Tackling the Problem of Self Selection in the Integration of Different Data Collection Techniques

Furio Camillo and Ida D'Attoma

Abstract A lot of studies are showing an increased tendency to use more than one mode of administration to collect data for a particular analysis ([10];[1];[16];[9]). Thus, understand if different data collection methods influence answers becomes a concern. The problem of self selection into different interview modes demands attention especially when the assignment to one or another collection method is not randomly controlled and respondents might self select in one data collection method over another. This paper shows an empirical case concerning the evaluation of two different data collection method: the CAWI (Computer Assisted Web Interview) method and the CATI (Computer Assisted Telephone Interview) method. If self-selection exists both mode of data collection and characteristics of the respondents influence answers. Hence, the mode effect may be confounded. In order to estimate an unbiased mode effect, this paper proposes a data driven multivariate approach to monitor self-selection that allows to disentangle interview's modes effects on answers from the effect of self-selection. We will work through the use of the monitoring system with an empirical case. In particular, we will use AlmaLaurea data and compare results of our approach to PS adjustment method that AlmaLaurea usually applies to control data quality as documented in various reports and analysis conducted by the AlmaLaurea Consortium .

Key words: self-selection; global imbalance; mixed interview mode; mode effects

---

Furio Camillo
Dipartimento di Scienze Statistiche dell'Università di Bologna, via Belle Arti 41, 40126 Bologna, e-mail: furio.camillo@unibo.it

Ida D'Attoma
Dipartimento di Scienze Statistiche dell'Università di Bologna, via Belle Arti 41,40126 Bologna e-mail: Ida.dattoma2@unibo.it

Acknowledgements

We acknowledge the AlmaLaurea researchers for giving us the possibility of comparing our approach to the classical Propensity Score method they used to apply for data quality control. Special thanks go to Silvia Ghiselli and Valentina Conti.

1 Introduction

This work stems from the consideration that using different interview modes might cause differences in answers. Consequently, if differences are caused by the interview mode an adjustment method could be needed. In applied research an increased tendency to use mixed data collection methods exists ([2], [23], [18]). Considering the case of the mixed use of the CATI/CAWI we aim at exploring the following research question:

- Do different treatments (CATI/CAWI) generate different outcomes (answers of interest)?

To answer to such question we need to tackle the influence of self-selection on the observed answers, being quite often the assignment-to-treatment not random.
On many fields, the problem of selection bias is mainly faced by referring to the Propensity Score(PS) approach proposed by Rosenbaum and Rubin (1983) ([20]). The authors scrutinized the problem by modeling the selection process given a set of pre-treatment covariates as a way of reducing bias in the estimation of the treatment effect of interest. In particular, they demonstrated that, having in hand several pre-treatment information which characterizes the units under analysis, it is possible to create groups of units having similar pre-treatment characteristics. These groups are, therefore, theoretically independent from the kind of undergone treatment. Within these groups becomes possible to compare the target variable among those who have undergone the treatment and those who have not or just have undergone a different treatment. Lee(2006)([17]), for example, suggested the use of Propensity Score Adjustment (PSA) as an approach of adjustment for volunteer panel web survey data. Another work which investigates Propensity Score as a method for dealing with selection bias in web surveys is that of Schonlau et al. (2006)([22]) who propose to construct weights based on the propensity scores to correct for selectivity. Despite PS is widely applied to correct for bias in many fields it suffers from some drawbacks: the main is that it is prone to model dependence. For this reason, to control for selection bias we propose the use of the Multivariate Approach introduced in Camillo and D'Attoma (2010)([4]), extended in D'Attoma and Camillo (2011) ([8]) and applied in Peck et al. (2010)([19]). Such Multivariate Approach has three main features. The first feature - the measure of Global Imbalance- is a single measure of imbalance in data, mainly based on the concept of inertia as a measure of dependence among categorical covariates and the assignment-to-treatment indicator variable. The second feature is the multivariate test of imbalance, that allows to

test Global Imbalance in data and represents an improvement over the variable-by-variable test. The last feature - the use of Cluster Analysis- enhances the possibility of finding local spaces in which balance holds according to the GI measure and its related test; within balanced groups the mode effect estimates are unbiased from selection. The idea is that once obtained groups of respondents whose pre-treatment characteristics are independent from the interview mode according to the GI measure and its related test, then any observed difference in the study outcome between CATI and CAWI respondents is attributable to the interview mode. On the basis of such an analysis if an effect of the interview mode on the study outcomes exists, an adjustment method will be needed. We have applied the Multivariate Approach to AlmaLaurea survey on second level graduates' employment condition as an example of potential application. We use such data for two main reasons. First, because data are related to a large-scale survey that includes a data collection process well controlled in each phase. Hence, other sources of non-sampling errors besides that due to the two data collection method (i.e. CAWI/CATI) are a priori minimized. Second, AlmaLaurea used to adopt in their surveys a complex system to monitor data quality that includes also the measure of the mixed modes effect. Therefore, our results can be compared to that of Almalaurea. In particular, we will first illustrate the Propensity Score Subclassification method they adopted and then apply to their data the multivariate approach and compare results. In section 2 we will briefly present the case of AlmaLaurea survey on graduates condition. Section 3 introduces our proposed methodology and the Propensity Score Subclassification used by AlmaLaurea. Section 4 presents results (AlmaLaurea PS subclassification results and the multivariate approach results). Finally, section 5 discusses and concludes.

