UNIVERSITÀ DEGLI STUDI DI PALERMO
dSEAS dipartimento scienze economiche aziendali e statistiche
department of economics business and statistics
Working Papers

ISSN 2611-0172, volume II, 2018

A Proposal to estimate the roaming–dog Total in an urban area through a PPSWOR spatial sampling with sample size greater than two.

Ornella Giambalvo · Roberto Sichera

Received: date / Accepted: date

Abstract Dogs roaming in urban areas constitute an issue for public order, hygiene and health. Proper planning of actions for health and security control, and allocation of financial funds require the knowledge of the roaming–dog–population size in a given urban area. Unfortunately, a reliable statistical procedure aimed to measure such population is not available yet in literature. This paper presents a simple, reproducible survey sampling procedure to estimate the number of roaming dogs in an urban area through the description of a real study carried out on a restricted area of the city of Palermo, in southern Italy. A sample of areas is drawn by means of a drawn–by–drawn spatial sampling with probabilities proportional to size and without replacement (PPSWOR). As inclusion probabilities are not available in closed form, they are estimated by Monte Carlo approach, which is of simple implementation and permits design–based variance estimation even when first–order inclusion probabilities are unknown.

Keywords PPS sampling · design–based variance estimation · inclusion probability · Monte Carlo estimation · roaming dogs

O. Giambalvo
Dipartimento di Scienze Economiche, Aziendali e Statistiche
Università degli Studi di Palermo
viale delle Scienze ed. 13, 90128
E-mail: ornella.giambalvo@unipa.it

R. Sichera
Dipartimento di Scienze Economiche, Aziendali e Statistiche
Università degli Studi di Palermo
viale delle Scienze ed. 13, 90128
E-mail: roberto.sichera@unipa.it

Riassunto *I cani che vagano sul territorio possono rappresentare un rischio per l'uomo in diversi modi: aggressioni, causa di incidenti stradali, rischio igienico. Azioni per il contenimento della crescita di questi gruppi di animali si rendono dunque necessarie. In Italia, la strategia più in uso è quella della cattura, sterilizzazione e successivo rilascio nel territorio dei randagi. Tuttavia, tale strategia ha successo solo se viene applicata all'intera popolazione randagia (Jackman and Rowan, 2007). A tal fine dunque, le autorità sanitarie hanno la necessità di conoscere il numero di randagi presenti sul territorio di competenza. Nonostante in passato siano stati effettuati alcuni tentativi di quantificazione del fenomeno, attualmente in letteratura non si dispone di una metodologia statistica adeguata.*
In questo articolo presentiamo un'indagine campionaria per la stima del totale di cani randagi tramite uno studio condotto nella prima circoscrizione della città di Palermo. I dati sono stati raccolti su un campione di aree selezionato per mezzo di un campionamento probabilistico con probabilità variabili. Poiché per il disegno campionario adottato le probabilità di inclusione non sono ottenibili in maniera esatta, queste sono state stimate tramite simulazione Monte Carlo e l'algoritmo di Bennett (Fattorini, 2009).

Parole Chiave *Campionamento a probabilità variabili – Stima della varianza – Probabilità di inclusione – Stima Monte Carlo – Cani randagi*

1 Introduction

The presence of roaming dogs in urban areas is an important issue that municipalities have to deal with, especially in big cities of Southern Italy where such phenomenon is widespread and represents a source of risk for people and public security, as well as being a concern for public hygiene. In these areas, policy-makers need to undertake actions focused on keeping under control the growth of such population and aimed at reducing its size in the long run. In Italy, the main control technique applied for this purpose is the capture–sterilization–release procedure which, to be effective, needs to be applied to most of the roaming dogs populating a given area (Jackman and Rowan, 2007). Estimating the roaming–dog–population size is thus necessary; however, despite its importance, the size estimation of roaming dogs is a little–covered subject.

