase. Ongoing training of interviewers is important both to maintain interviewers' motivation and to ensure consistent and continued levels of interview quality throughout the entire data collection phase.

Interviewers should be offered opportunities for regular debriefings, group discussions and persons to provide additional assistance and support on the field to effectively address any problem or difficulties that may arise during the work. They should also be offered debriefing with, for example, a psychologist or counsellor and trauma counselling as a group or on an individual basis as needed.

Debriefing aims at having the interviewers' general indication on the survey process from the point of view of content, as well as from a technical and a methodological standpoint and it allows a control to take place on the survey's effectiveness. Debriefings reveal their importance especially at the beginning of the survey, when the survey trend is still to be understood. Over the course of the survey, debriefings can become less frequent, about once a month, and they can be substituted by monitoring and group discussions. During debriefing, researchers solicit the interviewers to communicate and share any problems and difficulties they found when interviewing, related both to technical aspects and to content or relational ones, in order to carry out all the corrective interventions needed on the field.

Interviewer group discussions can be informally carried out during the entire data collection phase with about one meeting per week. They seek to offer support from

the psychological and also from the procedural point of view. Group discussion can be organised as a reference group in which everyone contributes, for example, to getting out of a difficult situation. The enrichment comes from the experience contributions from all participants.

Assistance on the field: when the survey is carried out from central unit, as in a Computer-Assisted Telephone Interviewing (Cati), it is convenient that researchers, trainers and the others responsible for the survey be present in the call centre also to demonstrate the importance given to the phase of data collection.

The survey leaders and their team have various tasks. They can be simply observers collecting impressions, observations, ideas, criticisms, feelings from interviewers and interviewees and recording them down on a notebook or on *ad hoc* module. Other times, instead, they can also take care of the personal relationship with the interviewers or they can intervene in various situations, giving suggestions, correcting the interviewers or expressing other points of view.

The staff involved as supervisors should be sensitised and accurately trained to show a welcoming and collaborative behaviour so that interviewers can see their presence not only as a control, but also as a support allowing them to feel free to ask for help and support.

Trauma debriefings provided for interviewers in the International violence against women survey (Ivaws)

It was recommended to national coordinators in all countries that interviewers in the field should have access to the following three levels of support for dealing with emotional distress:

1. the option of contacting the project coordinator at any time during the fieldwork to ask for assistance or to talk about distressing interview experiences. At least once a week, a counsellor should be available; on the other days, a qualified member of the research staff should be available. The counsellor should be on call during the survey period to provide emergency assistance to interviewers when necessary;
2. group meetings once a week during the fieldwork and as required afterwards;
3. individual meetings with a counsellor as required.

6.2.3 Training methodologies

It is important that interviewers be actively involved during the training, therefore it is advisable to alternate different techniques and methods. This can be done for example through the use of lectures, group discussions and group activities, film showings and/or trailers, testimonials, mock interviews, and role-playing.

All the training methodologies suggested are meant also to stimulate the debate on stereotypes, prejudices and myths about female/male victims of violence that may be widespread in the general population as well as to discuss any bias or stereotypical views that interviewers may have concerning those who are experiencing violence or their decision to leave or stay with their violent partner or to seek help.

A variety of different formats can be used to raise important issues during the training, such as films, documentaries, newspaper articles, statistical data, theories about the causes of violence and testimonials (abused women/men or advocates for abused women/men). Brainstorming and group discussions can also bring up concerns that interviewers

have about survey content, and establish ways to interact with respondents and to manage their reactions to the survey content as well as their own reactions. Holding discussions among interviewers about their own bias during the training is important in order to reduce the chances that interviewers express judgemental comments when interacting with respondents.

Role-playing is an effective way to develop the skills needed to recognise and react efficiently and professionally to challenging situations. It provides opportunities to rehearse a variety of scenarios and to discuss strategies for meeting challenges. It also lets interviewers observe effective strategies in others and to analyse their own reactions and behaviours. Role-playing and mock interviews can be video-recorded and reviewed for comments and 'corrections' together with the group.

Examples of role-playing scenarios should include the following:

- a threatening husband/wife interrupts the interview;
- a respondent has an emotional reaction to a disclosure of rape;
- a respondent reacts negatively to questions about intimate partner violence and wants to end the interview;
- someone in the household - child, teenager or adult - interrupts the interview;
- a selected respondent refuses to participate in the survey on account of lack of time or interest.

6.2.4 Organisational aspects

The training room should be hospitable, comfortable and well-equipped (boards, projector, and depending on the material to be used in the training, personal computers to practice the electronic questionnaire if this is the mode of survey administration, TV and/or video-tape and/or Dvd-player). Of course, each participant will also need a chair and a suitable desk to take notes.

The number of interviewers in each training session should be limited, no more than 10/15, in order to promote active participation and in-depth learning. Groups of 20 persons, however, are still manageable.

At the same time, it is preferable to allocate a few days for the training with no more than five or six hours per day, rather than cramming the whole content into a short time frame. This will facilitate the learning process and give interviewers the possibility to think through and soak up what they have learned during the training.

The proposal of training programme below includes a theoretical part, where objectives and topics of the survey are presented and discussed, and a practical part, in which the questionnaire administration and the management of the relationship with respondents are addressed. This kind of programme should ideally be carried out in at least four days: two days for the theoretical part and two days for the practical one.

Example of training programme

Module A. Introduction

- Introductions, meet and greet
- Interviewers' expectations about their job
- Objectives for the training
- Presentation of the Institute/agency and the team who carries out the survey
- Presentation of the training and work schedule

Module B. Objectives of the survey

- Definitions and types of violence
- Brief historical excursus about VAW and GBV surveys in Europe and worldwide
- Some data from previous surveys on violence
- Information on objectives of the GBV survey and survey peculiarities in the country in question
- What a statistical survey involves: sample, data collection, data processing and data release
- The role of the interviewers in the survey process
- Ethical considerations for the survey research
- Respondents' and interviewers' safety, including legal aspects
- Data collection technique and data collection timing
- Questionnaire structures and sections

Module C. Intimate Partner Violence

- Definitions and examples (use of presentations, readings, movie clips, video testimonials)
- Myths and explanations for domestic violence
- Effects of violence on victims and children witnesses of domestic violence
- Country legislation on domestic violence
- The questions on IPV asked in the questionnaire

Brainstorming and group discussions on the topic must be considered and included in the module

Module D. Sexual violence

- Definitions and examples (use of presentations, readings, movie clips, video testimonials)
- Myths and explanations for domestic sexual violence
- Effects of violence on victims
- Country legislation on sexual violence
- The questions on sexual violence asked in the questionnaire

Brainstorming and group discussions on the topic must be considered and included in the module

Module E. Sexual harassment at work and stalking

- Definitions and examples (use of presentations, readings, movie clips, video testimonials)
- Country legislation on sexual harassment at work and stalking
- The questions asked in the questionnaire

Brainstorming and group discussions on the topic must be considered and included in the module

Module F. The questionnaire

- The Questionnaire's structure and flow
- Each sections' objectives
- The questions' objectives
- Possible errors
- Questionnaire testing in pairs and role playing
- Homework: practice interviews

Module G. Administering the questionnaire

- How to read the questions
- How to code the answers
- The probing technique
- Examples of incorrect and correct interviews
- Questionnaire testing in pairs and role-playing
- Homework: practice interviews

Module H. Interviewer and respondent communication

- Encouraging interpersonal communication and respondent disclosure
- Responding to emotional trauma (providing psychological comfort when interviewing, preventing respondents' and interviewers' possible stress)
- Reasons for refusing the interview and how to deal with refusals
- Frequently asked questions and how to answer them
- Role-playing and exercises in pairs

6.2.5 Training materials

Questionnaire guide

The questionnaire guide is the main tool to help the interviewer understand the questionnaire both in terms of content and procedural aspects. It has to include a complete, easy and operational description of the data collection instrument.

As regards content, the guide should embrace one or more dedicated sections:

– describing the survey's aims;
– defining the types of violence;
– describing the methodologies used to collect experiences of violence (*i.e.* the screening technique);
– presenting each section and the more important questions and answers to each of them.

As regards *procedural aspects*, the guide should include complete information about:

– period and timing of the interviews;
– the survey technique (Papi, Capi, Casi or Cati);
– the reference sample;
– the survey units;
– the procedure to individuate the reference person, if any (the reference person is the person that answers the questionnaire about him/herself);
– how to administer the questionnaire. The quality of collected data depends on the carefulness and methodological rigour with which the questionnaire is administered. Therefore, the guide has to include instructions on how to read questions, the probing technique and, in the case of multi-response questions, on how to code spontaneous answers.

Interviewer's manual

This manual is intended as an aid where relational and communication aspects of the interview are addressed. It should provide indications and suggestions to help in problematic

situations that can occur at any phase of the interview. The interviewer manual should also include information on the background of the project and measures to address the ethical and safety issues associated with conducting interviews on violence.

This kind of didactic aid should provide a clear definition of the role and specificity of the interviewer's work through the use of examples. At the same time, it should offer tools and strategies useful for facing different situations and possible difficulties that can occur at any phase of the interview: how to efficiently introduce the interviewer, the survey and the Institute responsible for conducting the survey; how to encourage the respondent to participate, which information should be provided and which tones and arguments should be used; how to respond in difficult situations.

The manual should also include a list of frequently asked questions, according to previous field experiences, allowing the interviewers to effectively answer to any question or doubt the respondent may have.

A special section should be dedicated to the specific problems of interviewing a victim of violence. As speaking with victims poses specific problems related to their personal safety, the manual should indicate and recommend all of those tools to decrease such risks and help in reassuring the interviewee: the possibility of interrupting the interview at any moment, taking another appointment, or terminating or changing the subject of discussion if an interview is interrupted by anyone; the possibility of suggesting different phone numbers or places that are more suitable for such an interview, the possibility providing references of shelters, besides the need to provide particularly clear and detailed information about privacy and data confidentiality.

Other training materials

Some other materials should be distributed and/or illustrated during training: power point presentations; movie clips and video testimonials on violence to be used as an introduction to the subject for eliciting discussions; written exercise; short scripts to be used in role-playing; examples (recorded and/or videotaped) of "correct" and "incorrect" interviews.

Support services for the respondents

Prior to conducting the research, it is important that researchers meet with potential providers of support, which may include existing health, legal and social services and educational resources in the community and less formal providers of support, such as shelters and anti-violence centres. A list of governmental, private or not for profit resources and facilities should be prepared and offered to respondents in case of need or upon the respondents' request.

This list of services should be organised according to geographical area and may include a broad range of social services designed to prevent or treat domestic or non-domestic violence: counselling centres, emergency services, 24-hour hotlines, women's crisis centres, special police units or task forces supporting victims, legal advice services for victims.

In face to face interview can be used a simple list or a specially designed flyer or a brochure and it can be also offered to all the respondents, regardless of whether they need or ask for help.

When the survey use a Cati, the interviewers should offer to the respondent the name, address and the telephone number of the support services located in the respondent's area of residence only in case of need or upon the respondent's request.

What it is important to underline here is the interviewers' role: they don't have to push the respondents in any way. To take or not the flyer or the telephone number of the support services, to ask or not for help is up to them, it is their free choice. Any pressure done by the interviewers can be perceived as a control behaviour and can even give rise to doubts about the interviewers' true purposes.

Conclusion

In brief, the training principally aims at making interviewers:

- aware of the importance and sensitive nature of the survey;
- aware of the content of survey;
- able to follow the correct interview methodology;
- able to create a climate that facilitates disclosure by the respondent;
- able to manage potentially critical situations.

Another objective is to build team spirit and motivate interviewers to cooperate and support each other. Indeed, motivation is fundamental for the entire research process and particularly during data collection, because it increases the ability to tackle and solve problems. A deep knowledge of the survey content and mode of administration, coupled with interest and passion for the research, are prerequisites for carrying out the work successfully and engendering a sense of cooperation and co-responsibility.

6.3 Ethical considerations and support services for respondents

Population surveys, as well as research in general, are undertaken to expand knowledge, to provide evidence for policy-makers, legislators, practitioners and for others data users. Appropriate scientific methodologies have to be applied in all research to derive valid results about the phenomenon being studied, and those results have to be reported honestly. Therefore, scientific as well as ethical standards must be respected. Ethical principles require that any research involving human subjects is framed and conducted in a way

Recommendations regarding the ethical conduct of domestic violence research (Who, 2001)

The safety of respondents and the research team is paramount, and should guide all project decisions.
Prevalence studies need to be methodologically sound and to build upon current research experience about how to minimise the under-reporting of violence.
Protecting confidentiality is essential to ensure both women's safety and data quality.
All research team members should be carefully selected and receive specialised training and ongoing support.
The study design must include actions aimed at reducing any possible distress caused to the participants by the research.

Fieldworkers should be trained to refer women requesting assistance to available local services and sources of support. Where few resources exist, it may be necessary for the study to create a short-term support mechanism:

- Researchers and donors have an ethical obligation to help ensure that their findings are properly interpreted and used to advance policy and intervention development;
- Questions on violence should only be incorporated into surveys designed for other purposes when ethical and methodological requirements can be met.

that respects the human rights of the individuals concerned. For this reason, regulatory frameworks, guidelines and guidance for ethical procedures on scientific research have been constructed worldwide over the past decades.

There are research topics or target populations (such as children, prisoners, marginalised populations or the socioeconomically disadvantaged) that require particular attention in developing survey guidelines on ethical behaviour of the interviewers and ethical procedures to be followed. One of these topics is violence and, in particular, gender-based violence.

Not so many years ago, it was often felt that domestic violence against women is too sensitive a topic to be explored in a population-based survey. However, several studies show that research on domestic violence can be conducted with full respect of ethical and safety considerations and evidence suggests that many women find that participating in violence research is beneficial. Nevertheless, research on violence against women raises important ethical and methodological challenges in addition to those posed by any research. The nature of the topic means that issues of safety, confidentiality and interviewer skill and training are even more important than for other areas of research.