2 The mixed use of CAWI and CATI methods: the AlmaLaurea case

Increasingly, data are collected by mixing different survey modes. Mixed mode surveys combine the use of different data collection procedures (i.e. telephone, mail, web, face-to-face interview) in a single survey project. The reasons underlying the use of Mixed Mode Surveys are many. In some instances, a mixed mode survey may provide an opportunity for respondents to choose or switch method that may increase the participation rate. Some methods are significantly less costly to implement than others. Furthermore, in large-scale surveys, some methods like mail questionnaires may also allow to save time in conducting the survey. If on one hand the use of the mixed mode gives some advantages, on the other hand it may introduce a big disadvantage due to the fact that different modes potentially provide different answers and this problem must be minimized. Many articles document the increased use of mixed modes and propose methods to minimize such a problem (i.e. [14],[24], [22],[21], [23]). Also applications of mixed modes are increased. Here, we first briefly present the case of AlmaLaurea survey on graduates condition that consists in implementing a Propensity Score Subclassification Method to control for self-selection, then apply the multivariate approach here proposed to the same data and compare results. Every year the Inter-Universities Consortium AlmaLaurea [1] carries out a survey on graduates' condition. As reported in Camillo et al. (2009)([3]) and Cammelli et al. (2011) ([6]), the survey makes possible to analyze the most recent labor market trends through an examination of the career opportunities available for the Italian graduates of the universities taking part in the Consortium during the 5 years on from graduation. All graduates are contacted 1,3, and 5 years on from graduation. The data have been collected during the survey conducted by AlmaLaurea in 2008. More than 287,00 graduates were examined. This survey also involved all first and second level [2] graduates from the class of 2007 (about 140,000). The huge number of graduates involved has determined the necessity to use survey methods that allow the reduction of costs and duration. This objective has been achieved through the introduction of two survey methods: CAWI and CATI. The graduates having a mailbox (85% of the cohort) have been emailed and also included two e-mail reminders. Afterwards, all graduates who had not answered to the online questionnaire (and, obviously, graduates not having a mailbox) have been contacted by phone (Figure 1). More Precisely, given the survey's cohort, self-selected respondents voluntary participate to CAWI. Afterwards, the non-respondents are followed by CATI. In this way the coverage error and the non response error have been both drastically reduced [3]. But, at the same time being the two interview modes sequential, potential selection effects may arise that they aim to control.

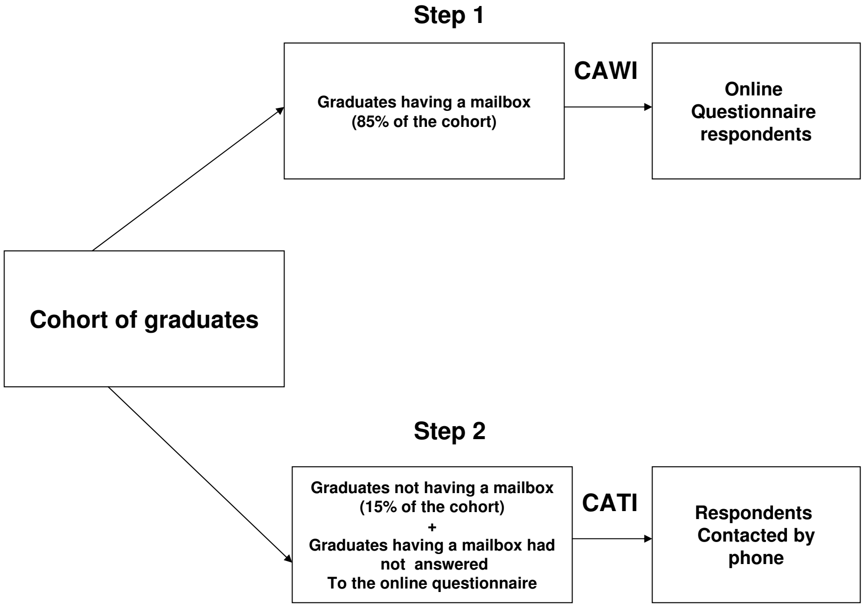
The use of CATI, for some aspects, overcomes the CAWI *under-coverage* problem ([1]), that means that only respondents with Internet access can complete the questionnaire form. The CAWI self-selection problem remains, meaning that it is completely left to individuals to select themselves for the survey. Based on the belief that any interview mode affects the probabilities of including respondents in a sample([22]), but also their answers, the AlmaLaurea aim was to determine if the observed differences in the answers were determined by self-selection (e.g. those who are most inclined to answer to CAWI interviews are the same who have specific characteristics) or just by the different data collection method. The survey enabled the Inter-universities Consortium AlmaLaurea to collect the main information related to academic and work experiences made after graduation: employment condition at the time of interview, characteristics of the job (contract, branch of activity, earning), time-to-entry into the labor market. These pieces of information are