Different sources presented figures about this phenomenon, but the methodology used is either omitted (partially or completely) or not statistically sound. One of the attempts made in Italy to find a measure for this phenomenon is a 2006 census arranged by the Italian Ministry of Health (Italian Ministry of Health, 2006). In accordance with this study, in 2006 about $590,000$ roaming dogs were evaluated to be in Italy. However, only the dogs that spent some time in public kennels were considered and the estimation/counting procedures are unknown. More data have been released by Italian regions and animal associations, but without relevant results. The *World Society for the Protection of Animals* (WSPA, now *World Animal Protection*) released a report titled *Surveying roaming dog populations: guidelines on methodology*, where a formal procedure to get an estimate of roaming dogs is proposed. WSPA suggests a way to elicit a sample of sub–areas from the area of interest and find an estimate of the total number of roaming dogs in that territory by means of a simple random sampling.

In this paper, we describe a study carried out in the city of Palermo, Italy, in collaboration with the local health authority. The study consisted in two subsequent ad–hoc surveys and had two aims: first, to formalize a simple and reproducible survey sampling procedure, in order to obtain a statistically reliable estimate and be able to monitor the phenomenon through time in a consistent and comparable way; second, to provide local authorities with an estimate of the roaming–dog–population size for planning proper prevention measures.

Section 2 describes the characteristics of the two surveys and of the data collection procedure. Sampling procedure and estimation methodology are introduced in Section 3, while results are presented in Section 5.

2 Features of the survey and data collection

The study started in the beginning of the 2010s with a pilot survey aimed at setting up an effective methodology for the study. We first defined assumptions concerning the distribution of the dogs over territory and their presence along the day considering, together with veterinarians of the local public health authority, their nature and habits.

These assumptions had naturally driven to the choices that defined the sampling design illustrated in this section.

However, during this survey a few data collectors were not accurate in their task, leading to some ambiguous observations that could not be correctly identified and thus to an inflation of the estimates.

As our main focus was to assess the effectiveness of the method, a few years later we set up a new survey with the same characteristics, except for data collection, which we updated with some modifications in order to remove, or at least reduce, non–sampling errors introduced by data collectors' imprecision.

The data collection procedure that we are going to describe is in some aspects similar to the method proposed in WSPA's report (2010).

2.1 Data collection settings and procedure

The surveys were carried out in the first district of Palermo, Sicily, a 249.7–hectares–large area, which was divided into 76 sub–areas, each large about 3 hectares and classified in *green–area* (e.g. parks), *open–area* (e.g. parkings and squares) or *other* (e.g. streets), in accordance with its internal pattern.

The number of roaming dogs was observed in a sample of 12 sub–areas. Data collection occurred during the second week of June, due to its stable weather; per each area in the sample, data have been gathered at 13:00 and 20:00 on Monday, Thursday and Sunday. These choices are in line with World Organization for Animal Health (2010) guidelines and they are aimed to maximize the chances of observing all roaming dogs which dwelt the sample areas.

At the times indicated, each person in charge for data collection visited the assigned location, where they followed a given path. Paths were designed to have approximately equal length among all areas and had been kept fixed over all observations. Data collection sessions lasted at least thirty minutes and the path were completed at least twice throughout each observation session. Data collectors were veterinarians and volunteers of the public local kennel, as well as components of local animal associations. In order to have consistent data entries over all the observations, each person was assigned to one specific area. Although being advised of the importance of collecting data with careful attention to any distinctive trait of the observed dogs and being asked to report all of them in the form, during the first survey, some of the field workers partially or completely neglected this task. The result was a number of observations that were not clearly ascribable to an unique dog.

The second survey was carried out following the same protocol, the only addition made to this procedure was taking a photo of the observed roaming dogs, made possible by the larger diffusion of devices capable of taking pictures of sufficient quality.

2.2 Data collection form

In both surveys, each data collector received six forms to fill in with observed information, one per each observation time.

The form used in the first survey included information about gender, size and health conditions (good, bad) of the dogs observed, as well as a section for the observer to write down any other significant information he would have found. Because the last point had been quite neglected, some changes were implemented in the second survey. Dog coat colour, a field to indicate whether or not a picture of the observed dog had been taken and one to write down the picture file name were added to the observed variables. Furthermore, the section dedicated to additional information was reorganised to be easier to fill in and highlighted to be simply reminded to the field workers. In addition, a meeting was held with veterinarians and future field workers in order to explain the details of the data collection procedure.

Applying these simple changes, we achieved a much higher quality of data, with no more issues in identifying different observations of the same dog over different times, and thus improving precision. The form of the second survey is shown in Appendix B.