In order to guide research in this area, the World Health Organization (2001) has developed some recommendations regarding the ethical conduct of domestic violence research. Even if these recommendations are stated to apply specifically to the collection of information on domestic violence against women, they also apply to ethical and safety issues associated with gender-based violence surveys against both women and men.

There is the danger that a well-intentioned but poorly conceptualised or implemented study may result in a serious under-reporting of violence. This raises both ethical and practical concerns.

Ethically, it is unacceptable to conduct a poorly designed study on violence against women, where women are asked to disclose difficult and painful experiences and where the nature of the subject matter may put women at risk.

Good morning /good evening, this is «Name and Surname», I am an interviewer and I work for Istat, the National Institute of Statistics. We are carrying out a survey on personal safety.

In recent days, you should have received our letter, addressed to Mrs. … saying that we would be calling for a telephone interview. Have you received it?

In the letter, we wrote that Mrs. … was randomly selected to take part in a survey conducted throughout Italy. The information collected will be used to find out more about the living conditions and the safety of people in the places where they live.

Am I speaking with Mrs. …?

IF THE RESPONDENT IS SUSPICIOUS OR IN DOUBT

If you want to check for yourself to verify that this survey actually exists or if you want more information, I can provide the Istat toll-free number and you can check the reliability of this sur-

vey. Your call will be answered by people able to give you all the information you need.

You can call the Toll-Free Number from Monday to Friday from 9:00 a.m. to 9:00 p.m. and Saturday from 11:00 a.m. to 5:00 p.m.

Furthermore, you can check if this is really the Istat toll-free number by calling the Istat switchboard in Rome: 06 … or in one of the regional offices of the Institute.

Practically, bad data may be worse than no data, because low prevalence estimates could potentially be used to question the importance of violence as a legitimate area of concern. Consequently, it is important that domestic violence surveys are methodologically sound and build upon current research about how to minimise under-reporting. Sound professional and scientific processes require established statistical sampling methods for drawing samples, unbiased questions that have been tested before their implementation, adequate training of the interviewers and other staff. Moreover, it is essential to provide data users enough information about the survey and its methodology to enable them to understand the limitations of the data as well as the strengths.

Safety of respondents

It is important that the survey is not introduced to the household and wider community as a survey on violence. Instead, at this level, the study should be framed in a different manner - such as a study on the living conditions, welfare and safety of people or a study on people's life experiences or family relations. However, the respondent has to be fully informed about the nature of the questions.

Below is a hint from the Italian survey on violence against women, on how the interviewer could introduce herself and the survey when calling or arriving at the household. The example is from the beginning of a Cati interview but it can be easily adapted to a face-to-face interview.

In instances where the sampling unit is the household, only one respondent per household should be selected randomly and interviewed about his/her experiences of violence.

Interviews should be conducted only in a private setting. Face-to-face interviews (Capi, Papi or Casi) are generally conducted at home, nevertheless, in some situations it is advisable to ask respondents if they prefer other places to be interviewed.

The participants should feel free to reschedule or relocate the interview to a time or place that may be more safe or convenient for them. Rooms of the Municipality, libraries, bars, a park, the interviewers' car or other public spaces can be good locations if interviewer and respondent feel at ease and free to talk without being heard or interrupted. If necessary, Government organisations also can be asked to provide private rooms for conducting interviews.

When the interview is conducted in the context of a household survey, interviewers should be trained to terminate or change the subject of discussion if an interview is interrupted by anyone – including children. Indeed, a short diversionary questionnaire on a less sensitive topic should be developed to assist with this situation. For Capi interviews it should be possible for interviewers to easily close the questionnaire or the laptop, or go to less sensitive questions.

Safety of interviewers

The safety needs of the interviewers should also be considered. In some circumstances it could be necessary for interviewers to travel in pairs, to have mobile phones and to use designated means of transport/drivers. In the case of women interviewers, a trusted male escort for certain neighbourhoods known to be unsafe can be required.

In forming teams of fieldworkers in face-to-face interviewing situations, research project managers must take into account local norms that may prohibit women from working in public spaces and other barriers that female interviewers may face when approaching households in order to obtain the interview. In these situations, teams of male and female interviewers working in tandem have been shown to improve household contact and to lower refusal rates; they may also be necessary to ensure the safety of the female interviewers (UNODC and Unece, 2010).

Furthermore, being involved in violence research is a very emotional experience. It could be hard to estimate the actual emotional effect on the interviewers and the burn-out risk, but this aspect must be considered during the selection and the training of the interviewers and all along the data collection phase.

Protecting confidentiality

The confidentiality of information collected during a survey is of fundamental importance. All interviewers should receive, during the training, strict instructions about the importance of maintaining confidentiality.

The potential respondents need to know how privacy and confidentiality will be maintained during all phases of the research project and for all information collected.

The research findings and the information presented in any format of research reports need to be sufficiently aggregated to ensure that no single community or individual can be identified.

The interviewer has to explain the procedures and obtain an informed consent by the respondents whenever applicable.

No names must be written on the questionnaires and no interviewers should conduct interviews in their own community. To protect the respondents' confidentiality and anonymity, the interviewers must inform the supervisor should they know any of the people they will be interviewing. This is critical (regardless of whether the interview is by telephone or face-to-face) and, should this be the case, those interviews will have to be assigned to other researchers. When face-to-face interviews are scheduled in very small municipalities, where people practically all know each other, interviewers must be called in from neighbouring municipalities and they also have to verify that there are no known people on the list. If having checked for acquaintances in the list, interviewers still encounter someone they know, for example someone who got married and changed last name, they have to tell to the respondent the reason why they cannot interview him/her and inform the field supervisor who will provide to reschedule the interview with another interviewer.

Reducing participants' distress

All questions about violence and its consequences should be asked in a supportive and non-judgemental manner. Particular care needs to be taken to ensure that the language of the questionnaire cannot be interpreted as being judgemental, blaming or stigmatising.

Interviewer training should include practice on how to ask the questions, how to support the respondents without becoming a psychotherapist, how to terminate an interview if the impact of the questions becomes too negative. All interviews should be ended in a positive manner; reinforcing the respondents' own coping strategies and reminding them that the information they shared are important and will be used to help other persons (Parker and Ulrich, 1990).

Fieldworkers should also be prepared to provide a list of available local services and sources of support for those respondents requesting assistance. In face to face interview can be used a simple list or a specially designed flyer or a brochure and it can be offered to all the respondents.

When the survey use a Cati, the interviewers should offer to the respondent the name, address and the telephone number of the support services located in the respondent's area of residence (and possible other information about opening days and hours) or, if it exists, a toll free number (a help line) for victims of violence, only in case of need or upon the respondent's request.

What it is important to underline here is the interviewers' role: they don't have to push the respondents in any way. To take or not the flyer or the telephone number of the support services, to ask or not for help is up to them, it is their free choice. Any pressure done by the interviewers can be seen as control behaviour and can even give rise to doubts about the true purposes of the survey.

Ensure findings' proper interpretation and use

It is a best practice to involve NGO and the experts of the topic that took part to the planning phase, to review and interpret the first survey results.

It is important that research findings are used for policy making, prevention and intervention activities. It is also important that the study community receive early feedback on the results of the research in which they have participated.

Researchers need to be pro-active in helping to ensure that research findings are interpreted appropriately by the public and the media. Although analysis may highlight sub-groups at higher risk of particular forms of violence, researchers should take care when presenting such findings not to feed into any negative stereotypes of particular ethnic or social groups, and to ensure that no single community or individual can be identified or stigmatised.

REFERENCES

Barletta, R., M.G. Muratore, and L. Scarnicchia. 2017. "Report of the Study Visits: Spain, Norway, Estonia". *Working paper*, prepared for the Eurostat Task Force Meeting. Luxembourg, 27-28 June 2017. https://circabc.europa.eu/sd/a/a8992f8d-d8bd-46b2-a07b-a23357f29ab7/Study%20visits_Report%20.pdf.

Claramunt, C. 1999. "Helping Ourselves to Help Others: Self-Care Guide for those who Work in the Field of Family Violence". *Gender and Public Health Series*. Washington, DC, U.S.: Pan American Health Organization – PAHO.

Ellsberg, M., and L. Heise. 2005. *Researching Violence Against Women: A Practical Guide for Researchers and Activists*. Geneva, Switzerland, and Seattle, WA, U.S.: World Health Organization - WHO and PATH.

European Institute for Gender Equality - EIGE. 2017. "Quality Criteria of the EIGE´s Gender Statistics Database". *Experts' Meeting on EIGE´s Gender Statistics Database*. Vilnius, Lithuania, 7 November 2017.

Jansen, H.A.F.M. 2010. *Swimming against the Tide: Lessons Learned from Field Research on Violence against Women in the Solomon Islands and Kiribati*. Suva, Fiji: United Nations Population Fund - UNPFA, Pacific Sub-Regional Office.

Johnson, H., N. Ollus, and S. Nevala. 2008. *Violence Against Women. An International Perspective*. New York, NY, U.S.: Springer.

Parker, B., and Y. Ulrich. 1990. "A protocol of safety: research on abuse of women. Nursing Research Consortium on Violence and Abuse". *Nursing Research*, Volume 39, Issue 4: 248-250

United Nations, Department of Economic and Social Affairs, Statistics Division. 2014. "Guidelines for Producing Statistics on Violence against Women - Statistical Surveys". *Economic & Social Affairs*. New York, NY, U.S.: United Nations Publication. https://unstats.un.org/unsd/gender/docs/guidelines_statistics_vaw.pdf.

United Nations Office on Drugs and Crime - UNODC, and United Nations Economic Commission for Europe - UNECE. 2010. *Manual on Victimization Surveys*. Geneva, Switzerland: United Nations.

World Health Organization - WHO. 2001. *Putting Women First: Ethical and Safety Recommendations for Research on Domestic Violence against Women*. Geneva, Switzerland: WHO.

7. THE QUALITY OF THE STATISTICAL PROCESS[1]

7.1 Strategies to reduce refusals

Realising good quality standards is based more on preventing errors than on correcting data ex-post, also because the correcting phase is more burdensome, requiring longer times and higher economic resources.

Among different non-sampling error components, non-response is an important one. Being as close as possible to the theoretical sample is a basic instance to assure a representative sample, but considering as inevitable a minimum of non-response percentage, specific strategies to prevent it are to be applied.

Considering the eventual total non-response, it is to be stressed the importance that the selected individuals are actually interviewed. Non-response can occur for a variety of reasons, including outright refusal to participate, inability to participate owing to language difficulties, illness or impairment and inability to contact selected respondents. Unit non-response will affect the representativeness of the sample because those who are not immediately available for an interview or who refuse to participate generally differ from those who agree to participate on characteristics that are relevant to the topic of the survey.

The partial non-response occurs when the questionnaire is only partially filled, that is, some questions or even entire sections are omitted or ignored. Partial non-responses occur more often in paper and pencil surveys, especially with a self-administered paper questionnaire, while when the interview is computer assisted they are less common, as the software does not allow the continuation of the interview without recording an answer to each question. Reasons for non-response are various, due to the interviewer or to the interviewee who do not respect the questionnaire pattern, or even due to lack of attention, or to the actual difficulty in following a complex and too structured or poorly structured questionnaire.

Many different strategies can help to prevent total or partial non-response: are contact rules, times to interview, call-backs or return visits to the sampled households, training of interviewers, introduction to the questionnaire, structure, sequence, wording and length of the questionnaire and the communication strategies aimed at motivating the selected individual and the family.

Indeed, it is important to establish a communication channel between the Institution carrying out the survey and the selected respondents, before the data collection phase start. General campaign against violence before the survey, for example, can create a positive climate and encourage participation. Large scale communication to inform population about the survey, instead, are discouraged because can expose the respondents at risk or alarm them.

The introductory letter and the toll-free number for respondent are the most usual communication strategies adopted.

1 This chapter has been drawn up by Maria Giuseppina Muratore

7.1.1 The introductory letter

The introductory letter is the first communicating tool with the respondent. Many studies and experience on the field have widely demonstrated that the introductory letter produces a positive context with the respondent, preparing him/her for a more effective cooperation: the interviewer's work is easier and the interview is more easily obtained, since much information has already been obtained from the respondents; the real focus of the interview can be more easily and quickly proposed and investigated.

It is important that the introductory letter be presented as an official letter, marked by logo of the Institute carrying out the survey, and signed by the Institute's President or Director. When the survey is managed by more than one Institute, for example the National Statistical Institute and other territorial Public Administrative Offices, it is better that the letter be signed by all or at least more than one of them. The involvement of territorial agencies can be important, as it is closer to people, making them more confident and helpful for the survey.

The introductory letter represents a preliminary introduction to the survey and the Institute carrying out it or the Agency promoting it. In this way, respondents are informed about the existence of the research project, its motivations and aims; explanations about how respondents were involved in the project are also provided, and reasons why the respondents' cooperation is so important are also highlighted. Respondents are to be reassured about the seriousness of the project, privacy and how respondents' data and information will be treated. Obviously, for safety reasons, it is important not to mention the word violence nor to reveal the true content of the survey.

Regarding arguments and questions that will be faced during the interview, it is important to give some general information about the main subject of the survey, without entering deeply into detail, especially if the addressed subject is so sensitive and personal, otherwise the respondent would be unnecessarily on guard, and consequently he/she could decide not to give any information.

Respondents are to be told how he/she will be asked to participate: if by a self-administered questionnaire or by a telephone interview or by face-to-face interview. Days and hours for interviews will be then indicated.

In many countries the collection and processing of personal data are protected by some legal rules. To assure clearness and completeness of information for the respondents, in the introductory letter, the main legal and normative references concerning response obligations and privacy have to be included. Respondents have to be informed regarding who will process their data, the possibility to access and modify their own personal data, and they are to be assured about the fact that all people involved in the survey have to ensure confidentiality.