[1] AlmaLAurea was set up in 1994. Up to March 2009 the number of its member universities amounts to 52,000, which correspond to 70 per cent of Italian Graduates. AlmaLaurea was set up to be at service of graduates, universities and companies. To reach this goal, a reliable and updated documentation on human resources which have received an advanced university education, is made available by the Consortium. This documentation, available also in English, constitutes the online graduates' databank, which also includes graduates having a long work experience, so that transnational mobility and the interaction of supply and demand are facilitated. Further information on survey methods can be found at: www.almalaurea.it

[2] Cycles of the Bologna Process

[3] AlmaLaurea does not use a sample survey methodology. Instead the idea is to interview all graduates. The non-response rate is around 10 %

Fig. 1 The Survey Protocol



integrated by the huge quantity of pre-survey data on socio-demographic characteristics of graduates (e.g. social origins, gender,age), pre-university studies, academic studies (e.g. degree course,graduation mark) and further experiences made during studies (foreign languages and IT skills, internships, study experiences made abroad and work experiences). The pre-survey information is mainly based on detailed administrative data from Universities archives and on a preliminary survey conducted close to the end of the university experience. All information that AlmaLaurea used in its study are summarized in table 1. As highlighted previously, they aimed at

Table 1 Collected data

Pre-treatment (X)	Treatment (T)	Post-treatment (Y)
Information on career		employment
Information on the family	CAWI	contract type
Information on the social class		skills
Geo-demographic information	CATI	importance of qualification
Expectations on the future after graduation		seeking employment
		earning

understanding if different survey modes generate different answers. Such questions stems from the consideration that being information collected through different survey tools with different peculiarities (CAWI or CATI), bias in answers may arise. On one hand, the presence/absence of interviewers is an important determinant for the quality of the information collected. On the other hand, because of the cultural level of the cohort involved in the interview, the contribution given by the interviewer may be limited: in some instances it may even be counterproductive, since they may influence the answers of the graduates. In consideration of the complexity of the subject that is dealt with, it has become important to determine if there are significant differences between the answers given by those who filled in the online questionnaire and those who gave their answers during the telephone interview (interview mode effect). This need has also been confirmed by the fact that these two groups of graduates turned out to be different in some preliminary analysis; for example, in terms of their studies and area of residence (self-selection effect). Furthermore, as reported in Schonlau et al. (2006) ([22]), internet users are most savvy computer users and therefore may be expected to be much quicker at understanding and answering Internet interviews than others. In order to understand if a mode's effect was present, AlmaLaurea first performed a Propensity Score subclassification and then if a mode effect was present AlmaLaurea in its public reports adopted the adjustment method proposed by Lee(2006) [17]. We apply the multivariate approach to the same data in order to avoid the Propensity Score Model dependence problem and understand if similar results could be obtained compared to PS subclassification. The method consists in finding groups whose pre-treatment characteristics are free of any dependence from the kind of undergone treatment. Therefore, within these groups of respondents any observed difference in the study outcome (e.g. the occupational status) among those who have undergone the CATI-treatment and those who have undergone the CAWI-treatment could be attributed to the interview mode. If an effect of the interview mode on the target variable (e.g. the occupational status) has been detected an adjustment of answers is needed.

3 Methodology

To understand if an effect of the interview mode exists, by dealing with the potential presence of self-selection, we propose the use of the data driven procedure introduced in D'Attoma and Camillo (2011) ([8]). This section begins by giving a brief introduction to the Propensity Score subclassification method adopted by AlmaLaurea and then to the multivariate approach and its notation.

3.1 Propensity Score Subclassification

In Rubin's Potential Outcome Approach, the aim of the resulting Propensity Score is to balance non-equivalent groups on observed pre-treatment covariates in order to reduce bias in causal effect estimation. Rubin demonstrated that, having in hand pre-treatment information that characterizes units under analysis, it is possible to create groups of units with similar pre-treatment characteristics. These groups are,