3 Sampling procedure and Estimation

From the population $U = \{1, \ldots, N\}$ of $N = 76$ areas, a sample $s = 1, \ldots, n$ of $n = 12$ units has been selected by means of a drawn–by–drawn spatial sampling scheme with selection probabilities proportional to a size variable X and without replacement (PPSWOR). The selection probability of contiguous units is reduced by means of a parameter β. This design, called PPS

FPDUST, draws a sample according to the following procedure (Barabesi et al., 1997; Fattorini, 2009):

1. Select first unit with probability
$$p_i = \frac{x_i}{\sum_{i \in U} x_i}$$

2. Select k–th unit, $k = 2, \ldots, n$, with probability
$$p_{i|i_1,\ldots,i_{k-1}} = \begin{cases} \frac{(1-\beta)x_i}{\sum_{i \notin s_k} x_i - \beta \sum_{i \in c(s_k)} x_i}, & i \in c(s_k) \\ \frac{x_i}{\sum_{i \notin s_k} x_i - \beta \sum_{i \in c(s_k)} x_i}, & \text{otherwise} \end{cases} \tag{1}$$

   where x_i are the values of the size variable, $i = 1, \ldots, N$, $s_k = \{i_1, \ldots, i_{k-1}\}$ is the set of units selected up to step k and $c(s_k)$ is the set of non–sampled units which are contiguous to at least one sampled unit. β is a parameter that modifies selection probabilities for units belonging to $c(s_k)$ and must take values in the interval $(-\infty, 1)$ to ensure the existence of the design. When $\beta = 0$ one obtains the simple random sampling.

In our study, the size variable X is the walkable area of each unit, that is, the total area excluding buildings and closed spaces such as private parks and parkings (where dogs can not be found). We set $\beta = 0.5$, in order to penalize selection of contiguous units, but not too much because sampling close units would have resulted in reduced costs.

Estimation of the number of roaming dogs is performed through the Horvitz–Thompson total estimator (Horvitz and Thompson, 1952):

$$\hat{Y}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i}, \tag{2}$$

where y_i are the values of the variable of interest, $i = 1, \ldots, n$, in the selected sample s. The variance of 2 is estimated by

$$\hat{v}\left(\hat{Y}_{HT}\right) = \sum_{i \in U} \frac{1-\pi_i}{\pi_i} y_i^2 + 2 \sum_{i \in U} \sum_{j > i} (\pi_{ij} - \pi_i \pi_j) \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} \tag{3}$$

However, estimators (2) and (3) require π_i and π_{ij}, the first and second–order inclusion probabilities, which cannot be directly computed under the sampling scheme adopted.

4 Estimation of inclusion probabilities

An effective way to estimate these quantities is by Monte Carlo simulation (Fattorini, 2006; Thompson and Wu, 2008), where estimation of inclusion probabilities is carried out by performing K independent sample drawings from the target population according to the desired sampling scheme. First–order inclusion probabilities π_i are then estimated as proportion of the number of occurrencies of unit i in the K trials:

$$\tilde{\pi}_i = \frac{I_i + 1}{K + 1} \tag{4}$$

while the second–order inclusion probabilities π_{ij} are estimated by the proportion of each couple of units (i, j) in the K replications

$$\tilde{\pi}_{ij} = \frac{I_{ij} + 1}{K + 1} \tag{5}$$

where I_i and I_{ij} are, respectively, the number of occurrences of unit i and of couple (i, j) over the K replications. Both numerator and denominator are incremented by one unit to ensure strict positivity of the estimators.

Estimates of the Total and its variance may now be obtained by substituting (4) and (5) in equation (3):

$$\tilde{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\tilde{\pi}_i}\,, \tag{6}$$

$$\tilde{v}\left(\hat{Y}_{HT}\right) = \sum_{i \in U} \frac{1 - \tilde{\pi}_i}{\tilde{\pi}_i} y_i^2 + 2 \sum_{i \in U} \sum_{j > i} (\tilde{\pi}_{ij} - \tilde{\pi}_i \tilde{\pi}_j) \frac{y_i}{\tilde{\pi}_i} \frac{y_j}{\tilde{\pi}_j}\,. \tag{7}$$

Fattorini (2006) showed that Monte Carlo estimates converge to the true values as K increases. Although this approach may be extremely computationally demanding, he showed that a few million replications may be sufficient to obtain reliable estimates.