Considering that the introductory letter should not to be too long, otherwise there is the risk that the citizen does not read it at all, here is a summary of the main points to be included:

- the institute logo;
- the presentation of the Institute carrying out the survey;
- the presentation of the survey and its main aims;
- the survey theme;
- motivations of the household/individual involvement in the survey;
- explicative lines for the involvements (type of interview, days and hours of interviews);

- references to normative about obligations to answer and the protection of privacy and personal data;
- the toll-free number, with hours indicated to call it;
- the further landline (not toll-free) phone number of the Institute;
- thanks;
- Institute President signature.

Due to the importance that the citizen, and overall the selected person, reads the letter and remembers it until he/she is interviewed, specific care must be taken to facilitate the reading of it. For making the letter more interesting, graphic aspects, different characters underlining different points and the most important aspects, different colours both for characters and for the logo, can be used. Normative indications, that could be distracting and difficult to read, can be framed and added at the end of the letter.

If the letter is addressed to the household, or if all the components have to be interviewed or the person to be interviewed is yet to be selected, it is important to highlight that this information should be spread to all the family members, suggesting to place the letter somewhere in the home where it can be easily seen.

Dear Madam,
this is to inform you that, together with 25,000 other people aged 16-70, you have been selected to participate in important statistical research, which the National Institute of Statistics (Istat) will conduct on the living conditions, welfare and safety of people in Italy.
This research is included in the National Statistical Program 2021-2024 - converted with amendments by Law no. 125 of 30 October 2013, and in the National Statistical Program 2023-2025 currently in the process of being approved (code IST-02260), which include all the statistical surveys that are deemed necessary for the country. The National Statistical Program is accessible on the Istat website at http://www.istat.it/it/istituto-nazionale-di-statistica/organizzazione/normativa.
Your collaboration is essential for understanding what is happening in the country and to its inhabitants.
Between May and October 2024, a person appointed by us will come to your home for an interview/will call you for a phone interview. We therefore ask you to keep this letter until you have the interview or, at the latest, until the end of October 2024.
Istat is required by law to carry out this research and the citizens' obligation to respond is prescribed by art. 7 of Decree Law no. 322/1989 and by the D.P.R. of 19 July 2013. However, as foreseen by current legislation, you may not answer some questions of a sensitive and judicial nature in sections B, C, D, E, F, G and in the Reports 1, 2, 3 and 4 of the questionnaire. The interviewer will flag each one of these questions. The complete list of questions, which you are allowed not to answer, is published on the Istat website.
Law guarantees the confidentiality of all information you will provide us.
The data collected, which is protected by statistical confidentiality and subject to the legislation on the protection of personal data, may be used by employees of the national statistical system also for subsequent processing, exclusively for statistical purposes. Furthermore, the data may also be communicated for scientific research purposes under the conditions and according to the modalities foreseen by art. 7 of the Code of Ethics for the treatment of personal data within the national statistical system. The same data will be disseminated in an aggregated form, so that no data can be attributed to the subjects who supply them or to whom they refer.
The Central Director of the Socio-demographic and Environmental Statistics of Istat is responsible for the statistical processing of data as well as for the processing of personal data. Please refer to the Central Director if you wish to exercise your rights about data collected.
During the interview period (April 2024-October 2024), for any further clarification please contact the toll-free number 800.046660, from Monday to Friday from 9:00 am to 9:00 pm and

on Saturday from 11:00 am to 6:00 pm or Istat's "Social structure and dynamics" service at 06/4673… or 4673 …..
We take this opportunity to remind you that information regarding Istat organisation and activities is published on the website http://www.istat.it. It is also possible to obtain information on the activity of the Institute by contacting the Statistical Information Centres active in the regional capitals.
Thanking you in advance for your cooperation, I welcome the opportunity to send my best regards.
Istat President signature
statistical confidentiality. response obligation.
protection of confidentiality and data subject rights
References to National Legislation and other official documents can be included here.

7.1.2 The toll-free number

Providing a helpful free phone number (toll-free number) allows respondents to contact the national statistical agency before, during or after the interview if they have questions about the survey or want to confirm the legitimacy of the survey. This tool can substantially improve response rates.

For this purpose the toll-free number should be activated soon after the mailing of the introductory letter, so that the citizen can communicate with the national statistical agency as soon as he/she has received it.

The toll-free number has to provide a good image of the Institute in addition to solving any doubts and problems of the interviewee to make him/her cooperate). If staff responding to the calls are very skilled and highly trained about the aims and content of the survey and are effective communicators, they can influence the willingness of callers to participate. In case citizens call to refuse, it is important that the operator try to convince the respondent by motivating him/her.

Furthermore, researchers have to be careful in detecting, from the calls to the toll-free number, the moods, complaints and doubts of the citizens. The flow of information from the calls to the toll-free number is very precious; such information can be recorded by a structured card collecting reasons for the call, caller profile (sex, age, education level, professional condition, residential region) and the operator's personal opinion regarding how the call was ended, in particular about the individual's willingness to be interviewed. Monitoring such information, day by day, makes it possible to control the situation, promptly detecting eventual negative situations that could invalidate the quality of the survey.

Very often citizens call to ascertain the seriousness of the survey and to have general information or to stress the unwillingness to participate or to communicate a change of address, or households' moving house or living for holidays, to state when someone of the family is at home to answer, or the mobile number if they prefer to be called there, or to ask for an appointment, or even to communicate the death of the person to whom the introductory letter is addressed. Also in these cases, the content of the survey needs to remain vague for respondents that have not yet started the interview.

The overall strategy joining the introductory letter and the effects of the toll-free number produces a substantial increase of response rates of the selected individuals. Indeed, such tools make it possible to involve citizens from the beginning of the data production process.

Indicating in the introductory letter a second not free phone number, whose reference can be found on the public phone list, will give citizens a higher assurance about the seriousness of the survey. Sometimes the most suspicious citizens think that the toll-free number can be a tool to mask an organised structure to carry out fraud, using personal data. A not free phone number that can easily be found with references of the Statistical Authority on the public phone list can reassure citizens that the survey is in fact institutional, serious and genuine.

It is quite important to monitor the calls received at the toll-free number. It represents a valuable opportunity to gain additional information about the quality of the interview process and it is recommended that the operators of the toll-free number record important information that can be used for quality control purposes.

The sheet can be organised in different sections, each of which summarising the development of the telephone dialogue: information about the person who called (name, address/phone number, socio-economic variables, reasons for calling as to verify information and institute carrying out the survey; having information on the survey; giving communications about availability (day and time preferred for the interview, new phone number/address); eventual problems in being interviewed, other informative notes.

7.1.3 Scheduling the mailing

The interval between the arrival of the letter and the moment when the person will be contacted to be interviewed should not to be too long, as the more time passes the less memory is retained about the received letter. The interview should take place no more than one week/15 days after the arrival of the letter. For this reason, it is important to schedule the mailing in various and different groups. The number of letters to be sent in each group is determined by considering various aspects related to what is called 'work time' of the selected households: length of the interview; number of interviewers, days and hours of the interviews and so on. An initial estimate of the number of those addressed that are closed each day (contacted and interviewed or substituted) is essential for planning the timing to send the groups of letters.

To verify the interval between the moment of sending the letter and its effective arrival, it can be useful to send some letters to people of the survey staff or known people not belonging to the sample, possibly located on different territorial areas to cover as many areas as possible. So it is possible to verify – with a small sample of areas – the correct sending of the letters, especially if mailing is made by an agency other than the Institute carrying out the survey and automatically, so as to control that the various steps are followed according what has been established.

In case the sample design foresees substitutions, letters have to be sent to substituted respondents. Sometimes more than one letter can be sent in order to increase the response rate.

Incentives and small gifts can also be useful in order to reduce refusals, but in case of surveys carried out by interviewers, their impact is not important (Singer, Ye 2013) and their use is also expensive.

7.1.4 Brochure with key statistics and information notes for the respondents

In some countries, an informative brochure with key statistics is prepared for citizens to maintain interest in participating in the survey (in particular for longitudinal surveys) or procedures facilitating access to publicly available information (for example, on a website, a guide to complete the questionnaire or helpline particularly for self-administered surveys) or a letter thanking them for their participation. These measures will help to sensitise the units selected in the sample to participate in the survey.

7.1.5 Find the best time and location for the interview

Interviewers have an important role to play in encouraging participation by creating a comfortable environment for respondents, presenting the survey objectives and benefits clearly, reassuring the respondents that their confidentiality will be strictly guarded, and showing flexibility in scheduling or rescheduling an interview at a time and place that is convenient for the respondents.

Determining the best time to call or to visit survey units is based on paradata acquired during previous iterations of the survey or from a similar survey. It is in fact important to manage calls or visits in such a way that respondents are contacted at the best time for them and that the number of attempts to call or visit does not exceed a useful maximum.

While for Cati household surveys it is generally preferred to call in the afternoon and in the evening, when it is more probable to find someone at home, in the case of a GBV survey it is suitable to call the respondents, especially in the case of women, also in the morning. Women indeed, can be at home alone while children are at school and her partner is at work.

To call very early in the morning is not very efficient, but it is important to give the chance to the interviewers to make an appointment also in the morning if the respondent prefers so. In addition, Saturday morning timetables are fruitful.

Face-to-face (Capi and Papi or Casi) surveys are generally conducted at home. Nevertheless, for GBV survey, in some situations, it is preferable to ask respondents if they prefer other places where to be interviewed. In the Italian VAW survey, for instance, that is face-to-face for migrants, some interviewees asked to be interviewed outside. In these cases, the interviews were carried out in rooms of the Municipality, libraries, bars, a park, in the interviewers' car or in other public spaces.

7.1.6 Contact rules

It is important to determine how many contacts to make at each sample unit. An effort must be made, through call backs or return visits to the sampled households, to ensure that groups typically overrepresented among non-responders or reluctant participants are not underrepresented in the survey.

The contact with the household or the individual to fill the questionnaire is different according to the interview mode used, as well as the interviewer's role, if the interview is face-to-face, by phone or self-administered. In any case, a contact has to be established between the respondent and the interviewer. How to establish such contact differs, depending on the chosen interviewing mode. In a face-to-face or self-administered questionnaire,

the interviewer has to go to the respondent's house (with the exception of Cawi), and that involves an important amount of time and costs; when considering a phone contact, obviously, time and costs are a great deal lower.

At this step, the quality control is focused mainly in minimising non-response and non-contacts/no answer; the effort has to be concentrated on obtaining the interview and, when the respondent is collaborative, on managing and concluding it. In terms of efficiency, it is recommended to realise the following three phases of the interview, obtaining the interview, keeping it, concluding it, in only one contact with the respondent. Nevertheless, there still persists a percentage of interviews that, for various reasons, it is not possible to conclude during only one telephone call/one visit, sometimes they cannot even start.

In the case of the Cati technique, a higher percentage of interviews are concluded during the first two contacts, while the percentage of concluded interviews in following contacts after the first two decreases more and more. After a specific number of contacts, it is more efficient for the overall economy of the survey to abandon that household, working the following scheduled nominative instead of insisting on the present one. To evaluate the maximum number of contacts for each unit, a costs-benefits evaluation has to be made: if, on the one hand, there are quantitative and qualitative aims related to the number of filled questionnaires to be obtained and the sampling rules to be attended, on the other, side costs, budget and time constrains are to be faced and considered (Fuchs, Bossert, Stukowski 2013). As a consequence, choices are to be made and general rules have to be defined that will be adopted by all interviewers.

To establish effective and useful rules it is important to analyse the reasons why an interview is not completed in a single contact and sometimes even not started.

First of all, depending on the sample design, the interviewer needs to ascertain if the respondent unit is in the target (and he/she does not correspond for example to a company, a shop, a professional study) and to ascertain if the respondent unit is the right one (resident or usually living in that house). This issue is even more complex if the sample is a household sample (where all individuals in the household have to be interviewed).

Second, it should be noted that each trial does not always translate in a contact with a family. In face-to-face interviews, a maximum number of visits to the family need to be fixed, while in case of phone interviews, the maximum number of calls to contact the family needs to be higher. We suggest that at least 10 attempts have to be made before quitting the number that means, depending on the timing of the rules of the calls, to try to interview the respondent for at least 3/4 days.

However, the concrete situation is very complex in case of the Cati survey, when a lost call has different meanings: no one is at home, persons are at home but no one answers or the number is busy or it is an answering machine/fax. Usually rules are structured, such as the type of non-contact and the following actions to be carried out; for example, if there is no answer at the phone, it is likely that nobody is at home, so it is recommended to re-call after one hour at least, if the line is busy, it is likely that someone is at home using the phone, so it is useful to re-call after a few minutes (5/10 minutes). Therefore, it is not possible to define a priori a constant number of trials to be made, as it depends on the type of non-contact.

If, on the contrary, a contact is established with someone, various different situations could arise. The person could be out of the target, it could be someone declaring the family is untraceable or that the family members are all severely sick or deceased. On the other hand, if the targeted person was contacted, he/she could be not the one to be interviewed

(this can happen for sample designs where the respondents' selection is made during the call/visit and not a priori), or he/she is the respondent unit but he/she wants to refuse, or he/she starts the interview but definitely stops it after a short time; he/she could also ask for an appointment or finally, at the best, could carry out the entire interview.

Rules for Cati surveys: Suggestions from the Italian Violence against Women Survey

A) RULES FOR THE MANAGEMENT OF TELEPHONE CONTACTS

- The composition of telephone numbers, if possible, must be managed automatically without the manual intervention of the interviewers;
- telephone contacts and any substitutions must be managed automatically;
- the outcome of the calls with no answer, answering machine, fax, busy, must be assigned automatically. Telephone / fax, answering machines must be treated as unresponsive.