therefore, theoretically independent from the treatment. Within these groups, one than compares the target variable among those who have undergone the treatment and those who have not.
Consider a population of n units. Denote T the assignment-to-treatment indicator vector ($T = 0/1$). More formally, in Rubin's perspective, each unit i has two potential outcomes, the potential control outcome $Y_i(0)$ under the control condition ($T_i = 0$) and the potential treatment outcome $Y_i(1)$ under treatment condition ($T_i = 1$). After treatment, only one of the two potential outcome is observed, the outcome corresponding to the treatment condition of the i unit ($Y_{i,obs} = T_iY_i(1) + (1 - T_i)Y_i(0)$). Because is not possible to observe both potential outcomes, the causal effect for each unit, defined as $\tau_i = Y_(1) - Y_i(0)$ could not be determined. Thus, at the group level, we can only observe the expected treatment outcomes for the treated, $E(Y_i(1)|T_i = 1)$ and the expected control outcomes for the untreated, $E(Y_i(1)|T_i = 0)$. Therefore, the simple difference in observed groups means:

$$\hat{\tau} = \frac{1}{N_{T_1}} \sum_{i \in T_1} Y_{i,obs} - \frac{1}{N_{T_0}} \sum_{i \in T_0} Y_{i,obs} \tag{1}$$

is a biased estimator for the Average Treatment Effect (ATE) when the assignment-to-treatment is not random and each potential outcome could belong to a different population, where T_1 indexes the set for the treated units, T_0 the set for the control units, N_{T_1} is the number of treated and N_{T_0} the number of control units. Propensity Score is a very popular technique that overcomes such a problem. In particular, with the propensity score e($\mathbf{X}$) defined as the conditional probability of treatment exposure given the observed covariates $\mathbf{X}$, that is $e(\mathbf{X}) = Pr(T = 1 \mid \mathbf{X})$, units under different treatment conditions are comparable if their probabilities to get assigned to one treatment given the observed covariates are the same. Comparability is allowed given that it has been demonstrated that the propensity scores are balancing scores ([20]), meaning that it balances all pre-treatment group differences in observed covariates. Propensity Score (PS) Subclassification can be used to find groups of treated-comparison units with similar characteristics. The estimated PS is used for subclassifying all units into Q homogeneous bins. The underlying rational is that treated-comparison cases within each bin (or stratum) are homogeneous on both the PS and the observed traits.

3.2 The Multivariate Approach to monitor self-selection

Here we suggest the use of the multivariate approach introduced in D'Attoma and Camillo (2011) ([8]) to monitor self-selection applied to the problem of the evaluation of different data collection methods. The key aspect underlying the proposed monitoring system, involves measuring and testing global imbalance under non-experimental conditions. The appendix details the computation of the measure of the Global Imbalance, $GI = \frac{1}{Q}\sum_{t=1}^{T}\sum_{j=1}^{J_Q}\frac{b_{tj}^2}{k_{.t}k_{.j}} - 1$, and the related multivariate imbalance test. Combining these tools we provide a three step strategy for estimate the effect of the interview mode on answers in an unbiased way. The first step involves measuring imbalance via the GI measure and testing the extent to which there is imbalance in data. In other words: are differences between CATI and CAWI groups such that a simple comparison of their answers may be biased by selection? Such difference is measured in terms of between-group inertia, which represents the global measure of imbalance in data. As reported in Peck et al. (2010)([19]) the advantage of this GI measure stems from the consideration that most common variable-by-variable imbalance measures, such as differences in means or in proportions between treatment groups, might fail to detect imbalance since they do not take into account any interactions between or among variables. If imbalance in data exists, within data that demonstrate the presence of selection bias, we proceed to the second analytic step, which involves executing a Cluster Analysis that identifies homogeneous groups on the basis of the continuous Multiple Correspondence Analysis (MCA) coordinates. Using MCA coordinates before clustering exploits the advantage of working with continuous variables (MCA coordinates) rather than categorical covariates (original variables), which need to be treated with unusual metrics. In the third step we assess the balance within Step 2's resulting clusters, computing local effects within balanced groups and pruning observation in unbalanced clusters. In particular, within balanced clusters, the observed differences in the study outcomes will be attributed to the interview mode.

4 Empirical Results

4.1 Multivariate Approach Results

We analyze 26997 respondents [4] of the AlmaLaurea survey on graduates' employment condition: 15749 have been contacted via CAWI method and the remaining via CATI method (table 2).

Table 2 CATI and CAWI respondents

Survey Method	Graduates	%
CATI	11248	41,7
CAWI	15749	58,3
TOTAL	26997	100

The aim is to measure the influence of the interview mode on the answers of respondents. Due to the possible presence of self-selection the ultimate intent is to

[4] All respondents are 2nd level graduates interviewed one year on from graduation

find balanced groups of respondents, whose pre-treatment characteristics (i.e., information on career, information on the family, information on the social class, geodemographic information, expectations on the future after graduation) are free of any dependence from the kind of undergone treatment (CAWI/CATI). To determine if dependence exists, we consider the same 18 categorical pre-treatment covariates used by AlmaLaurea (table 1) and the CAWI/CATI treatment indicator. The main information is related to academic and work experiences made after graduation; but it also concerns socio-demographic characteristics of graduates, pre-university studies, academic studies and further experiences made during studies. Two of the 18 covariates considered were previously discretized (age and graduation mark). To offer a sense of AlmaLaurea database's characteristics, Table 3 shows baseline traits versus interview mode chi-square test. It is clear that dependence between interview mode status and each baseline covariate exists, since across all traits the chi-square is significant with the exception of *gender*. Assessing if the interview mode causes an