Moreover, Fattorini (2009) proposed an adaptive algorithm for the estimation of Monte Carlo inclusion probabilities, which runs until a given threshold for the precision of the estimates is reached. Given an arbitrary $\varepsilon > 0$, the algorithm ensures that the probability of the relative difference between the Monte Carlo estimate $\tilde{Y}_{HT}$ and the exact Horvitz–Thompson estimate $\hat{Y}_{HT}$ is larger than ε with probability α:

$$P\left\{ \frac{|\,\tilde{Y}_{HT} - \hat{Y}_{HT}\,|}{\hat{Y}_{HT}} > \varepsilon \right\} = \alpha\,.$$

Using Bennett inequality (Bennett, 1962), Fattorini, 2009 shows that

$$P\left\{ \frac{|\,\tilde{Y}_{HT} - \hat{Y}_{HT}\,|}{\hat{Y}_{HT}} > \varepsilon \right\} \leq 2 \sum_{i \in s} e^{-c(\varepsilon) M \pi_i}\,,$$

where

$$c(\varepsilon) = \left(\frac{\varepsilon}{2 + 2\varepsilon} + 1\right) \ln\left(\frac{\varepsilon}{2 + 2\varepsilon} + 1\right) - \frac{\varepsilon}{2 + 2\varepsilon}$$

and $M > (2+2\varepsilon)/(\varepsilon\,\pi_0)$, with $\pi_0 = \min_{i\in s} \pi_i$. The algorithm takes the name of Bennett algorithm after this last result.

Denote with L the number of Monte Carlo replicates to perform at each step $j = 1, 2, \ldots,$ and with δ the maximum acceptable error for the estimates, such that

$$\frac{|\,\tilde{Y}^{j}_{HT} - \hat{Y}^{j-1}_{HT}\,|}{\hat{Y}^{j-1}_{HT}} < \delta \qquad \text{and} \qquad \frac{|\,\tilde{v}^{j} - \tilde{v}^{j-1}\,|}{\tilde{v}^{j-1}} < \delta\,. \tag{8}$$

Moreover, denote with $K \geq 1$ the minimum acceptable number of consecutive steps that satisfy (8). Then, the Bennet algorithm can be implemented as described below:

1. Choose some values for ε, α, δ, K and L;

2. Compute
$$c(\varepsilon) = \left(\frac{\varepsilon}{2+2\varepsilon} + 1\right) \ln\left(\frac{\varepsilon}{2+2\varepsilon} + 1\right) - \frac{\varepsilon}{2+2\varepsilon}\,;$$

3. For $j = 1, 2, \ldots$:

   (a) Draw L Monte Carlo samples from the population U according to the original sampling scheme and compute the Monte Carlo estimates $\tilde{\pi}_i$ and $\tilde{\pi}_{ij}$;

   (b) Compute $\tilde{Y}_{HT}$ and $\tilde{v}\left(\hat{Y}_{HT}\right)$ as in (6) and (7), respectively;

   (c) Compute $\hat{M} = (2+2\varepsilon)/(\varepsilon\,\pi_0)$, and $\hat{P} = 2\sum_{i\in s} e^{-c(\varepsilon)\hat{M}\tilde{\pi}_i}$, with $\pi_0 = \min_{i\in s} \tilde{\pi}_i$;

   (d) Define
$$k(j) = \begin{cases} k(j)+1, & \text{if } \frac{|\,\tilde{Y}^{j}_{HT} - \hat{Y}^{j-1}_{HT}\,|}{\hat{Y}^{j-1}_{HT}} < \delta \text{ and } \frac{|\,\tilde{v}^{j} - \tilde{v}^{j-1}\,|}{\tilde{v}^{j-1}} < \delta \\ 0, & \text{otherwise} \end{cases}$$

4. If $k(j) \geq K$, $L \times j \geq \hat{M}$, and $\hat{P} \leq \alpha$ stop the algorithm, otherwise return to step 3.