B) RULES FOR THE MANAGEMENT OF THE APPOINTMENTS

In the case that a telephone contact has ended with an "appointment", except in cases when the interview has already started (in this case, there are other rules), the management of subsequent appointments and the number of attempts to be made before replacing the person to be interviewed, will follow the rules below:

B.1 - 4 appointments for the family (called "family appointments" because the interviewee has not yet been selected or when there has not yet been a direct contact with him/her) and 5 appointments on the selected respondents are possible.

B.2 - In the case of the call resulting in a non-contact (nobody answers, busy, fax, answering machine) between one appointment and another or between an interruption and the resumption of the interview, the call must be repeated in a different time band or on different days, according to the rules of contactless outcomes.

B.3 - In particular, if after the appointment there is a result of no contact (no answer - NR, answering machine, fax), at least 4 NRs and 4 sets of busy signals must be completed before the call is dropped.

B.4 - In case the interview has started (starting point has to be decided, generally refers to questions regarding the core variables) and an appointment is necessary, the interviewee must be called up until the interview is completed.

In particular situations in which the interview cannot be completed (for example, after the interview has already been resumed for at least 5 times), it will be necessary, before giving up, to submit the case to the researchers' managers.

C) RULES FOR THE SUBSTITUTION OF SAMPLE UNITS

The substitution of a sample unit will occur only if the following conditions are met: wrong telephone number; prolonged non-availability of the name; interruption of the interview; refusal of the interview.

C.1 - A unit is considered untraceable when it is not possible to establish a contact after at least 10 non-response results, as defined in the following rules:

C.1.1 - In the event of a NR outcome, symptom of the absence of the family from the home, it is better not to insist and to contact the family in another time slot or another day. If, for example, a family is called at 5:00 pm and is not available, it will be called again in the next band, or after 7:00 pm. The outcome of the answering machine or fax are automatically considered as non-response (NR).

The call time of the midweek days has been divided into four bands:

1. 11:00 am - 1:59 pm;
2. 2:00 pm - 4:59 pm;
3. 5:00 pm - 6:59 pm;
4. 7:00 pm - 9:00 pm.

An attempt with the result of NR in the first band can only be repeated from 2:00 pm, while an

attempt in the fourth band can only be tried again the day after between 11:00 am and 1:59 pm. On Saturday, the call time has been divided into three bands:
1. 11:00 am - 1:59 pm
2. 2:00 pm - 3:59 pm
3. 4:00 pm - 6:00 pm

C.1.2 - The outcome of OC, which instead mainly implies a presence of the family at home, suggests trying to place the again at a close time interval. The call will then be repeated after five minutes for at least seven consecutive times.

In the event that after seven attempts the number is not yet available, the family will be counted 1 NR outcome, which will follow the rules of the own call back of the "no answer" (the number is, therefore, called in another band).
If the sequence of seven consecutive busy signals is interrupted by a non-response (NR) from that moment on, the call follows the NR rule.
If a busy number becomes NR and subsequently it is busy again, it starts a new series of seven consecutive busy signals according to the rules.
As for the non-response results, the unit will come out of the recycling not before ten non-response results. That is, if the number is always busy, this will come out of the recycle after ten sets of busy signals that correspond to ten NR.

Reasons for non-response should also be recorded. These can include failure to find the selected respondent at home despite repeated attempts, outright refusal to participate, difficulty communicating because of language and the existence of logistical problems or security issues that prevent the inclusion of certain locations in the interviews. A complete accounting of the outcomes of all contacts with sampled households in terms of status as respondents, refusals or ineligible units is essential for calculating the weights for the survey data.

It is clear how different outcomes can be obtained in each contact trial and different specific rules to manage them are to be defined for each of them.

Contact rules can change according to specific requirements related to the subject to be surveyed, to technical reasons such as the software structure managing the calls and to the overall survey organisation. For the Gender-based violence survey, it is recommended not to limit the number of possible appointments due to the particularly sensitive subject to be faced, so as to give to the respondent the possibility of finding the best moment to freely and safely speak about it, for example, when they are alone at home.

In a Cati system of interviews, the most complex system, it is useful to adopt the calls' results, and the indicators, suggested by the AAPOR international standards. This is useful in the monitoring phase, but also for evaluating the overall quality of the survey at the end of it, and to assure the comparison of the results with other data sources.

7.2 Possible outcomes of a phone interview

Possible analytic and synthetic outcomes of a phone call are indicated below (Table 7.1). They can be different according to the different sample design, and for face-to-face surveys the analytic structure is somehow less complex, but not so different (Table 7.1a).

Table 7.1 a – Analytic outcome and synthetic outcome (a) of phone contacts

Synthetic outcome	Analytic outcome
1 - Completed	1 = Completed
	2 = Stopped but valid
2 - Stopped	3 = Definitively stopped
3 - No-response	4 = Refusal before the question about the introductory letter arrival
	5 = Refusal after the question about the introductory letter arrival
	6 = Refusal of selected person
	7 = Refusal of respondent to involve selected person
4 - Provisional outcome without contact	8 = Not answer
	9 = Busy
	10 = Answering machine
	11 = Fax/Data
5 - Appointment	12 = Appointment with family
	13 = Appointment as selected is temporarily not present
	14 = Appointment as selected is temporarily not available
	15 = Appointment as selected has to temporarily stop the contact/ for privacy issue
	16 = Appointment fictitious
6 - Out of target	17 = Group of individuals
	18 = second house
	19 = office, shop, company,…
	20 = disconnected phone number
	21 = dead
	22 = not eligible individuals in the family
7 - Other outcomes	23 = Severely sick
	24 = Answering but the family is untraceable
8 - Maximum number of trials	25 = Maximum number of possible trials without contact
	26 = Maximum number of possible family appointments
	27 = Maximum number of possible selected individual appointments

Source: Istat, Italian Violence against Women Survey
(a) Outcomes 17, 18, 19, 22 are possible only for some sample design frames

Table 7.1 b – Analytic outcome (a) and synthetic outcome of Capi/Papi contacts

Synthetic outcome	Analytic outcome
1 - Completed	1 = Completed
	2 = Stopped but valid
2 - Stopped	3 = Definitively stopped
3 - No-response	4 = Refusal of household member
	5 = Refusal of selected person
4 - Provisional outcome without contact	6 = Nobody at home
5 - Appointment	7= Appointment as selected person is temporarily not available
	8 = Appointment as selected person has to temporarily stop the contact/ for privacy issue
6 - Out of target	9 = Group of individuals
	10 = second house
	11 = office, shop, company,…
	12 = wrong address
	13 = dead
	14 = not eligible individuals in the family
7 - Other outcomes	15 = Severely sick
	16 = someone is present but not a member of the family (the family is untraceable)
8 - Maximum number of attempts	17 = Maximum number of possible attempts without contact
	18 = Maximum number of possible selected individual appointments

Source: Istat, Italian Violence against Women Survey
(a) Outcomes 9, 10, 11 are possible only for some sample design frames

Interview

Completed interview (analytic outcome = 1) and not completed interview still valid (analytic outcome = 2) are clearly cases where the interview has been obtained; indeed, in the second case it has been stopped when the last non-basic questions are asked. When

designing the questionnaire, taking into account the knowledge aims of the research and consequently organising its structure, the exact point where the interview can be stopped without losing relevant information can be defined.

Interrupted Interview

If the interview is stopped before the defined point, the interview has to be considered not valid, definitively stopped and lost (analytic outcome = 3).

Refusal

Refusals are classified in four types in order to better understand the points where difficulties arise in doing the interview. There are clue moments which define the conquest of the interview. If refusal happens before asking if the family/the respondent received the introductory letter – which is the first real question which is asked (analytic outcome = 4) – it is an absolute refusal: the respondent does not want to give any information, not even about the introductory letter. Other refusals happen after some interactions have been established and some information has been asked, for example, after the question about the introductory letter (analytic outcome = 5). In the following two cases (analytic outcome = 6, 7) the interview continues until the individual to be interviewed is selected, but he/she refuses to be interviewed, or if the respondent himself or herself does not allow the selected person to answer. Different analytic outcomes specify in detail when the refusal happened, allowing reflection about strategies to persuade the reluctant. In case of refusal, the nominative is substituted.

No contact

Non-answer, busy, answering machine, fax/data (analytic outcome = 8, 9, 10, 11) all are temporary trials without contact. Such nominatives will be called again according to specific rules of re-calls.

For face-to-face surveys they are generically non-contact.

Appointment

In the case of appointment (analytic outcomes = 12, 13, 14, 15, 16) different kinds of appointments are defined, depending on the situations in which they are taken. A family appointment is when the person to be interviewed has not yet been individuated and an appointment is scheduled with the respondent who can be a member of the family or not, a targeted person or out of target (for example a person aged less than the minimum age required) or he/she can be someone not belonging to the family (a cleaning lady, a baby sitter, another relative, a friend). A fictitious appointment is a not agreed one in case of a first refuse of the respondent. Instead of immediately quitting the nominative, a second chance is preferred.

Such second trial could seem too much bother for the family, but it is to be stressed that the final outcome for a household is strongly affected by the behaviour of the single component of the family that first answers. Such behaviour could cause a filtering or

dragging effect for the other family members. A drastic refusal of this first respondent to participate in the survey, or even to listen to the interviewer's motivations, without giving any chance to speak with other family members, suddenly stopping the call, produces what is named as the filter effect, causing the loss of the entire household. On the contrary, if the first respondent, even if he/she is not selected, is collaborative with a positive attitude toward the interview, he/she helps in building good communication and a positive mood in the further communication with the actual family member selected to be interviewed.

Appointments due to temporary absence or unavailability are scheduled after the selection of the individual to be interviewed, in the first case he/she is not at home, while in the second one he/she is temporarily busy but well-disposed to make the interview in another moment. The appointment due to a temporary stop is scheduled when the interview is started and the interviewee needs to stop it due to tiredness or other incoming and pressing activities or because of safety issues (for instance, in the meantime some other members of the household could have arrived).

For the face-to-face survey, no specific type of appointments could be identified.

Out of target

Accordingly to the sample design, the out of target code can be differently identified. For instance, if the sample unit can be identified out of the sample of phone numbers or of addresses with "Group of individuals" not corresponding to the households, as well as the second homes or "office, shop, company…" or if the phone number is disconnected, or in case there are no eligible people in the household because of death or other reasons.

Other residual outcomes

This category includes eligible households that cannot answer (severely sick or who is present at home says that the household is untraceable).

Maximum number of trials

The last categories consider outcomes of no interviewed households/individuals because the maximum foreseen attempts had been reached: maximum number of possible trials without contact, maximum number of possible family appointments and maximum number of possible selected individual appointments.

The maximum attempts for trials without contact can be, for instance, ten no answers, answering machine or fax, or ten series' of consecutive busy signals (where a busy outcome at least means seven consecutive busy signals, re-called, for example, every five minutes).

The maximum number of possible appointments - too many appointments is actually a way to refuse the interview - could be for instance three or four times, but it depends on the appointment typology. In fact, in the case of the appointment for safety issues, when the interview is already started, it is better to not have a strict schedule.

However, field supervisors and interviewers must be aware of the risk that persistence on their part could pose to women who are living with a violent partner. Repeat calls or visits to households where either the selected respondent or another household member has refused to participate must be undertaken cautiously so that violent partners are not

inadvertently alerted to the content of the survey and so that respondent safety is not put at risk as a result.

If a scheduled appointment results in no answer or answering machine or fax/data, the number could be called again after 30 minutes; while, if the number is busy it should be called again after five minutes. Also in these cases there will be a limit. For example, eight trials with outcomes of no answer or answering machine or fax or a sequence of busy signals, after the appointment, the nominative can be quit and substituted with the corresponding reserve number.

Furthermore, the contacting rules are not established once and for all. In fact, they have to be tailored according to the data collection phase: at the beginning of the survey they can be lenient, but when the data collection is almost ended, they need to be more restricted.

The Final disposition Codes are derived from the phone contact outcomes, and they have to be strictly monitored (Table 7.2).

Researchers and field supervisors need to pay close attention to the "consumption" of sample units. It is necessary to ensure that the sample is as representative of the target population as possible, thus minimising non-response bias. Field supervisors will also be required to work closely with sampling specialists in order to monitor sample allocation and to ensure that a sufficient number of call backs are made to selected households prior to new sample households being released. This issue also relates closely to the project budget since, if it is proving difficult to meet objectives related to the number of interviews within a specified time period, a decision can be made to extend the time spent in the field or to reduce the number of interviews that must be completed. For huge samples, these operations are made with sophisticated and automated sample selection processes, in the case of non-automated procedures, field supervisors may be responsible for overseeing both the sample selection and the release of the sample and they are required to monitor the sampling process closely.

Weighting procedures at the end of the process adjust for non-response among certain groups in order to render the sample more closely representative of the population. However, this is only a partial solution, since weighting cannot resolve the problem of having lost particular hard-to-reach populations, which are challenging to reach because of age or other characteristics such as race, ethnicity, language, or extreme poverty.

Capi surveys present the same final codes with only two exceptions, the EF3 codes are reduced in just one code: "No one at home" and EF4.1 – Non-working or disconnected telephone number – becomes "Wrong address".

Cawi surveys have a simplified version, because of the absence of interviewers. Final codes concern, above all, a complete interview, refusal, broken-off interviews, achieved time limit to complete the questionnaire, but also in this case there are codes for "no Internet" and other kinds of respondents' ineligibility reasons.

In the case of the mixed-mode technique, it is very useful to have at least the same macro codes, such as EF1, EF2, EF3, EF4 in order to monitor the coverage lists, the respondents' availability, the interviewers' performance (in case the survey uses the interviewer support).

