The Annals of Applied Statistics
2013, Vol. 7, No. 4, 2293–2314
DOI: 10.1214/13-AOAS672

HETEROGENEITY AND BEHAVIORAL RESPONSE IN CONTINUOUS TIME CAPTURE–RECAPTURE, WITH APPLICATION TO STREET CANNABIS USE IN ITALY

BY ALESSIO FARCOMENI AND DARIA SCACCIATELLI

Sapienza—University of Rome and University of Rome—Tor Vergata

We propose a general and flexible capture–recapture model in continuous time. Our model incorporates time-heterogeneity, observed and unobserved individual heterogeneity, and behavioral response to capture. Behavioral response can possibly have a delayed onset and a finite-time memory. Estimation of the population size is based on the conditional likelihood after use of the EM algorithm. We develop an application to the estimation of the number of adult cannabinoid users in Italy.

1. Introduction. Capture–recapture experiments have been adopted in a wide range of applications, including ecology, agriculture and veterinary science, public health and medical studies, software engineering, behavioral research and, in general, in the estimation of the size of hidden populations. Detailed reviews can be found in the International Working Group for Disease Monitoring and Forecasting (1995a, 1995b) and Amstrup, McDonald and Manly (2003). Capture–recapture experiments are based on repeatedly capturing subjects over time. The counting process of captures is modeled so as to obtain an estimate of the number of subjects never captured or, equivalently, of the size of the catchable population. In discrete time there is a fixed number of capture occasions. In continuous time each subject is at risk of capture in any moment of a fixed-length period.

In this paper we focus on continuous time models, motivated by a large scale study on the population of drug users in Italy. The development of new policies for tackling drug abuse is considered a concern in the European Union. There is extremely limited information on the number of cannabis users in Italy. Population size estimates are usually based on indirect estimation methods, for instance, through chemical analyses of waste waters [European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (2009)]. Our formal capture–recapture experiment will provide a direct estimate at least for the catchable subpopulation. Note that closed population capture–recapture models are often used for estimation of the number of illicit drug users [Böhning et al. (2004), Bouchard and Tremblay (2005), Chiang et al. (2007), Khazaei et al. (2012), Vaissade and Legleye (2009)]. For a fixed time period, beginning with the introduction of a new law on drug control, officers of different Italian police departments identified and reported drug

Received November 2012; revised July 2013.

Key words and phrases. Behavioral response, capture–recapture, drug abuse, frailty, heterogeneity, Horvitz–Thompson estimator.

users. According to the new law, Art. 75, possession of even a minimal quantity of any drug is a crime, which can be punished with administrative sanctions. We have access to the entire database, to the date of each capture, but only to part of the information relating the drug user (e.g., sex, age, province), and no information about the officer. The database records information on different substances (cannabis, cocaine, heroin, other). In this paper we focus on cannabis, whose prevalence of use in public places is the largest and is in most cases the first drug used. See, for instance, Antidrug Policies Department (2009) and Rey, Rossi and Zuliani (2011). The study of cannabis use is particularly important for planning prevention, evidence-based interventions and understanding how abuses should be handled. Furthermore, drug dealers often sell multiple drugs and the number of drug dealers is expected to be proportional to the most requested drug [Bouchard and Tremblay (2005), Reuter and Kleiman (1986)]. Therefore, a study on the prevalence of cannabis may also be useful in planning actions for tackling the illicit drug market.

Subjects were continuously at risk of being captured for the entire observation period. As a consequence, we have a continuous time capture–recapture experiment with observed covariates, possibly unobserved heterogeneity, and time-heterogeneity. It is also reasonable to allow for the possibility of arrests having consequences leading to modifications of the future risk of capture, therefore having a behavioural component in the model. Our catchable population is made of cannabis users in Italy who use, buy and/or carry less than 500 mg of drug in the streets. It shall be noted that frequent and long-term users mostly belong to this population, and these are at higher risk of health consequences [e.g., Semple, McIntosh and Lawrie (2005)]. Drug use and/or dealing in the streets causes a degree of public nuisance [e.g., de Jong and Weber (1999)], generates stress and increases the risk of psychological distress even to subjects that are not involved [Mirowsky and Ross (2003)]. We will also provide indirect estimates of the entire population of cannabis users, that is, also including subjects who use cannabis in private houses and never buy or carry it in the streets.