In our study, first and second–order inclusion probabilities were estimated through Bennett algorithm with $\varepsilon = 0.005$, $\alpha = 0.005$, $\delta = 0.0001$ and $L = 1000$.

Computations were performed through the R software and the packages sampling[1], jipApprox[1], bootstrapFP[1], fpdust[2], and robustHT[2]. The Bennett algorithm was implemented through an R function, available in Appendix A.

[1] Available on CRAN: https://cran.r-project.org
[2] Available on GitHub: https://github.com/rhobis

Table 1 Sample data and estimates for the roaming dogs surveys carried out in the two surveys.

ID	x_i	$\tilde{\pi}_i$	Y_1	Y_2
1	3.17	0.31	5	6
2	2.74	0.29	15	7
3	1.20	0.13	5	1
4	1.13	0.14	3	0
5	1.14	0.14	1	3
6	2.94	0.24	14	9
7	1.46	0.16	1	2
8	2.07	0.22	6	4
9	1.07	0.12	36	13
10	2.66	0.20	8	2
11	1.28	0.15	6	5
12	1.14	0.12	12	1
		$\tilde{Y}_{HT}$	700	300
		$\sqrt{\tilde{v}(\hat{Y}_{HT})}$	227.5	81.8

5 Results

Table 1 presents sample data for the two studies. The table reports the unit ID, the values x_i of the size measure X (the walkable area of each unit), the Monte Carlo inclusion probabilities $\tilde{\pi}_i$, and the observed number of roaming dogs in the two surveys. The two last rows of Table 1 show the estimates of the Horvitz–Thompson total and of its standard error.

It can be seen that the number of observed dogs in the first survey presents some values that are noticeably higher than others, in particular unit 9. As we mentioned earlier, indeed, the first study was affected by errors during data collection, which nonetheless seem to have been successfully reduced in the second survey by the modifications applied to the data collection form. In fact, besides a general reduction in the presence of roaming dogs, which was likely caused by the interventions of local health authorities, the observed values seem more regular. Unit 9 still exhibits an unusually high value, however it is now much closer to the general distribution of the observations. By analysing data collection forms it seems to be a correct outlier, probably due to a particular concentration of dogs in that area.

The estimated total number of roaming dogs in the first district of Palermo, Italy, more than halved in the four–year–long period between the two studies. The reduction appears too large to be due only to outliers. It seems that the interventions made by the local health authorities to stop the spread of the dog population were effective.

Table 2 Sample data and estimates for the roaming dogs surveys, considering a reduced sample with area "3c" excluded.

ID	x_i	$\tilde{\pi}_i$	Y_1	Y_2
1	3.17	0.29	5	6
2	2.74	0.27	15	7
3	1.20	0.12	5	1
4	1.13	0.12	3	0
5	1.14	0.12	1	3
6	2.94	0.22	14	9
7	1.46	0.15	1	2
8	2.07	0.20	6	4
10	2.66	0.18	8	2
11	1.28	0.13	6	5
12	1.14	0.11	12	1
		$\tilde{Y}_{HT}$	444	212
		$\sqrt{\tilde{v}(\hat{Y}_{HT})}$	68.9	28.6

5.0.1 Correction for non–sampling errors

As mentioned in Section 2.1, during the first survey some observations were collected poorly, which caused a number of dogs not to be clearly identifiable among different observations on different data collection times. Most of such imprecisions concentrated in one single area, labelled as "3c", which was very likely to host a much lower number of dogs than the one obtained by the data collection forms. In the particular case of area "3c", we isolated some groups of observations that likely referred to the same dog, but due to said lack of precision, we could not unequivocally identify them, so that the observed roaming dogs amounted to 36.

A closer inspection to area "3c" shows that, in spite of being the smallest area in the sample, it accounts for the highest number of observations in both studies. This high concentration of dogs may be influenced by geographical reasons such as its proximity to a park and to the sea, an unusual availability of food or other causes. The estimates on obtained on the reduced sample where area "3c" is excluded are reported in Table 2, and show that even in this scenario the estimated number of roaming dogs for the first survey is twice as much as the estimate for the second one.

6 Conclusions

The widespread of roaming–dog populations in urban areas is an important issue for local administrations. Despite being essential for a proper planning of health and security actions, no reliable and reproducible procedure for the estimation of such phenomenon have been adopted yet.