Table 7.2 - Final Disposition Codes (Cati version)

Final Disposition Codes			
EF1	Interviews		
		EF1.1	Complete interview
		EF1.2	Partial interview
EF2	Eligible, No Interview	(Non-response)	
		EF2.1	Refusals
			EF2.1.1 Refusal before introduction
			EF2.1.2 Refusal after introduction
			EF2.1.3 Refusal of household member
			EF2.1.4 Refusal of selected respondent
		EF2.2	Break-offs
		EF2.3	Worked-out call-counting meter for appointments
EF2.3.1	Too many appointments of household members		
EF2.3.2	Too many appointments of selected respondent;		
EF2.3.3	Too many attempts after an appointment		
		EF2.4	Other cases
EF2.4.1	Eligible household, but household is never available		
EF2.4.2	The respondent's physical and/or mental inability to do an interview		
EF3	Unknown Eligibility, Non-Interview		
		EF3.1	No answer
		EF3.2	Always busy
		EF3.3	A telephone answering message /fax
		EF3.4	Call-screening
EF4	Out target		
		EF4.1	Non-working and disconnected number
		EF4.2	A group of friends
		EF4.3	Non-residential household
		EF4.4	Office, concern,...
		EF4.5	Death

Source: Istat, Italian Violence against Women Survey

About item non-response

Non-response can occur in any question owing to ambiguity in the question wording or response categories, respondents' lack of knowledge about the topic of the question, non-exhaustive or overlapping response categories, respondent resentment at being asked certain questions, lack of rapport with the interviewer, lack of time, interviewers skipping questions, respondent boredom and dropping out and coding or data entry errors. Item non-response is a function of the questionnaire design, the quality of the interviewers' training and performance and the monitoring of interviewers' performance (UNSD 2014).

To cope with these aspects, special attention needs to be given to each of these phases of the project. In fact, analysis carried out on different types of respondents and interviewers shows interesting interviewers' effects on data quality. Bad and good interviewers are the same from the beginning to the end of the interview: very often, in fact, interviewers with high refusal rates also have a higher item non-response rate, short interviews and low disclosure rates.

To monitor item non-responses, refusals and "I don't know" response to key questions, it is important to calculate item non-response rates. Item non-response rates can be reported by type of violence, by type of investigated content, by kind of population (sex, age, education, work situation, geographical area...).

7.3 Quality control during the survey process

Since interviewers are so important to achieve the quality of the entire data collection phase, monitoring their performance and supporting them is a very important issue.

The monitoring process can be split into two typologies, the qualitative actions and the quantitative ones.

Qualitative actions refer mainly to the interviewers' support, especially in the case of centralised data collection. These can be synthesised into mainly three activities: daily assistance during the entire data collection phase (made by researchers or their co-workers), debriefings with interviewers, psychological support. For Papi and Capi surveys, where interviewers are spread throughout the territory, field supervisors can take on these activities with periodic meetings and also the phone conference or skype conference, for example, can be useful.

7.3.1 Daily assistance

While in the case of centralised data collection, the daily assistance can be coded in a specific data room sheet with the aim of recording significant situations occurring during interviewing, evaluating the level of professional skills acquired and developed through the training received, discovering likely systematic and/or random errors that can occur due to the context and/or the way of managing the interview itself.

This form should be filled daily by appointed monitoring staff at the location where the data are collected. This can allow researchers to analyse the interviewers' performance and quickly intervene to correct and/or resolve on the field eventual unplanned problems and/or erroneous patterns.

The sheet is organised in three parts, the opening portion of the interview, the development of the interview, and the closing of the interview. For each phase the room assistant is asked to provide his/her personal judgement by answering some questions, with the possibility of adding some personal doubts or thoughts/notes.

Analysing the interviewers' behaviour in this systematic way is important, as it allows one to get an understanding of the quality of the interviews, the mood between interviewees and interviewers, the atmosphere of the centre from which the interviews are conducted, the risk of burn-out or anything else that can be problematic.

7.3.2 Re-contact calls

Another tool that can be used by researchers in order to monitor the interviewing process and to assess quality standards are the respondent re-contacts and their recording in a specific data sheet.

Respondents are randomly selected daily to be re-contacted and it's suitable to report the information gained in a re-contact data sheet. At any rate, attention needs to be directed to women's/men's disclosure of violence during the interviews. It is better to exclude them from the re-contact calls.

The aims of the respondent re-contacts concern the qualitative controls of the interviewer's work (whether the interview was fully administered and whether it was conducted in the proper manner); respondent opinions about the survey and about the interviewer's professionalism.

The re-contact data sheet should be organised as to collect both general respondent information as well as more specific questions to verify the correctness of the administered questionnaire and the respondent judgement about the interviewer's behaviour. If the outcome of these re-contact calls is negative, the responsible interviewer has to be monitored in an intensive way.

The re-contact sheet

In the re-contact data sheet there will be some variables able to identify the respondent and the interview, taken from the microdata file of the interviews:
date of the interview;
name of the interviewer;
identification number of the respondent;
sex of the respondent;
age of the respondent;
some information about what she/he said during the interview (for instance some information regarding the violence screening).

First, the variables regarding the respondent have to be controlled.
Then information captured during the re-contact call will be recorded in order to monitor the interviewers' behaviour and to find out the respondents' opinion about the survey itself, as:
the respondent has been interviewed: 1. YES; 2. NO
the questionnaire was administered in a proper way : 1. YES; 2. NO
the questionnaire was administered in its entirety (to do so, some questions that are located at the end of the questionnaire are asked in order to compare the answers) : 1. YES; 2. NO
how was the interviewer: 1. Kind; 2. Professional; 3. Too familiar; 4. Rude; 5. Not prepared; 6. Irritating; …
how was the survey: 1. Interesting; 2. Useful; 3. Annoying; 4. Useless; 5. Inappropriate …
respondent judgement about the opportunity and procedure of asking some delicate and sensitive questions: open question other information …

7.3.3 Quantitative indicators on the quality levels

Using statistical quality control methods to assess and improve the quality of the data collection and coding operations is extremely important for assessing the quality process.

Everything needs to be monitored and evaluated in the field, for example, regarding the interviewers' performance (if any), the respondent burden, the chosen strategies about the interviewing time scheduling, the interviews' mode, the effectiveness of the adopted questionnaire, the mailings, and the disclosure.

Monitoring data collection while it is in progress is very useful because the knowledge of productivity helps survey managers in their decision-making throughout the collection period. It is also useful to monitor the costs, so that it is possible to decide which decisions can be undertaken in order to improve quality (for instance, to increase response rates).

A system of quality indicators allows one to identify operational efficiency and cost-efficiency opportunities (e.g. sequence of calls, best time to call, optimal limit for calls or visits, etc.) in order to improve current and future collection processes and practices on the one hand, and provides feedback at the interviewer or operator level, on the other.

Important quality measures include response rates, processing error rates, follow-up rates and rates of non-response by reason. When these measures are available at all levels at which estimates are produced and at various stages of the process, they can serve both as performance measures and measures of data quality.

There are several variables to be controlled, but the strategic elements to be monitored are the respondents' characteristics (gender, age, and geographical area), interviewers' characteristics (gender, age, education, previous work experience), and the interviewer-interviewee relationship (assessed availability, relationship difficulty...). Each interview phase, from the introduction to the survey, to the interview itself until its conclusion, need to be monitored and this is much easier when the process is automated (for Cati, Capi, Casi, Cawi).

The data collection can be outsourced or in-house and in both situations, the quality has to be under control. The set of quality indicators is decided upon during the planning phase.

Tools that can be used to monitor the survey quality *in itinere* can be the following:

indicators, whose daily elaboration can be automated and made available on-line (if computer assisted systems are used);

control charts, based on the above indicators, also to be elaborated daily.

For computer assisted surveys, files are designed that must be produced by the external company or in-house, so that everything that happens can be reconstructed, possibly codified beforehand and documented in electronic format. The analysis of the indicators must be daily, since the survey progresses rapidly, in order to correct any erroneous situation. The response rate needs to be analysed, starting from the first days of data collection, comparing it with external sources or previous editions of the survey, so that any distortion can be immediately detected.

For the entire data-gathering period, the interviewers' activity has to be constantly kept under control via the analysis of a specific daily report, with indicators calculated singularly for each interviewer. This tight control is not only to supervise the performance of the interviewers, but also to give them support if they must face particularly complex or difficult situations, immediately identifying the need for additional or remedial training.

In case of refusals, it is important to collect data on them. The questionnaire can contain a part regarding some information can be requested directly from the person who is refusing about the reason for the refusal, the age and gender.

In Statistics Canada various software tools are available to monitor quality. These include the Quality Control Data Analysis System (QCDAS) and NWA Quality Analyst (Mudryk, Bougie and Xie 2002).

7.3.4 Monitoring computer assisted interviews

Monitoring data quality is easier for Cati surveys, starting from the centralised collection phase, but all Computer Assisted data collection can use similar strategies in order to control the quality in the ongoing process.

In a Cati system of interviews, the results of calls can be monitored on a daily basis also by crossing them with time and space variables, and with the interviewers' code, many indicators suggested by international standards AAPOR (American Association for Public Opinion Research) can also be calculated. Such indicators are useful during the monitoring phase, but also for evaluating the overall quality of the survey and assuring the comparability of the results with other data sources.

Such indicators are calculated on the basis of the final outcomes (table 2), that is relative to the closest contacts, but also on the basis of provisional results, that is relative to all the contacts made for each phone number called.

For Capi and Cawi surveys, the monitoring could be also carried out twice a week.

Indicators relate to each interview phase: family contact, selection of the person to be interviewed (if any), asking if respondents/households have received the introductory letter and the interview. Such indicators are expressly developed to evaluate *in itinere* the suitability and respect of the adopted methodology and to discover eventual problematic situations to be quickly addressed.

The quantitative indicators' set, whose daily analysis and study allows one to continuously verify the respect of the fixed qualitative standards and the evaluation of the interviewers' performance, also allows one to best organise the first contact with the family, that is, the mailing of the introductory letters, to assure that households will receive them in the days closely before the interview.

Rates assure a further analysis in all phases of the interviewer-respondent communication, from the introductory contact with the family and the selection of the person to be interviewed, up to the managing of the interview, providing an objective check about the actual understanding of the definitions and adopted classifications.

Numerous outcome rates are commonly cited in survey reports and in the research literature. The same names are used to describe fundamentally different rates and different names are sometimes applied to the same rates. The most used are response rates, co-operation rates, refusal rates, and contact rates.

For calculating outcome rates, the final disposition codes above described are used with the denominators described below:

EF = EF1+EF2+EF3+EF4 =	All telephone numbers called/all addresses visited
EF1+EF2+EF3 =	Eligible household/respondents and unknown eligible household/respondents
EF1+EF2 =	Eligible household/respondents
EF1+EF2.1+EF2.2+EF2.3+ EF2.4 =	Eligible household/respondents who are not free to make the interview
EF1+EF2.1+EF2.2 =	Eligible household/respondents who are free to make the interview

Outcome rates are computed dividing the final disposition codes by all denominators, but only some of them are the most useful.

Response rate A is the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews (refusal and break-off plus worked-out call-counting meter for appointments plus others) plus all the cases of unknown eligibility (unknown if household). It is clearly the most important meta-data in a sample survey and plays has a strategic role in the whole monitoring process, but the joint analysis with other indicators and rates surely allows better accuracy and overall vision.

Response rate B is the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews (refusal and break-off plus worked-out call-counting meter for appointments plus others). This is a particular rate that assumes there are no eligible cases among the cases of unknown eligibility.

Refusal rate A is the proportion of all cases in which a housing unit or respondent refuses to do an interview, or breaks-off on interview of all potentially eligible cases.

Refusal rate B is the most appropriate and consider the proportion of all cases in which a household or respondent refuses to do an interview, on completed interview, interrupted interview and refusals.

Interruption rate considers the proportion of all cases in which a household or respondent has interrupted the interview, on completed interview, interrupted interview and refusals.

Contact rate measures the proportion of all cases in which some responsible member of the household was reached by the survey.

Out of target rate is defined on the basis of the familiar and individual eligibility criteria. So the out of target rate, calculated through the ratio between the number of calls/visit to ineligible people and the total number of calls/visits, is an optimal control tool of the correct application in the field of the specific methodologies adopted.

A further qualifying aspect is the length of the completed interviews that allows one to discover eventual deficiencies in managing the questionnaire when the interview lasted too long, or on the contrary, when there is a too rushed management of the communication with the respondent, carelessness of the data quality when the interview was too brief.

In gender-based violence surveys, for example, monitoring that indicator is also helpful for studying interviews where at least one episode of violence is collected, to consider their higher number due to more in-depth sections, and the consequent greater difficulties that interviewers and interviewees can face.

Analysis of the interview length can also be used to evaluate part of the respondent burden.

The territorial analysis of such indicators allows one to control for specificities due to cultural and social aspects related to coexistence of different realities in the country, while the analysis by interviewers allows one to verify their specific skills for dealing with such a particular theme, their skills in gaining respondents' trust and collecting their experiences.

Rates in table 3, based on the outcome of the calls, allow one to evaluate the process quality and are necessary for building control charts.

Table 7.3 - Main survey rates

Response rate	No. Interviews/Total dials
Response rate by eligible	No. Interviews/Total dials to eligible households
Out of target rate	Out of targets/Total dials
Not contacted rate	Not contacted/Total dials to eligible households
Appointment rate	Appointments/ Total dials to eligible households
Refusal rate	No. Refuses/Total dials
Refusal rate of selected	No. Refuses of the selected woman/Total dials to eligible households

Source: Istat, Italian Violence against Women Survey

About the mode effect, some studies try to identify measurement bias that is attributable to the mode of data collection. Ideally, mode effects can be investigated using experimental designs where sample units are randomly assigned into two or more groups. Each group is surveyed using a different data collection mode. All other survey design features are controlled. Differences in the response distributions for the different groups can be compared and assessed. Other methods, such as the propensity score method or regression analysis, can be used to assess mode effects

when experimental designs cannot be applied. Others directly carrying out mixed-mode, by looking at their results (about this topic, see what has already been done in Canada, Italy, Netherlands).