Table 3 Variable-by-variable balance checking: covariates vs interview mode

Pre-treatment covariates	Chi-Square value	Prob	Balance
Internet surfing skills	2236.83	$< .0001$	No
Willingness to accept mobility	2169.69	$< .0001$	No
Attended class on a regular basis	2277.03	$< .0001$	No
Intended to pursue postgraduate studies	2274.02	$< .0001$	No
Command of spoken English	2067.13	$< .0001$	No
Command of written English	2172.37	$< .0001$	No
Not yet graduated	2210.40	$< .0001$	No
Did the student study abroad?	2288.79	$< .0001$	No
Social class	2099.60	$< .0001$	No
Pure-hybrid graduates	2275.39	$< .0001$	No
Regular Attendance during studies	185.94	$< .0001$	No
Educational qualification of parents	2179.71	$< .0001$	No
Gender	0.0072	$< .0001$	Yes
Degree-course group	598.86	$< .0001$	No
Geographical area of the University	172.72	$< .0001$	No
Geographical area of residence	112.99	$< .0001$	No
Age at graduation (in class)	451.02	$< .0001$	No
Graduation Mark (in class)	57.02	$< .0001$	No

effect on answers is not easy, and it requires, as highlighted in the previous sections, the effect of the interview mode to be disentangled from the influence of the respondents characteristics. We expect that these differences in characteristics might explain differences in answers that are distinct from the contribution of interview mode, although the expected direction of the bias is not obvious. With this as context, we begin by implementing the three step analysis by computing the GI measure for this data set. As reported in table 4 the resulting value of 0.0531 can be interpreted as demonstrating the presence of imbalance in data. The GI measure falls in the critical region, thereby demanding adjustment in order to estimate the presence of the interview mode effect that is not biased by self-selection. The second step in

Table 4 Global Balance Checking

n	$n_{T=1}$	$n_{T=0}$	GI	Interval	Balance
26997	15749	11248	0.0531	(0; 1,28E-08)	No

our analytic process is to use Cluster Analysis to identify homogeneous groups on the basis of the MCA coordinates. The cluster analysis was carried out on the SAS system employing Ward's algorithm on the MCA coordinates where the proximity between two groups is taken to be the square of the Euclidean distance between them. We most closely examined different clusters solutions with the aim of identify which one appears to meet the criteria of achieving balance in an acceptable number of clusters. We moved from a 2-clusters partition to a 28-clusters partition. Finally, we retain the 28-cluster solution because it provides balance within a suitable number of clusters with fewer pruned observations (around 19%), compared to larger cluster solutions. With the 28-cluster solution, we test balance within each group, again using our computation of the GI and whether it falls in the critical region, as described in the prior step[5]. Table 5 shows the results of the cluster analysis in terms of balance. In this illustration, five of the clusters result in having unbalanced characteristics by our GI measure. In total these five clusters represent about 19 percent of the observations (5264 units) being excluded from the third analytic step. During the final stage of the procedure, we estimate the effect of the interview mode on the outcome variables: employment, contract type, use of skills, importance of graduate, earning, seeking employment, seeking actions. Being all outcomes categorical, for each balanced group we compare the observed frequency of each answer to the corresponding expected frequencies under the hypothesis of independence between the answer and the interview mode (Table 7).

4.2 *Propensity Score Subclassification AlmaLaurea results*

AlmaLaurea adopted a PS Subclassification method to control for the interview mode effect. To estimate the PS a logit model was specified, where the interview mode indicator variable is a function of the observed pre-treatment covariates [6] (Table 3) as in 2:

[5] The procedure to check and test balance is completely automatic. The %Macro Balance could be downloaded at amsacta.cib.unibo.it/2874/1/balance.pdf ([5])

[6] The logit of the estimated PS, also called linear propensity score, is more frequently used than PS itself since the logit is typically more linearly related to the outcome of interest than the PS ([**?**])