With this study we have proposed a first attempt to define a simple, reproducible survey sampling procedure to estimate the total of roaming dogs on an urban area. The procedure proposed was first implemented through a pilot survey on a reduced portion of the city of Palermo, Italy. Some imprecisions in data collection were identified, so a few years later the study was replicated with same characteristics and an enhanced data collection form.

A sample of sub–areas from the first district of Palermo has been drawn by means of a drawn–by–drawn spatial sampling with selection probabilities proportional to a size variable. For this sampling design exact inclusion probabilities are not available, however they can be easily estimated through Monte Carlo simulation and the Bennett algorithm.

Results appeared satisfactory to the local health authorities, who found the estimates to be close to their expectations. A reduction in the estimated size of the roaming–dog population was registered over the time period between the two studies. This was expected due to the interventions promoted by the health authority after the results of the first survey.

However, further studies are required to improve the efficiency of the survey, for example by defining stratification variables and better size measures that better capture the nature of the phenomenon. Also, more recent spatial sampling designs might be employed, such as the *doubly balanced spatial sampling* (Grafström and Tillé, 2012). More work is also needed to extend the survey to larger areas; this could be done, depending on available human and financial resources, either by including all the city districts in the study or by means of a two–stage design with districts as primary units and their sub–areas as second–stage units.

A R function for the Bennett algorithm

The R function written to perform the Bennett algorithm is reported below. It requires package jipApprox, available on CRAN. The input objects required are the vectors of sample values of the target and auxiliary variables, a vector with the indices of sampled units, the sampling design that produced the original sampling, a list with the arguments to pass to the sampling design function and the values of L, K, δ, ϵ and α.

```
bennet <- function(y, x, sample_labels, design, design_pars=NULL, K=2,
                   delta=0.001, eps=0.1, alpha=0.1 ){
n <- length(y)
k <- 0
i <- 0
P <- 100
M <- Inf
occurrences <- matrix(1, n, n)
ce   <- (eps/(2+2*eps) + 1) * log( eps/(2+2*eps) + 1 ) - eps/(2+2*eps)

while( k<K | L*i<M | P>alpha ){
    i    <- i+1
    occurrences <- (occurrences - 1) +
                   (L+1)*jipApprox::jip_MonteCarlo(x, n, replications = L, design,
                                sample_labels, seed = NULL, as_data_frame = FALSE,
                                design_pars, progress_bar = FALSE )
    pikl <- occurrences/(L*i+1)
    pik  <- diag(pikl)
    ht   <- drop(crossprod(y, 1/pik)) # Horvitz-Thompson total estimator
    yy   <- outer(y,y)
    pp   <- 1/outer(pik, pik) - 1/pikl
    vv   <- sum(yy*pp)                # Horvitz-Thompson variance estimator
    pi0  <- min(pik)
    M    <- 16*log(2/alpha)*eps^(-2)/(pi0)
    P    <- 2 * sum( exp(-ce * M * pik))
    if( i>1 ){
        unbiasedness <- abs(ht - ht0)/ht0 < delta
        precision    <- abs(vv - vv0)/vv0 < delta
        k <- ifelse( unbiasedness & precision, k+1, 0)
    }
    ht0 <- ht
    vv0 <- vv
}

return( list( replicates = L*i, k = k, tot = ht, var = vv, probs = pikl ) )
}
```

B Data collection form

Figures 1 and 2 show the form used to collect data in the second survey.

Compilare in stampatello

Rilevatore: Nome ______________________

Cognome ______________________

Data: ____________

Ora inizio: ________

Ora fine: ________

Sede Rilevazione: ____1B____

Presenza rifiuti ‘liberi’ ☐ SÌ ☐ NO

Condizioni meteorologiche ______________________

	Taglia	Sesso	Stato di Salute	Colore	Segni particolari (se presenti, specificare sul retro)	Foto
1	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
2	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
3	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
4	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
5	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
6	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
7	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
8	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
9	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
10	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
11	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
12	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
13	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
14	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
15	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
16	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
17	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
18	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________
19	☐P ☐M ☐G	☐M ☐F ☐NS	☐S ☐M ☐NS	☐N ☐B ☐M ☐Altro__________	☐Sì ☐No	☐No ☐Sì ________