7.3.5 Monitoring paper and pencil interviews

The organisational structure of a personal interview (Papi or Capi) is more complex than the ones of a telephone interview (Cati) and generally foresees several levels of territorial articulation.

Sometimes the territorial offices are linked in a network throughout the country with the centralised office. The control can be made during work in progress and ex-post. The control during work in progress (or at the same time) regards the observance of methods, procedures and times set for the carrying out of the interviews by the local office and the interviewers. Suitable data sheets could be prepared to collect data on households' outcomes, such as refusal, interruption, ineligibility or being unreachable.

Tolls to speed up the monitoring can be, for instance, the Interactive Voice Response. For each contact with the household, interviewers have to complete a monitoring sheet transcribing the contact outcome codes (contact made both by phone and in person). Then, every two days, for example, the interviewer is asked to transmit the monitoring sheet to a centralised toll-free phone number to record the control data.

7.3.6 Control charts

Control charts are very useful as an additional tool for monitoring the data collection. The experience made suggests integrating controls on the field with a massive use of Shewart Control Charts. Such techniques, much used in the industry for making statistical quality controls during production processes, allow one to verify the stability of a process, the coherence with specific qualitative instances and to evaluate the outcome of corrective actions to bring the process variability under control again. The control of the work in progress is based on daily measurements of basic characteristics of the process quality.

If, in the case of industrial production, identifying an out of control parameter of the process translates in maintenance actions for the installations, when dealing with a social survey, actions to be made will be firstly to better train the interviewers, and in more severe cases, they will be to remove the worst interviewer's sources of qualitative errors. It is also possible that out of control points shed light on interference related to territorial characteristics, the household contact time, or other logistic aspects that can be quickly and rapidly resolved or improved.

A requirement for applying this technique is the individuation of key variables of the process, that is, those more correlated with the output process quality to be monitored, in the case of the social survey, during the phase of data collection. Key variables are to be found among quantitative indicators and rates built taking into account the methodological specificity of the survey and trying at the same time to harmonise them with the AAPOR international standards, to assure comparisons with other sources of data.

Control charts referring to the entire period of the data collection can be built according to various time variables: for all data collection into day, week, two weeks, and months. This allows the time trend of the survey to be analysed with a greater number of sample units, considering the more stable and significant outcomes, and if the survey lasts for a long

time, improved summaries can be provided.

The analysis of the charts allows monitoring of the collection process with respect to the time, following the evolution of the indicators, evaluating some specific events such as, for example, inserting a new group of interviewers or new debriefing actions, if needed.

Beyond the analysis of the time trend, it is important to evaluate the performance of the Cati/Capi operator, so it is necessary to build control charts by each operator code and by each different training group they belong to.

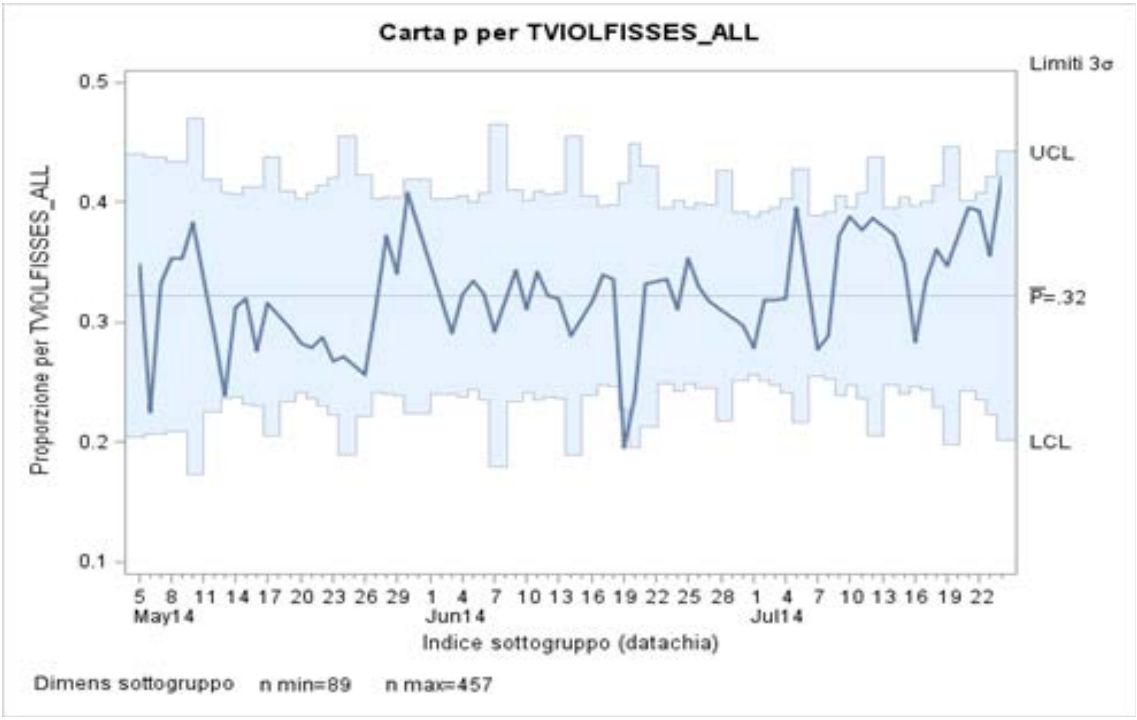
The training group has a specific importance. On the one hand, there are no differences in training modules and number of training days for interviewers, but on the other hand, characteristics of each interviewer can affect the outcomes, such as specific dynamics that may have occurred and developed during the training for some of the training groups.

To evaluate the performance and better individuate critical situations by the operators, control charts are to be analysed reporting the interviewer's code on the X axes.

The analysis of charts can suggest the need to study in more detail particular situations relative to specific interviewers or specific days, presenting out of control values for one or more rates, or presenting outlier time trends.

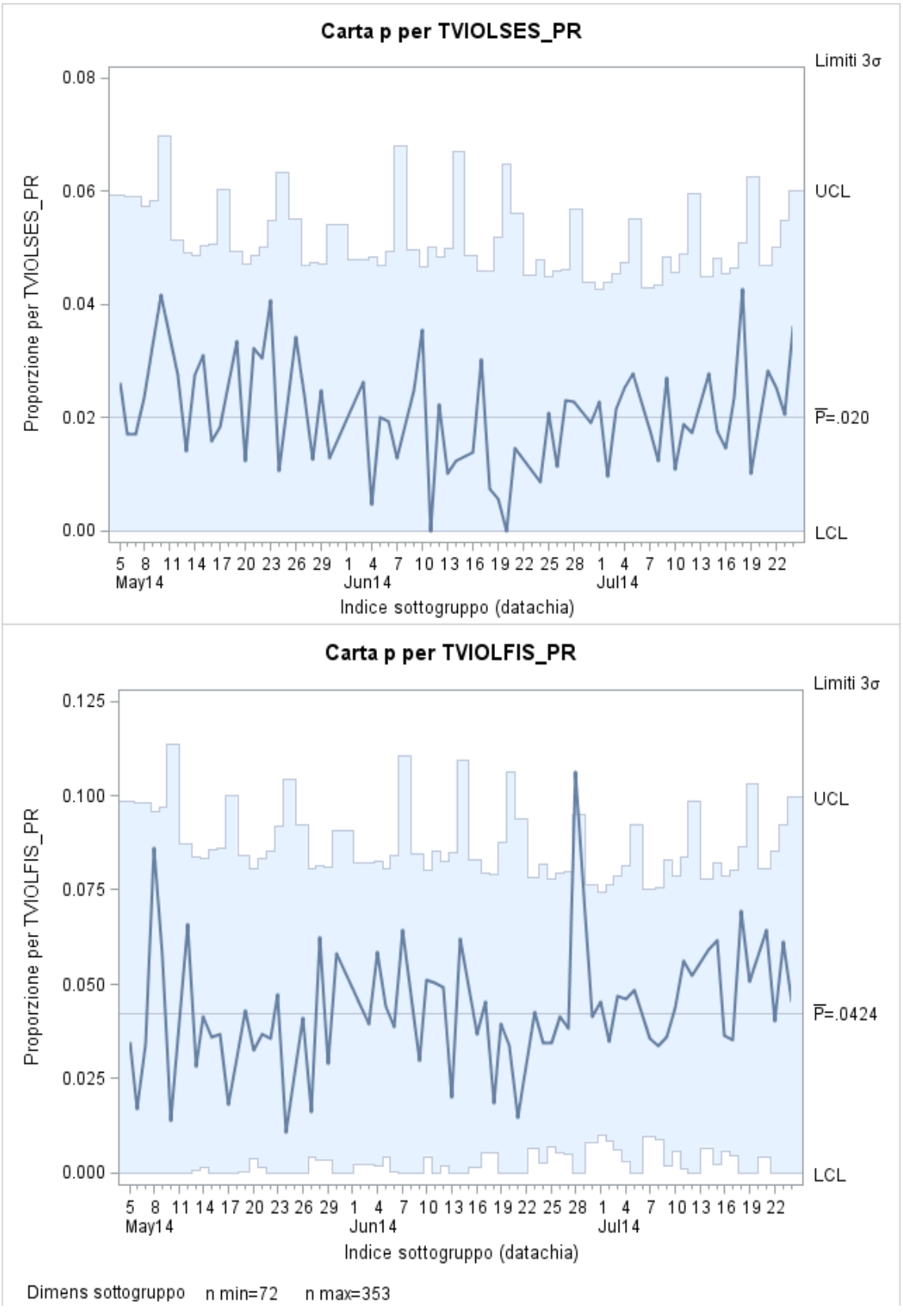
The violence indicators - prevalence rate of violence by time and by interviewers - can be calculated as total or by form of violence – physical and sexual - or by kind of perpetrator (Figures 7.1 and 7.2).

Figure 7.1 - P- Chart of prevalence rates (physical or sexual violence by any men) by time. Istat, Violence against women survey. Year 2014



Source: Istat, Italian Violence against Women Survey

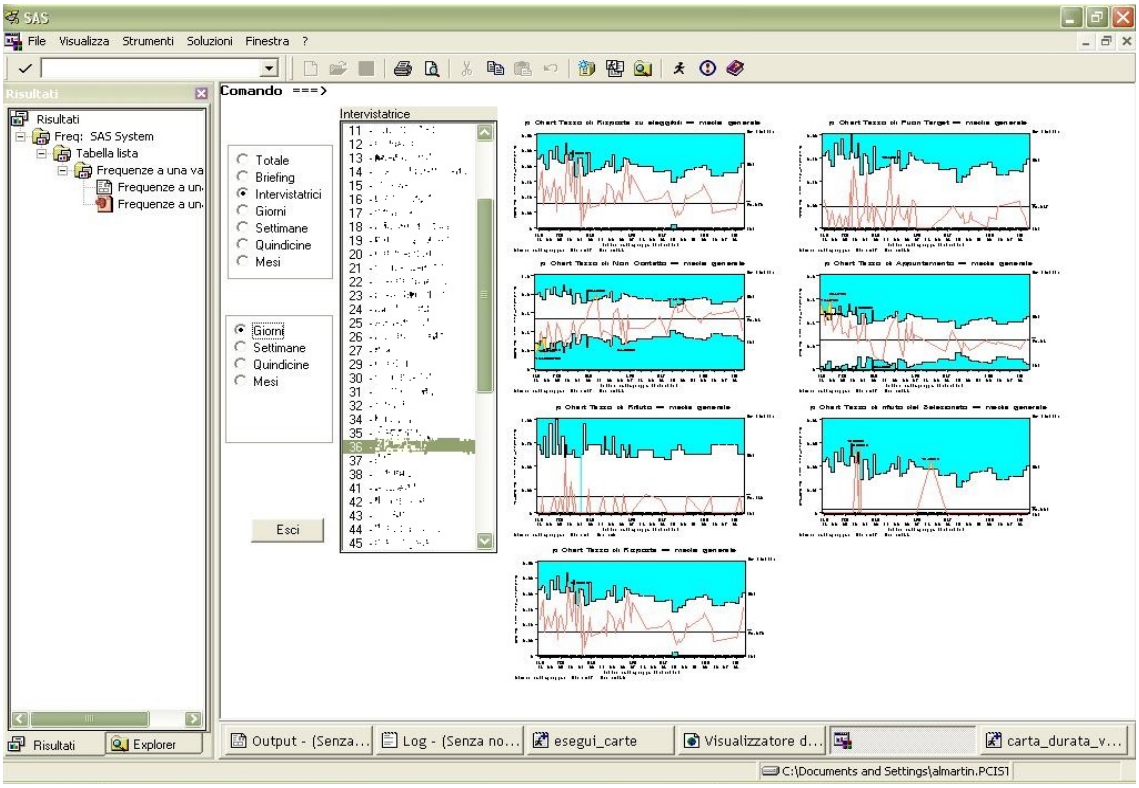
Figure 7.2 - P-Chart of sexual and physical violence rates by current partner and time. Year 2014



Source: Istat, Italian Violence against women survey.

Focusing the analysis on a single interviewer, the contemporary vison of all the charts provides an exhaustive picture of the situation regarding the correct managing of contacts, the ability of communicating with the family and of obtaining the interview, and gives further important information regarding time trends on continuity, efforts and motivation of the interviewer (Figure 7.3).

Figure 7.3 - The comprehensive box of control charts for a single interviewer. Year 2014



Source: Istat, Italian Violence against women survey. :

The ease of interpretation and reading of the charts allows a well-timed scheduling of debriefings to promptly face problematic elements that could arise or be found on the field.

When the "out of control" arises at the beginning of the survey there is the opportunity to focus efforts on keeping down the refusal rate, by improving the communication skills of the less practiced interviewers, and by giving a constant help to new groups of interviewers that, having finished the initial training, have been inserted in the ongoing survey. The efficacy of "targeted" debriefing is attested by the fact that when all the interviewers' groups were added and the survey was established, the process was then under control.