With assumptions of time-homogeneity and absence of behavioral response to capture, one could simply work with the individual number of captures at the end of the observation period [e.g., Chao (1987)]. When any of these two assumptions may not be met or shall be verified, the conditional hazard function shall be explicitly modeled through a Cox-type model [Cox (1972)]. There are only few, but very neat, works dealing with continuous time capture–recapture models along these lines. The most relevant to our application are Hwang and Chao (2002), who allow for time-heterogeneity, covariates and a behavioral response; and Xi, Yip and Watson (2007), who also included a frailty in the Cox-type model but no covariates. In summary, the available approaches do not allow for simultaneous modeling of two different kinds of individual heterogeneity: the observed one, that can be modeled with covariates, and the unobserved unmeasured heterogeneity, that shall be modeled through a random frailty component. Another limitation of the available

models is that the behavioral effect necessarily implies a *long-term* memory: the hazard function is multiplied by a fixed factor after first capture, and this factor does not vary over time. We believe that, especially in our application in which a capture corresponds to a police officer dealing with a drug user, individual response to capture could be more complex. We generalize the long-term memory assumption in two directions: first of all, we allow for a delayed onset of the behavioral response; second, we allow this effect to disappear after a fixed number of additional captures, after a fixed time frame from the onset, or after the minimum between the two conditions. Not surprisingly, we will estimate a peculiar but easily interpretable behavioral response on our data.

We denote our most general model by M_{hotb}, where the h stands for the unobserved heterogeneity, o for the observed heterogeneity, t for the time-heterogeneity and b for our flexible behavioral response to capture. We use, as is common practice, an unspecified baseline in the hazard function, thus nonparametrically modeling time-heterogeneity. Unlike available models, we explicitly distinguish and allow for the two different sources of individual heterogeneity. For estimation, we employ the conditional likelihood (and Horvitz–Thompson estimator). We set up a general expectation-maximization (EM) algorithm in order to maximize the conditional likelihood, after employing an augmentation scheme to tackle the issue of unknown frailty terms, and drastically reducing the number of parameters involved by obtaining an implicit estimate of the nonparametric baseline. Notably, we also obtain a closed form expression for the integrals involved, thus avoiding numerical integration at the E-step. Our model extends Xi, Yip and Watson (2007) to the use of covariates. It can also be seen as an extension of the model of Hwang and Chao (2002) to inclusion of unobserved heterogeneity. Finally, given our flexible behavioral response to capture, we extend all previous works in this direction in the spirit of the work of Farcomeni (2011), who proposes a general M_b class of models in discrete time. The rest of the paper is as follows: in the next section we describe our M_{hotb} model and discuss all submodels in the class. In Section 3 we outline generalized behavioral responses. In Section 4 we show how to perform inference on model parameters. We describe and analyze our data in Section 5, and state conclusions in Section 6.

2. General recapture models in continuous time. Suppose we have a population of N subjects, at risk of capture in the time interval $[0, \tau]$. We assume the population is closed, that is, there are no new users, no quitters and no migration during the observation interval. We will discuss below the implications of these assumptions for the data at hand. A common understanding is that if τ is small enough, this assumption is safe. Let $N_i(t)$, $i = 1, \dots, N$, denote the number of times the ith subject has been captured in the time interval $[0, t]$, $0 \le t \le \tau$. We have n subjects for which $N_i(\tau) > 0$, with captures at time t_{ij}, $i = 1, \dots, n$; $j = 1, \dots, N_i(\tau)$.