Table 5 Balance by clusters

Cluster ID	n	$n_{T=1}$	$n_{T=0}$	GI	Interval	Balance
1	1736	883	853	0.0029	(0;0.003)	yes
2	1236	784	452	0.004	(0;0.0044)	yes
3	1499	735	764	0.0045	(0;0.0039)	no
4	1059	538	521	0.0056	(0;0.0054)	no
5	1122	634	488	0.0036	(0;0.0051)	yes
6	1708	891	817	0.00319	(0;0.0032)	yes
7	751	365	386	0.0056	(0;0.0072)	yes
8	1392	687	705	0.0034	(0;0.0039)	yes
9	829	444	385	0.0051	(0;0.0069)	yes
10	930	471	459	0.005	(0;0.0066)	yes
11	582	314	268	0.010	(0;0.010)	yes
12	550	315	235	0.009	(0;0.011)	yes
13	112	74	38	0.0312	(0;0.05)	yes
14	321	211	110	0.0155	(0;0.018)	yes
15	1043	518	525	0.0052	(0;0.0056)	yes
16	1327	655	672	0.0039	(0;0.0045)	yes
17	575	333	242	0.0130	(0;0.0112)	no
18	693	345	348	0.0061	(0;0.0086)	yes
19	761	420	341	0.0065	(0;0.0081)	yes
20	792	495	297	0.0054	(0;0.0077)	yes
21	245	157	88	0.0171	(0;0.0246)	yes
22	768	404	364	0.0063	(0;0.0079)	yes
23	154	106	48	0.0028	(0;0.041)	yes
24	497	311	186	0.0135	(0;0.0117)	no
25	865	420	445	0.0041	(0;0.0066)	yes
26	677	351	326	0.0059	(0;0.0078)	yes
27	1634	822	812	0.0046	(0;0.0036)	no
28	3139	3066	73	0.0126	(0;0.0182)	yes

$$log[\frac{e(X)}{1-e(X)}] = \alpha + \beta^T f(X) \tag{2}$$

Thereafter, a subclassification on the estimated PS was performed. This was first done sorting units by the estimated propensity score and partitioning units in a predefined number of strata, where each stratum has approximately the same number of units. Based on Cochran(1968) ([7]) they have first divided the estimated range of propensity score in 5 strata [7]. Afterwards, being the common support not satisfied in one of them, they divided the range of the estimated propensity score in 4 strata. One of these resulted unbalanced and units within it were discarded. The effect of the interview mode on the outcome variables was estimated comparing the observed frequency of each answer to the corresponding expected frequencies under the hy-

[7] Based on Cochran results (1968) we may expect a 90% bias reduction for each of the 18 covariates when we subclassify at the quintile of the distribution of the propensity score

pothesis of independence between the answer and the interview mode (Table 7) within each balanced bin (or stratum).

Table 6 Balance within Propensity Score Strata

Stratum	CAWI	CATI	Total	Balance
1	1248	5501	6749	no
2	2810	3939	6749	yes
3	3372	3378	6750	yes
4	3818	2931	6749	yes
Total	11248	15749	26997	

4.3 Results Comparison

Both methods (PS Subclassification and Multivariate approach) provided very similar results. As reported in table 7 the effects of interview mode (CAWI/CATI) on the answers are small: never more than 2%. For expository reasons, we report only results concerning the effect of the interview mode on *Contract Type* and we do not report results by clusters or strata, but only the aggregate results. The aggregate observed frequencies are obtained as the sum of the observed frequencies by clusters (or strata). Differences between observed and expected frequencies have to be interpreted as an aggregate measure of the mode effect. Results shows differences in answers. CATI and CAWI operate in opposite directions: where CAWI *underestimates* the expected frequencies, the CATI *overestimates* them and viceversa. Those differences could be attributable to the fact that the question about the type of contract could be perceived in different ways. In fact, CATI involves an oral and long list of categories; whereas, CAWI allows respondents to analyze and compare each category. Furthermore, using the word *other* to list the *fixed term contract* (Table 7) might generate a wrong perception. Especially with the CATI method it might be intended as a residual category. In sum, both methods do support the conclusion that answers are affected by the interview mode. For such reason AlmaLAurea in their public reports adopted the adjustment proposed by Lee(2006) [17][8].

5 Discussion and Conclusions

The main aim of this work has been to introduce a multivariate monitoring system of self-selection that allows to understand if observed differences in the answers are

[8] The CATI-sample was used as reference sample

Table 7 Differences between observed and expected frequencies

Outcome	PS Approach		Multivariate Approach	
Contract Type	Diff.CAWI	Diff.CATI	Diff.CAWI	Diff.CATI
Permanent	-0.543	0.462	-0.776	0.477
Work/training	1.598	-1.358	1.505	-0.925
Apprenticeship	0.329	-0.279	0.204	-0.126
Contract employment agencies	0.235	-0.199	0.320	-0.196
Continuous and coordinated collaboration contract	1.083	-0.921	1.399	-0.860
Occasional collaboration contract	-0.140	0.119	-0.098	0.060
Socially useful work	0.064	-0.055	0.080	-0.049
Intermittent work	0.063	-0.053	0.096	-0.059
Job sharing contract	0.027	-0.023	0.031	-0.019
Casual incidental work	0.181	-0.154	0.192	-0.118
Other fixed-term contract	-3.102	2.638	-3.136	1.927
Self-employed	-1.056	0.898	-1.038	0.638
Working without a contract	0.272	-0.231	0.237	-0.146
Prestazione d'opera[1]	0.842	-0.716	0.817	-0.502
Pip [1]	0.071	-0.060	0.068	-0.042
Associazione in partecipazione[1]	0.052	-0.044	0.067	-0.041
Non response	0.026	-0.022	0.032	-0.020