Other out of control values can also be observed in correspondence with holiday periods, when families showed lesser willingness in being interviewed. This is the usefulness of these kinds of tools, allowing them to monitor and adjust the data collection. In this regard, some improvements can be obtained, for example, by defining specific strategies regarding the management of shifts and parameters managing the Cati system algorithm.

The analysis by interviewer, made throughout the survey period, often highlights some pejorative elements in previous more efficient interviewers. In such cases, specific talks with them were planned to encourage and motivate those interviewers,

or groups to share with each other ways of better managing tiredness and difficult moments, and in this way, helping them to reduce burn-out.

The use of control charts to monitor the overall process of surveying is now a well-established tool, so it can be considered a recommended practice for the main surveys administered by Cati and Capi techniques.

7.4 Studying the interviewers' effect

To decide strategies to be adopted to best work with interviewers, the following instances have to be addressed:

which variables of the interviewer most greatly affect the interview's success?

what is the supervisor effect?

which training methods are to be used?

One proposal to study the above instances could be the use of a *multilevel cross classified* statistical model on data recorded in previous waves of the survey, when possible, or in similar surveys in terms of methodology and possibly of content.

The aim could be to separate the interviewer effect from other variables when analysing, for example, the item non-response rate.

The application of such models generally highlights that the preliminary work of interviewers to create a confidential context is strategic for predicting the success in answering a sensitive question. To better define such a work, it can be useful to divide the interview into two moments: the initial moment to convince respondents to cooperate and the interview itself.

The use of these models is very useful for defining in advance the best strategies to be implemented in order to achieve quality.

These studies are also very useful in order to assess the most suitable metodology for carrying out the survey.

REFERENCES

American Association for Public Opinion Research - AAPOR. 2016. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition*. Alexandria, VA, U.S.: AAPQR. https://aapor.org/wp-content/uploads/2022/11/Standard-Definitions20169theditionfinal.pdf.

Baker, J., C. Gentile, J. Markesich, S. Marsh, E. Panzarella, and R. Weiner. 2013. "Ensuring Data Quality: What Criteria Do Monitors Use to Rate Interviewers?". *Survey Practice*, Volume 6, N.1. https://doi.org/10.29115/SP-2013-0004.

Bethlehem, J., F. Cobben, and B. Schouten. 2009. "Indicators for the Representativeness of Survey Response". *Statistics Canada's International Symposium Series: Proceedings*. Ottawa, Canada: Statistics Canada/ Statistique Canada. https://www150.statcan.gc.ca/n1/pub/11-522-x/2008000/article/10976-eng.pdf.

Billiet, J., M. Philippens, R. Fitzgerald, and I. Stoop. 2007. "Estimation of Nonresponse Bias in the European Social Survey: Using Information from Reluctant Respondents". *Journal of Official Statistics (JOS)*, Volume 23, N. 2: 135-162. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/estimation-of-nonresponse-bias-in-the-european-social-survey-using-information-from-reluctant-respondents.pdf.

Dillman, D. A. 2006. *Mail and Internet Surveys: The Tailored Design Method*. Hoboken, NJ, U.S.: John Wiley & Sons Inc.

Dielman, L., and M.P. Couper. 1995. "Data Quality in a CAPI Survey: Keying Errors". *Journal of Official Statistics (JOS)*. Volume 11, N. 2: 141-146. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/data-quality-in-a-capi-survey-keying-errors.pdf.

Durrant, G.B., R.M. Groves, and F. Staetsky. 2010. "Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys". *Public Opinion Quarterly*, Volume 74, N. 1: 1-36. https://doi.org/10.1093/poq/nfp098.

Eurostat. 2015. *Quality Assurance Framework of the European Statistical System. Version 1.2*. Luxembourg: Publications Office of the European Union. https://ec.europa.eu/eurostat/documents/64157/4372717/Eurostat-QAF-V1-2.pdf.

Eurostat. 2014. *European Statistical System handbook for quality reports - 2014 edition*. Luxembourg: Publications Office of the European Union. http://ec.europa.eu/eurostat/documents/3859598/6651706/KS-GQ-15-003-EN-N.pdf.

Eurostat. 2011. *European Statistics Code of Practice - revised edition 2011*. Luxembourg: Publications Office of the European Union. https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-32-11-955.

Fuchs, M., D. Bossert, and S. Stukowski. 2013. "Response Rate and Nonresponse Bias: Impact of the Number of Contact Attempts on Data Quality in the European Social Survey". *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, Volume 117. N.1: 26-45.

Istat, 2001. Indagini sociali telefoniche: metodologia ed esperienze della statistica ufficiale, edited by Maria Giuseppina Muratore, Luciana Quattrociocchi, Linda Laura Sabbadini, Metodi e Norme,n.10.

Istat. 2005. Il monitoraggio del processo e la stima dell'errore nelle indagini telefoniche. Applicazione all'indagine sulla sicurezza dei cittadini, edited by Maria Giuseppina Muratore e Marina Signore, Metodi e Norme, n. 25.

Laflamme, F. 2009. "Data Collection Research using Paradata at Statistics Canada". Statistics Canada's International Symposium Series: Proceedings. Ottawa, Canada: Statistics Canada/ Statistique Canada. https://www150.statcan.gc.ca/n1/pub/11-522-x/2008000/article/10997-eng.pdf.

Laflamme, F., M. Maydan, and A. Miller. 2008. "Using Paradata to Actively Manage Data Collection". *Proceedings of the Survey Research Methods Section*. Alexandria, VA, U.S.: American Statistical Association. http://www.asasrms.org/Proceedings/y2008/Files/300608.pdf.

Lyberg, L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.). 1997. *Survey Measurement and Process Quality*. New York, NJ, U.S.: John Wiley & Sons, Inc.

Muratore, M.G., L. Quattrociocchi, e L.L. Sabbadini. 2001. "Indagini sociali telefoniche: metodologia ed esperienze della statistica ufficiale". *Metodi e Norme*, N. 10/2001. Roma, Italia: Istat.

Muratore, M.G., e M. Signore. 2005. "Il monitoraggio del processo e la stima dell'errore nelle indagini telefoniche. Applicazione all'indagine sulla sicurezza dei cittadini". *Metodi e Norme*, N. 25/2005. Roma, Italia: Istat. https://ebiblio.istat.it/digibib/Metodi%20e%20norme/IST0048279Ed2005N25.pdf.

Peytchev, A. 2013. "Consequences of Survey Nonresponse". *The Annals of the American Academy of Political and Social Science*, Volume 645, N. 1: 88-111.

Schmich, P., and F. Jentsch. 2012. "Fieldwork Monitoring in Telephone Surveys". In Hader, S., M. Hader, and M. Kuhne (*eds.*). *Telephone Surveys in Europe. Research and Practice*: 295-313. Berlin Heidelberg, Germany: Springer.

Singer, E., and C. Ye. 2013. "The Use and Effects of Incentives in Surveys". *The Annals of the American Academy of Political and Social Science*, Volume 645, N. 1: 112-141.

United Nations Office on Drugs and Crime - UNODC, and United Nations Economic Commission for Europe - UNECE. 2010. *Manual on victimization surveys*. Geneva, Switzerland: UN. https://www.unodc.org/unodc/en/data-and-analysis/Manual-on-victim-surveys.html.

United Nations Statistic Division - UNSD. 2014. *Guidelines for Producing Statistics on Violence against Women* - Statistical Surveys. New York, NJ, U.S.: UN. https://unstats.un.org/unsd/gender/docs/guidelines_statistics_vaw.pdf.

World Health Organization - WHO, García-Moreno, C., H.A.F.M. Jansen, M. Ellsberg, L. Heise, and C. Watts (*eds.*). 2007*a*. *WHO Multi-Country Study on Women's Health and Domestic Violence: Interviewer's manual*. Geneva, Switzerland: WHO.

World Health Organization - WHO, García-Moreno, C., H.A.F.M. Jansen, M. Ellsberg, L. Heise, and C. Watts (*eds.*). 2007*b*. *WHO Multi-Country Study on Women's Health and Domestic Violence: Study protocol*. Geneva, Switzerland: WHO.

World Health Organization - WHO. 2001. *Putting Women First: Ethical and Safety Recommendations for Research on Domestic Violence Against Women*. Geneva, Switzerland: WHO. https://www.who.int/publications/i/item/WHO-FCH-GWH-01.1.

8. DISSEMINATION PLAN[1]

The aim of the dissemination plan is to ensure the correct, accurate, complete and accessible statistics and data of the GBV Survey.

A mainstream dissemination strategy is essential for achieving the greatest impact, visibility and credibility of the results. This strategy needs to answer the following five questions:

1. Who is our audience?
2. What information do we want to disseminate?
3. How will we disseminate the results?
4. When will we disseminate these results and raise awareness?
5. What dissemination of new digital data?

8.1 Who is our audience?

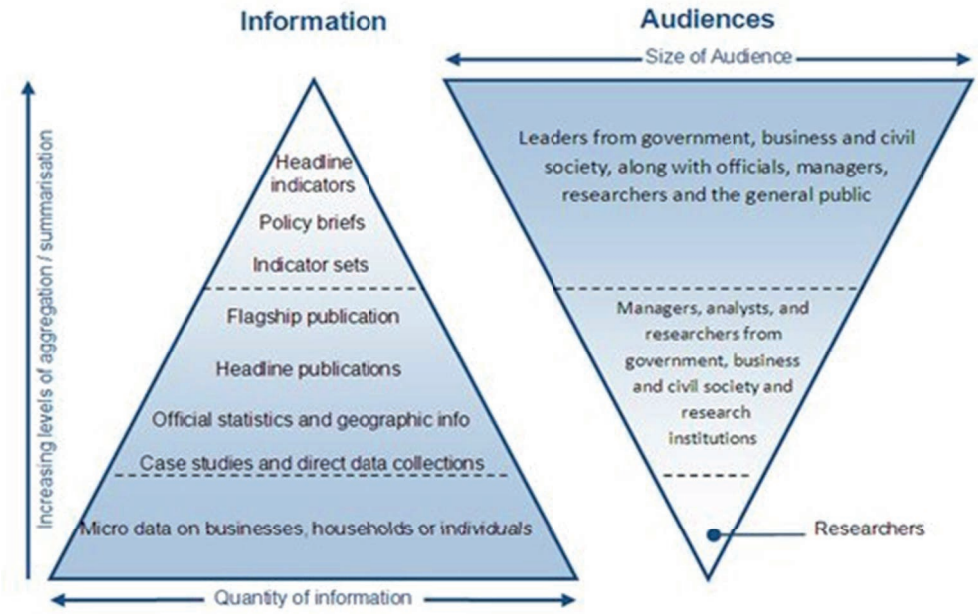
Following the United Nations Economic Commission for Africa (Uneca) indications, as the following Figure shows, we have to take into account that the choice of indicators and data to be disseminate is different in relation to the audience.

Who are the key users and what are the key means by which the system will inform them? Targeting users with specific publications or other dissemination media can be effective. Their effectiveness should be measured in order to continually assess demand.

One key aspect that a statistical agency should keep in mind is the need to provide relevant and effective data visuals for informing certain sectors in society. Journalists and citizens may require specific ways to view statistical concepts; policy-makers may have their own preferred ways to visualise and researchers need microdata for more in-depth analysis of the phenomenon. Developing creative approaches to data visualisation can have an impact in the process of making the use of data widespread, especially on a topic as sensitive as gender-based violence, that takes on different forms and meanings depending on whether one takes into consideration the violence suffered by men and that suffered by women. It is important to recognise that communicating data is a special case of communicating in general. In this context, use of the word “data” here is done in a more explicit sense of “information in numerical form” and not in the general sense of factual information. Hence, the goal of effective “data” communication is to ensure that data are transmitted, decoded, and understood accurately, and acted upon. So, different and customised deliverables meet their appropriate target, as Figure 8.1 shows.

1 This chapter has been drawn up by Claudia Villante and Maria Giuseppina Muratore.

Figure 8.1 – Different information for different audiences



Source: Uneca, 1984

For instance, for media deliverables, it is very important to lead them to the comprehension results through the info sheet with the main issues that allow them to better understand the complexity of the phenomenon.

Using info sheet one can afford a comprehensive overview of complex thematic, providing different information and data in a synthetic framework. For example, in order to better understand how gender-based violence is deeply connecting to the prejudice, attitudes and stereotype both data should be illustrating at the same time, as the following example shows.

Figure 8.2 - Overview of VAW incidence, gender stereotype persistence and social image of GBV

Source: https://www.istat.it/en/archivio/249195

Organising workshops with stakeholders, policy-makers, media and specialised services such as shelters allows better comprehension and better use of the data.

With this perspective, a broader communication strategy needs to be developed. In most cases, the main purpose of this strategy would be to systematically guide the efforts in raising awareness and mobilising support to the policy-makers and decision-makers so that they are able to provide effective measures to monitor and combat GBV.

8.2 What information do we want to disseminate?

With greater potential of the data, two aspects should be considered upon official dissemination, that is, related to the data analysis, and dissemination-at-large, related to the consumption of these data.

Official dissemination relates to the data analysis. The basic steps that have to be considered in analysing consists in identifying an issue related to gender-based violence.

Gender-based violence against women and men provides itself a crucial question to build an overall approach to the survey in relation to a definition of “gender-based violence”.

A second issue is related to pointing out differences of violence against women and against men, asking meaningful relevant questions, answering questions on gender-based violence through the examination and interpretation of data and communicating the key findings and results to the users. The definition of violence against men/women is very different in respect to the concept of gender-based violence against men/women.

Data analysis also has an important role as part of the survey development and revision process (in the case of longitudinal or updated surveys). Analysis is essential for understanding the results from studies, for planning new statistical activities, for providing information on data gaps, for designing surveys and for formulating quality objectives.