Each $\{N_i(t); 0 \le t \le \tau\}$ is then a continuous time counting process, along the lines of Hwang and Chao (2002) and Xi, Yip and Watson (2007). We denote its intensity function, conditionally on the capture history up to time t, by $\lambda_i(t)$. Note that $\{N_i(t); 0 \le t \le \tau\}$ would be a Poisson process under an assumption of absence of memory, that is, absence of behavioral effects. Let Z_i denote a vector of subject specific covariates, of size p, which is known only for the subjects captured at least once. Our general M_{hotb} model can be stated as follows:

$$\lambda_i(t) = \rho_i e^{\beta' Z_i} \phi^{I(N_i(t) \ge 1)} \omega(t), \tag{2.1}$$

where $I(C)$ is the indicator function for condition C. In (2.1), β denotes a vector of log hazard-ratio coefficients describing the observed heterogeneity; $\omega(t)$ an unspecified nonnegative baseline function describing the time-heterogeneity; ϕ measures a proportionality effect on the risk after first capture, hence quantifying a behavioral response; and ρ_i is a subject-specific frailty term corresponding to unobserved heterogeneity. All parameters, except for the vector β, are assumed to be strictly positive. The frailty term is assumed to arise from a distribution with support on $\mathcal{R}^+$. This distribution shall not be left unspecified in general [see, e.g., Link (2003)], even if it is possible for certain models in discrete time [Farcomeni and Tardella (2010, 2012)]. A common choice [e.g., Xi, Yip and Watson (2007)] is a $\text{Ga}(\alpha, \alpha)$ distribution, that is,

$$f(\rho_i) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \rho_i^{\alpha-1} e^{-\rho_i \alpha}. \tag{2.2}$$

This choice, that is, assuming that both parameters of the Gamma distribution are identical, implies that a priori $E(\rho_i) = 1$, as with most frailty models, and $\text{Var}(\rho_i) = 1/\alpha$. As usual, a larger variance corresponds to a more important role of unobserved heterogeneity, that is, a *smaller* α can be interpreted as a larger unobserved heterogeneity. The model, as stated, involves the following restrictions: first, we make the assumption of proportionality of hazards, that is, that the effects of covariates included in the model are time-constant and log-linear. The formulation is furthermore restricted to time-constant covariates as well. Second, we assume a multiplicative effect of the frailty ρ_i, which is also time constant; and we formulate a parametric assumption on its distribution. In (2.1) we also make the usual long term memory assumption for the behavioral effect. We will generalize (2.1) to more complex behavioral effects in Section 3 below. It is worth noticing here that straightforward assumptions can be used to obtain simpler classes of models, some of which are commonly used in the capture–recapture literature. Assuming, for instance, that $\omega(t) = \omega$ is constant over time leads to time-homogeneity. Fixing $\beta = 0$ leads to assume that there is no observed heterogeneity, while assuming $\rho_i = 1$ corresponds to no unobserved heterogeneity. Finally,

TABLE 1
Models nested in M_{hotb} and the corresponding restrictions needed to obtain them

Model	**Equation**	**Restrictions on M_{hotb}**
M_{hotb}	$\rho_i e^{\beta' Z_i} \phi^{I(N_i(t)\geq 1)} \omega(t)$	
M_{hob}	$\rho_i e^{\beta' Z_i} \phi^{I(N_i(t)\geq 1)} \omega$	$\omega(t) = \omega$
M_{htb}	$\rho_i \phi^{I(N_i(t)\geq 1)} \omega(t)$	$\beta = 0$
M_{hot}	$\rho_i e^{\beta' Z_i} \omega(t)$	$\phi = 1$
M_{otb}	$e^{\beta' Z_i} \phi^{I(N_i(t)\geq 1)} \omega(t)$	$\rho_i = 1$
M_{ho}	$\rho_i e^{\beta' Z_i} \omega$	$\omega(t) = \omega, \phi = 1$
M_{ht}	$\rho_i \omega(t)$	$\beta = 0, \phi = 1$
M_{hb}	$\rho_i \phi^{I(N_i(t)\geq 1)} \omega$	$\beta = 0, \omega(t) = \omega$
M_{ot}	$e^{\beta' Z_i} \omega(t)$	$\phi = 1, \rho_i = 1$
M_{tb}	$\phi^{I(N_i(t)\geq 1)} \omega(t)$	$\beta = 0, \rho_i = 1$
M_{ob}	$e^{\beta' Z_i} \phi^{I(N_i(t)\geq 1)} \omega$	$\omega(t) = \omega, \rho_i = 1$
M_h	$\rho_i \omega$	$\omega(t) = \omega, \beta = 0, \phi = 1$
M_o	$e^{\beta' Z_i} \omega$	$\omega(t) = \omega, \rho_i = 1, \phi = 1$
M_t	$\omega(t)$	$\beta = 0, \rho_i = 1, \phi = 1$
M_b	$\omega \phi^{I(N_i(t)\geq 1)}$	$\omega(t) = \omega, \rho_i = 1, \beta = 0$
M_0	ω	$\omega(t) = \omega, \rho_i = 1, \beta = 0, \phi = 1$