[1]Not translated Italian Contract Type Name

attributable to self selection or just to interview mode. We worked through the use of the multivariate approach with an application to AlmaLaurea 2008 survey on Graduates' condition. In conducting such survey AlmaLaurea has to deal with the strong growth in the reference population (almost 300.000 graduates in 2008). As a consequence, the need to reduce survey costs and duration has led to the introduction of a mixed data collection strategy (CATI+CAWI). The adoption of such mixed strategy has been undoubtedly facilitated by the increasing availability of e-mail addresses. However, on one hand the mixed strategy led to a reduction of data collection costs; on the other hand the use of two data collection methods led to the need of implementing a method to control for self-selection into CATI-CAWI treatments. In this paper we have briefly discussed and then compared two alternative methods to control for self-selection: the monitoring system based on the multivariate approach introduced in Camillo and D'Attoma (2010) [4] and the PS subclassification applied by AlmaLaurea, and both led to similar results. We report two main findings corresponding to the research question we posed initially. First, respondents self-select in one of the two interview modes. Second, the interview mode affects answers being all else characteristics of respondents equal. The innovative aspect of the implemented system is that it allows to disentangle the effect mode from self selection effect and thus, to understand if an adjustment of estimates is needed. In this way bias is reduced, or eliminated, also when a mixed data collection mode is adopted. We address that the multivariate approach may overcome the PS model dependence problem and may facilitate the way to control for selection bias being completely model free. We hope to whet the appetite of researchers interested in the problem of self-selection in mixed mode surveys and encourage them to consider the multivariate approach here proposed in future works.

Appendix

The GI measure

D'Attoma and Camillo (2011) ([8]) reports that the between-group inertia of a cloud of units denotes the GI measure expressed as:

$$GI = I_b = \frac{1}{Q}\sum_{t=1}^{T}\sum_{j=1}^{J_Q}\frac{b_{tj}^2}{k_{.t}k_{.j}} - 1 \tag{3}$$

where Q denotes the number of pretreatment covariates, T denotes the number of treatment levels, J_Q denotes the set of all categories of the Q variables considered, b_{tj} is the number of units with category $j \in J_Q$ in the treatment group $t \in T$, $k_{.t}$ is the group size $t \in T$, $k_{.j}$ is the number of units with category $j \in J_Q$. The GI measure is the result of using Conditional MCA([11]) that allows to quantify the between-group inertia. Such a measure originates from the consideration that when the dependence among **X** and T is outside the control of researchers, displaying the relationship among them on a factorial space represents a first step for discovering the hidden relationship. In fact, if dependence among **X** and T exists, any descriptive factorial analysis may exhibit this link. A conventional method dealing with the factorial decomposition of the variance related to the juxtaposition of the X matrix and the T variable is Multiple Correspondence Analysis (MCA) framework. Given that the inertia of a data matrix can be decomposed into eigenvalues and eigenvectors, and referring to MCA for the study of the relationship among variables and of the structure induced by variables on the population, the presence of a conditioning variable (T) could strongly influence the structure of the matrix decomposition process. Hence, a conditional analysis could be useful in order to isolate the part of the variability of the **X**-space due to the assignment mechanism. Conditioning applied to problems arising from the dependence between categorical covariates and an external categorical variable was first studied by Escofier (1988)([11]) with the resulting Conditional Multiple Correspondence Analysis (MCA_{Cond}). Referring to Huygens' overall inertia decomposition of total inertia (I_T) as within-groups (I_W) and the between-groups (I_B), MCA_{Cond} consists in a factorial decomposition of the within-group inertia. MCA_{Cond} could be also considered as an *intra* analysis since the inertia induced by the conditioning variable T is not taken into account.An *inter-group* analysis considers the relative position of groups, whereas and *intra-group* analysis detects and describes differences among units within each group by not considering the effect due to the partition's structure. In the evaluation context, with observational data, this structure is induced by the non random selection process. An intra-analysis allows measuring the influence of conditioning, which means to obtain a measure of comparability between groups. The key result of using MCA_{Cond} is represented by the quantified between-group inertia that represents our measure of global imbalance in data. An assumption underling the approach assumes a crucial role: the assignment mechanism is assumed to be known, which means that the **X** matrix includes all baseline variables associated with both the treatment assignment and the observed outcome.