Fig. 1 Data collection form used in the second survey - front

Legenda

Taglia: P = piccolo, M = medio, G = grande;

Sesso: M = maschio, F = femmina, NS = non so;

Stato di salute: S = sano, M = malato, NS = non so

Colore: N = nero, B = bianco, M = marrone, Altro;

Età: A = adulto, C = cucciolo;

Segni particolari: qualsiasi informazione aggiuntiva (razza, microchip, …);

Foto: Specificare se è stata scattata una foto o meno. In caso affermativo, inserire il nome del file nello spazio apposito accanto al campo "Sì".

Segni particolari:

1. ____
2. ____
3. ____
4. ____
5. ____
6. ____
7. ____
8. ____
9. ____
10. ____
11. ____
12. ____
13. ____
14. ____
15. ____
16. ____
17. ____
18. ____
19. ____

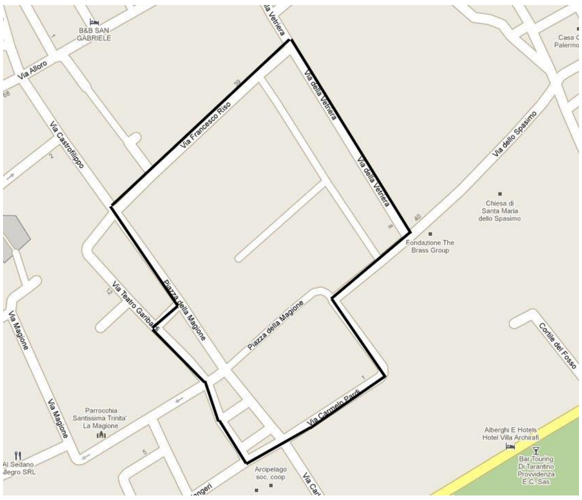

Fig. 2 Data collection form used in the second survey - back

Acknowledgements The paper is the result of the productive collaboration between the authors. In particular, sections 1 and 2 can be ascribed to Ornella Giambalvo, and sections 3 and 4 to Roberto Sichera. The authors wish to thank Dr Francesco Francaviglia for the support provided in the organization of the two surveys.

References

Barabesi, L., L. Fattorini, and G. Ridolfi (1997). "Two-phase surveys of elusive populations". In: *Proceedings of the Statistic Canada Symposium*. Vol. 97, pp. 285–287.

Bennett, G. (1962). "Probability Inequalities for the Sum of Independent Random Variables". In: *Journal of the American Statistical Association* 57.297, pp. 33–45. DOI: 10.1080/01621459.1962.10482149.

Fattorini, L. (2006). "Applying the Horvitz: Thompson Criterion in Complex Designs: A Computer-Intensive Perspective for Estimating Inclusion Probabilities". In: *Biometrika* 93.2, pp. 269–278.

Fattorini, L. (2009). "An adaptive algorithm for estimating inclusion probabilities and performing the Horvitz–Thompson criterion in complex designs". In: *Computational Statistics* 24.4, p. 623. DOI: 10.1007/s00180-009-0149-9.

Grafström, A. and Y. Tillé (2012). "Doubly balanced spatial sampling with spreading and restitution of auxiliary totals". In: *Environmetrics* 24.2, pp. 120–131. DOI: 10.1002/env.2194.

Horvitz, D. G. and D. J. Thompson (1952). "A generalization of sampling without replacement from a finite universe". In: *Journal of the American statistical Association* 47.260, pp. 663–685.

Italian Ministry of Health (2006). *Rendicontazione annuale sul randagismo*.

Jackman, J. and A. N. Rowan (2007). "Free–roaming dogs in developing countries: The benefits of capture, neuter, and return programs". In: Washington, DC: Humane Society Press., pp. 55–78.

Thompson, M. E. and C. Wu (2008). "Simulation-based randomized systematic PPS sampling under substitution of units". In: *Survey Methodology* 34.1, pp. 3–10.

World Organization for Animal Health (2010). *Stray dog population control*.

World Society for the Protection of Animals (2010). *Surveying roaming dog populations: guidelines on methodology*.