$\phi = 1$ corresponds to no behavioral effects. An account of all possible models and corresponding assumptions is given in Table 1.

3. Flexible behavioral effects. Model (2.1) includes classical behavioral effects: the risk of capture changes, and then remains constant after the first capture. The assumption that risk changes only at first capture may be restrictive in some cases; see, for instance, Farcomeni (2011), Ramsey and Severns (2010), Yang and Chao (2005) and the references therein. In continuous time capture–recapture, we outline two generalizations of the usual M_b, which can be summarized in our modification of (2.1) as follows:

$$\lambda_i(t) = \rho_i e^{\beta' Z_i} \phi^{I(c_1 \leq N_i(t) \leq c_2 \text{ and } t \leq t_{ic_1} + \Delta_b)} \omega(t), \tag{3.1}$$

where $c_1 < c_2$ are fixed integers smaller than the maximum number of recaptures $\max_i N_i(\tau)$, and $\Delta_b > 0$. The model is identifiable as soon as there is at least one subject with at least $c_1 + 1$ captures, with the $c_1 + 1$th in the interval $(t_{ic_1}, t_{ic_1} + \Delta_b)$. Note that assuming $c_1 = 1$ and $c_2 = \Delta_b = \infty$ gives back (2.1). In (3.1) we allow for a delayed onset of the behavioral response. For instance, $c_1 = 3$ means that a behavioral effect is not expected after the first or second, but only after the third capture. Further, we allow for a finite time memory: after the minimum between $c_2 - c_1$ captures and a time frame of Δ_b, the risk of capture returns to

the pre-behavioral-response state. In our application, for instance, it is reasonable to expect that $c_1 = 2$, given that more serious legal consequences are experienced after the second capture, and that $c_2 = \Delta_b = \infty$ (purely delayed onset model). There would be a finite time behavioral response if, for instance, a subject who is not captured for a Δ_b time frame is forgiven the previous offenses. A finite c_2 can be expected in applications of capture–recapture to animal populations, in which the animals may get used to traps after being trapped a few times.

4. Inference. Given that we cannot measure covariates for subjects never captured, the only practical possibility for inference is via maximization of the conditional likelihood, that is, the likelihood obtained conditioning on the event that $N_i(\tau) > 0$. The estimated parameters are then used to build an Horvitz–Thompson (HT) type estimator for N [Horvitz and Thompson (1952), Sanathanan (1972)]. For the sake of conciseness and simplicity of notation we here outline inference for model M_{hotb} in (2.1). The strategy for submodels can be found along similar lines.