The Imbalance Test

To determine the significance of the detected imbalance, we perform an Imbalance test. We specify the null hypothesis of no dependence between **X** and T as $H_0 : I_W = I_T$. To establish an interval of plausible values for I_B under the null hypothesis, we use results obtained by Estadella et al. (2005)([12]), who have studied the asymptotic distribution function of I_B as:

$$I_B \sim \frac{\chi^2_{(T-1)(J-1),\alpha}}{nQ} \tag{4}$$

Thus, the interval of plausible values for GI is defined as:

$$GI \in (0, \frac{\chi^2_{(T-1)(J-1),\alpha}}{nQ}) \tag{5}$$

Then, if the GI is outside the interval, then the null hypothesis of no dependence among **X** and T is rejected and data are deemed unbalanced.

References

1. Bethlehem,J.: Selection in Web Surveys. International Statistical Review, **78(2)**, 161–188 (2010). http://www.unisi.it/eventi/dmq2009/presentation.htm
2. Borkan, J.M.: Mixed Methods Studies: A Fundation for Primary Care Research. Annals of Family Medicine, **2(1)**, 4-6, 2004.
3. Camillo,F., Conti, V. and Ghiselli, S.: Impact Evaluation of different data collection methods using causal inference approaches. Proceedings of ITACOMS09, June 10-12, Siena University. (AVAILABLE VIA DIALOG) http://www.unisi.it/eventi/dmq2009/presentation.htm
4. Camillo,F. and D'Attoma, I.: A New Data Mining Approach to Estimate Causal Effects of Policy Interventions. Expert Systems with Applications. **37(1)**, 171–181 (2010)
5. Camillo,F. and D'Attoma, I.: A SAS Macro for measuring and testing global balance of categorical covariates. Available VIA DIALOG.
6. Cammelli, A. Antonelli, G., Camillo, F., Di Francia, A., Ghiselli, S. and Sgarzi, M.: Graduates' employment and employability after the "Bologna Process" reform. Evidence from the

Italian experience and methodological issues. AlmaLAurea Working Papers, 1 , June 2011. http://www.almalaurea.it/universita/pubblicazioni/wp.
7. Cochran, W.G.: The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. Biometrics. **24**, 205–213 (1968)
8. D'Attoma,I. and Camillo, F.: A Multivariate Strategy to Measure and Test Global Imbalance in Observational Studies. Expert Systems with Applications, **38**, 3451–3460 (2011)
9. De Leeuw,E.D.: To mix or not to mix data collection modes in surveys. J.Official Statist., **21**, 233–255 (2005)
10. Dillman, D.A., Sangster, R.L., Tarnai, J. and Rockwood, T.H.: Understanding Differences in People's Answers to Telephone and Mail Surveys. New Direction for Evaluation. **70**, 45–61 (1996)
11. Escofier,B.: Analyse des correspondances multiples conditionelle. In: E.Diday(Ed.), Data Analysis and Informatics. North Holland, Amsterdam: Elsevier Science. 333–342 (1988).
12. Estadella,J.D., Aluja, T. and Thi-Henestrosa, S.: Distribution of the inter and intra inertia in conditional MCA. Computational Statistics. **20**(3), 449–463 (2005).
13. Ho, D.E., Imai, K. and King, G.: Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis. **14**, 131–159 (2006).
14. Jansen, B.: Web data collection in a mixed mode approach: an experiment. Proceedings of Q2006. European Conference on Quality in Survey Statistics.
15. King, G. and Zeng, L.: The Dangers of Extreme Counterfactual. Political Analysis. **15**, 199–236 (2007).
16. Kreuter,F., Presser, S. and Tourangeau,R.: Social desirability bias in CATI, IVR, and Web Surveys.The effects mode and question sensitivity. Public Opinion Quartely **72(5)**, 847–865 (2008).
17. Lee,S.: Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Survey. Journal of Official Statistics. **22(2)**, 329–349 (2006)
18. Mora, M. : Understanding the pros and cons of mixed-mode research. Quirk's Marketing Research Review, 50 (2011).
19. Peck, L.R., Camillo, F. and D'Attoma, I.: A Promising New Approach to Eliminating Selection Bias. Canadian Journal of Program Evaluation **24(2)**, 31–56 (2010)
20. Rosenbaum, P.R. and Rubin, D.B. : The Central Role of Propensity Score in Observational Studies for Causal Effects. Biometrika. **70**, 41–55 (1983)
21. Schonlau,M. , Fricker, R.D. and Elliot, M.N.: Conducting research survey via e-mail and the web. Rand Corporation (2002).
22. Schonlau,M. Van Soest, A., Kapteyn, A. and Couper, M.P., Selection Bias in Web Surveys and the Use of Propensity Score. (Available via DIALOG, 2006).
23. Vannieuwenhuyze, J., Loosveldt, G. and Molenberghs, G. : A Method for Evaluating Mode Effects in Mixed-Mode Surveys. Public Opinion Quartely. **74(5)**, 1027–1045 (2010).
24. Woltman, H. F., Turner, A. G. and Bushery, J. M.: A comparison of three Mixed-Mode Interviewing Procedures in the National Crime Survey. Journal Of the American Statistical Association. **75(371)**, 534–543 (1980)