This document contains a collection of strategies, identifying target audiences and corresponding key messages, by which the task force, or the agencies, would seek to achieve its data advocacy and communication goal on gender-based violence against women and gender-based violence against men.

In this respect, it’s crucial to provide documentation and metadata along with the disseminated material that contains, as appropriate, descriptions of its quality and the methodology used to ensure the right focalisation of the gender-based violence against women and gender-based violence against men, so that users do not draw conclusions from the data that are inaccurate or inappropriate.

Indicators and topics to disseminate will have to respect the difference of data among men and women. Below a proposal for data dissemination.

EXAMPLE OF INFORMATION TO DISSEMINATE BY MAIN TOPIC

Overview of the phenomenon of violence (psychological, physical and sexual) by sex
- prevalence by perpetrator
- reference period (lifetime, during the last 12 months)
- type of violence
- duration
- frequency

Victims and perpetrators characteristics by sex
- type of perpetrators
- type of violence by age, education, citizenship

Reporting and seeking help by sex
- to whom one talked about domestic violence (physical/sexual violence by any partner or any relative)

Dynamic of violent action by sex
- place of occurrence
- frequency of the act
- type of consequences
- drug or alcohol abuse
- reaction of the victims
- people under 18, witnessed

Severity and effects of the violence by sex
- type of severity (injuries, psychological consequences, seriousness perception, life in danger, etc.)
- effect and limitations due to violence

The cycle of violence by sex
- witnessing of violence
- being abused

Risk factors by sex

Stalking phenomenon by sex

Sexual harassment at work by sex
- number of forms
- frequency
- support available in case of sexual harassment at work for employees by sex, availability of support and violence experience

The way out of the violence
- knowledge on support services or free legal aid by sex and violence experience

8.3 How will it be disseminated?

Dissemination occurs when information obtained through a statistical activity is released to users. Forms of dissemination can vary, depending on resources and demand. The question is not simply the way to render the data public (it can be via the Internet, infographics, ebook, paper publication, microfiche, microdata file, telephone, facsimile, public speech and/or presentation, or radio and/or television interview) but what are the aims that have to be followed in order to achieve the best result of dissemination.

Solutions to disseminate data

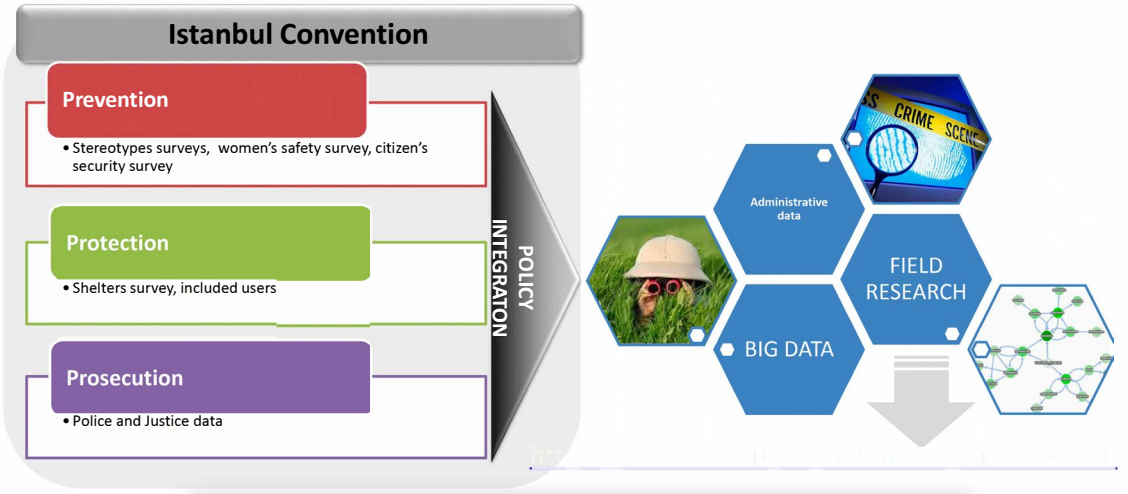
Official data can be disseminated in different ways:

1. **Using a multi-source approach at national level and at the international level**, through the agency's website from the producer to the users with a dedicated hub on VAW (*e.g.* https://www.istat.it/en/violence-against-women).

This hub is an example of a multi-source approach to data dissemination, where statistical data are published with short reports[2].

Multi-source approach follows 3P Istanbul Convention Strategy: **Prevention** of violence, addressing its causes and changing attitudes, gender roles and stereotypes that make violence against women acceptable, providing data on stereotype surveys, women's safety survey, and citizen's security survey. **Protecting** women and girls who are known to be at risk and setting up specialist support services for victims and their children, reporting data on shelters and anti-violence centres. **Prosecuting** the perpetrators, updating data from administrative sources, such as Police and Justice data. In order to give a holistic response to violence against women it should be adopt and implement state-wide integrated Policies that are effective, co-ordinated and comprehensive and it should be adopt an integrated data ware house, where that information can be available and update (Figure 8.3)

Figure 8.3 – Istanbul Convention and data needs



Source: https://www.istat.it/en/violence-against-women

2 The "Micro Strategy" software allows an interactive system matching supply with demand.

A Multi-source approach provides information not only to measure the phenomenon but also to better understanding it. It means that it suitable using data produced by others official sources (such as Police data), and in meantime carrying out a periodical survey to observe the trend of the gender-based violence, matching administrative and survey data.

Adopting a multi-source approach to data dissemination means collecting in a unique hub data, information and others relevant documents to better understand the evolution of the phenomenon over time.

Figure 8.4 – Example of Violence against Women dissemination hub

Source: https://www.istat.it/en/violence-against-women

2. **Organising workshops and events**, publishing info sheets and papers, also in coincidence with the main celebration days, such as the international day of violence against women (the 25th of November) or international women's day (the 8th of March).
3. **Finding key messages for preparing the awareness campaign**, helping policy-makers and stakeholders to fine tune the aims of the intervention.

As an example, below is shown the main Italian campaign designed after the first VAW survey (2006) in which the main messages were:

Key messages from the VAW research	Campaign pictures
DATA ANALYSIS – Repeated violence is more common by partner than by non-partner (67.1% compared to 52.9%). – The peak is for sexual violence by the current partner (91.1% of repeated violence). – Injuries, life felt in danger and heavy consequences are more frequent for domestic violence. KEY MESSAGE **Violence is more often from someone you know**	

Key messages from the VAW research	Campaign pictures
DATA ANALYSIS – Partner psychological violence is common and often means escalation of violence. – 7 million 134 thousand women (43.2% of women with current partner) were or are victims of psychological violence. – Isolation and attempted isolation (46.7%), control (40.7%), financial violence (30.7%) and berating (23.8%), followed by intimidation (7.8%). – 21.1% of total victims have suffered, always or often, this kind of violence. – 6 million 92 thousand women were only victims of psychological violence by their current partner (36.9% of women living as a couple). KEY MESSAGE **Violence is hidden and pervasive**	
DATA ANALYSIS – Not reported violence: 96% by a non-partner; 93% by a partner. Not reported rapes: 91.6%; 94.2% for attempted rapes. – Victims talk of violence with no one: 33.9% of victims by partners – 45% for victims of current violence – and 24% by a non-partner. – Victims do not seek help (only 2.8% of victims have reported to shelters or keep contacts with other associations supporting women). – Only 27.3% of women raped by their partner consider the suffered violence as a crime (one of the reasons for under-reporting, also for new generations of women). – 20% of victims of violence by partner say they feared for their life, a very at risk set, since domestic murders often happen after an escalation of violence. KEY MESSAGE **Violence is not reported**	

Key messages from the VAW research	Campaign pictures
DATA ANALYSIS Injuries, life felt in danger and heavy consequences more frequent for domestic violence. – Partner uses or threatens to use pistols or knives (52.5%). – Attempted strangulation and burning (78%). KEY MESSAGE **Violence is very serious**	
DATA ANALYSIS – Risk factors: partners' behaviours, women's considerations and previous experiences of violence in childhood. – When current partner is violent physically or verbally 35.6%, versus 7.2% compared to 25.7% versus 7.2% outside the family. – when women suffer psychological violence (in particular when partners humiliate and berate them) 35.9% versus 7.2%. – When men were witnesses or victims of violence in childhood 30-35% versus 6% for no childhood experiences. Same tendency when women experienced violence in childhood. – Women have to be aware that their children exposed to violence will tend to reproduce learned models and the violence itself too, daughters tolerating it, sons perpetrating it. KEY MESSAGE **Children reproduce models they learn at home**	

The aim of the advertising campaign is to make the phenomenon visible, widespread, for the public, also using blogs, web sites, social media, by private citizens, foundations, cultural organisations, businesses and companies, local administrations, schools, editors, TV, political parties and associations.

Planning a dissemination strategy also means identifying tools and indicators useful for assessing their impact on the different targets identified.

In this respect, the dissemination plan should be regularly monitored with specific indicators. These indicators should respond to certain evaluation questions and in relation to the purpose of the plan, as the following Table suggests (Table 8.1).

Communicating official data and statistics requires concerted efforts. When data are communicated well, it is easy to appreciate their positive influence on the consumption.

Table 8.1 - Indicators to monitoring effectiveness of dissemination on the data

Effectiveness of dissemination of the data	Indicators
Are data portals/hubs effective ways to disseminate aggregate data?	– Measure public access and use of certain sites by "google analytics"
Are there other ways to disseminate the data?	– Number of informal phone calls received for data; – Number of emails requesting data and what data is requested
Which institutions in the country are most active in data use?	– Number of researchers and agencies actively engaged in a partnership with statistical agencies
What are the most effective public information campaigns for increasing data use?	– Increase in website traffic after promotional campaigns; – Using a sentiment analysis

Source: https://www.istat.it/en/violence-against-women

8.4 When will we disseminate these results and raise awarness?

In order to increase the likelihood that a target audience will pay attention and use the data made available, one must consider the appropriate timing and channels for the data delivery. For example, as already mention, in coincidence with the main celebration days, such as the international day of violence against women (the 25th of November) or international women's day (the 8th of March).

Nevertheless, in the last two decades, following the digitalisation of our societies new types of digital data sources become available. In the official statistics community, the term "Big Data" is often used to refer collectively to non-traditional data sources, as an "umbrella term" for all kinds of data other than survey and administrative data. Those new data require not only a new way to processing methodologies, computation paradigms, data access models, regulations, organisational aspects, but also to communicate and disseminate approaches and time. In this respect the question "when will disseminate" takes on a different meaning because the new digital data assume that they have to disseminate any time.

8.5 What dissemination of new digital data?

This is new frontier of the research aimed to improve statistical analysis methodology using Big Data[3]. This challenge is very relevant for social science and for the gender-based violence related topics. Timely data, in fact, are difficult to collect. In order to give an answer to this challenge, researchers, added also the Big data, monitoring and analysis of data coming from social and web. One of the way to analyse and communicate the results of this analysis is the sentiment analysis and the emotion detection. Thanks to complex methodological procedures and through ad hoc reading tools, it is possible to collect and observe the opinions of "social contents" and in general of messages conveyed by social channels on gender stereotypes and the phenomenon of gender violence. Social networks have the ability to create content, which can generate positive or negative effects on public opinion, amplifying or diminishing the informative scope of certain messages. Knowing the reactions to specific information campaigns, observing the characteristics of those who

3 The term "Trusted Smart Statistics" (TSS) was put forward by Eurostat to signify the evolution and officially adopted by the European Statistical System (ESS) in 2018 in the so-called Bucharest memorandum (European Statistical System Committee, DGINS2018 - Bucharest Memorandum adopted - ESS (europa.eu).

social content messages with a greater following, can help to increase our knowledge of how the public perceives the phenomenon of gender violence and how gender stereotypes are represented.

Using the new digital data helps the exercise of awareness's campaign monitoring and evaluation as above mentioned. For example, on the 25th of November (international day fighting violence against women), some monitoring indicators could be used, based on the evidences of social contents collected, focussing on:

- relevant themes linked to gender-based violence such as femicide, body shaming, rape and sexual violence;
- follow-up to awareness campaign linked to helplines;
- type of users profiling (institutions and politicians, non-governmental associations, influencers, common users).

REFERENCES

Kuonen, D., and B. Loison. 2019. "Production Processes of Official Statistics and Analytics Processes Augmented by Trusted Smart Statistics: Friends or Foes?". *Statistical Journal of the IAOS*, Volume 35, Issue 4: 615-622.

Radermacher, W.J. 2018. "Official statistics in the era of big data opportunities and threats". *International Journal of Data Science and Analytics*, Volume 6, Issue 3: 225-231.

Schulz, T. 2009. "Guidelines on the Presentation of Statistical Maps". United Nations, Statistical Commission, and United Nations Economic Commission for Europe - UNECE, Conference of European Statisticians - *CES. UNECE Work Session on the Communication and Dissemination of Statistics*, Warsaw, Poland, 13-15 May 2009. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.45/2009/crp.1.e.pdf.

Smith, A.D., and S. Rogers. 2008. "Web 2.0 and official statistics: The case for a multi-disciplinary approach". *Statistical Journal of the IAOS*, Volume 25, 3-4: 117-123.

Statistics Canada. 2006. *Policy on Informing Users of Data Quality and Methodology (approved July 27, 2006)*. Ottawa, Ontario, Canada: Statistics Canada.

United Nations Economic Commission for Europe - UNECE. 2004. *Communicating with the Media. A Guide for Statistical Organizations*. Geneva, Switzerland: United Nations.

United Nations, Economic and Social Council, and United Nations, Economic Commission for Africa – UN. ECA. 1984. "Report writing and dissemination of data". UN. ECA *Workshop on Household Surveys*, Lusaka, Zambia, Africa, 1-6 October 1984. http://hdl.handle.net/10855/21992.