In what follows, let $\Omega(t) = \int_0^t \omega(s)\,ds$ and $\gamma_i = e^{\beta' Z_i}$. Denote the probability of having at least one capture for individual i, conditional on ρ_i, as

$$P_i = 1 - e^{-\rho_i \gamma_i \Omega(\tau)}. \tag{4.1}$$

With an argument similar to the reasoning in Crowder et al. (1991) and Hwang and Chao (2002), we can thus obtain a likelihood contribution for the ith subject, conditional on ρ_i, as follows:

$$L_i \propto \phi^{N_i(\tau)-1} \gamma_i^{N_i(\tau)} \left[\prod_{j=1}^{N_i(\tau)} \omega(t_{ij}) \right] \rho_i^{N_i(\tau)} e^{-\rho_i \gamma_i \Omega_i^*(\tau)} / P_i,$$

where

$$\Omega_i^*(\tau) = \phi\Omega(\tau) + (1-\phi)\Omega(t_{i1}). \tag{4.2}$$

When ρ_i is a random effect arising from a parametric distribution $F(\rho_i)$, we shall integrate the above expressions with respect to this distribution. We consequently have a likelihood that is conditional on having at least one capture, and marginal with respect to random effects. The expression for this likelihood is

$$L_c = \prod_{i=1}^{n} \phi^{N_i(\tau)-1} \gamma_i^{N_i(\tau)} \left[\prod_{j=1}^{N_i(\tau)} \omega(t_{ij}) \right] \int_0^\infty \frac{x^{N_i(\tau)} e^{-x\gamma_i \Omega_i^*(\tau)}}{1 - e^{-x\gamma_i \Omega(\tau)}}\, dF(x). \tag{4.3}$$

The above expression is a function of the covariate parameters β, the behavioral parameter ϕ, the baseline function $\omega(t)$ and any parameter involved in $F(\rho_i)$. Estimates of the above parameters can be obtained by maximizing L_c. Note that this

maximum likelihood estimator will be consistent and asymptotically normal given that it satisfies properties in Nielsen et al. (1992) and Sanathanan (1972). See also Gill (1992).

When $F(\rho_i)$ is a $\mathrm{Ga}(\alpha, \alpha)$, we can solve the integral and obtain a closed form expression for L_c. Note, in fact, that under the assumption of Gamma distributed random effects

$$L_c = \prod_{i=1}^{n} \phi^{N_i(\tau)-1} \gamma_i^{N_i(\tau)} \left[\prod_{j=1}^{N_i(\tau)} \omega(t_{ij}) \right] \frac{\alpha^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{x^{N_i(\tau)+\alpha-1} e^{-x(\gamma_i \Omega_i^*(\tau)+\alpha)}}{1 - e^{-x\gamma_i \Omega(\tau)}} \, dx.$$

The integral is proportional to the integral definition of the generalized Zeta function [Gradshteyn and Ryzhik (2007), Section 3.411, formula 7] $\zeta(s, a)$, which is defined as the series $\sum_n 1/(n+a)^s$, and can be evaluated simply. See Magnus, Oberhettinger and Soni (1966) for details.

It can be shown that

$$\int_0^\infty x^a e^{-bx} / (1 - e^{-cx}) \, dx = \frac{\Gamma(a+1)}{c^{a+1}} \zeta(a+1, b/c). \tag{4.4}$$

Hence, after straightforward algebra we obtain that, under assumption of Gamma distributed random effects,

$$L_c = \prod_{i=1}^{n} L_i, \tag{4.5}$$

where

$$\begin{aligned} L_i = {} & \phi^{N_i(\tau)-1} \gamma_i^{N_i(\tau)} \left[\prod_{j=1}^{N_i(\tau)} \omega(t_{ij}) \right] \frac{\alpha^\alpha \Gamma(\alpha + N_i(\tau))}{\Gamma(\alpha)(\gamma_i \Omega(\tau))^{\alpha + N_i(\tau)}} \\ & \times \zeta\left(\alpha + N_i(\tau), \frac{\gamma_i \Omega_i^*(\tau) + \alpha}{\gamma_i \Omega(\tau)} \right). \end{aligned} \tag{4.6}$$

4.1. *Maximization of the conditional likelihood.* Even if, with Gamma distributed random effects, we have a closed-form expression for the conditional likelihood, its direct maximization is cumbersome, as it would require a precise evaluation of the derivatives of the generalized Zeta function, which are not available in closed form. More importantly, our approach would be limited to the assumption of Gamma distributed random effects. We give in this section a maximization strategy based on the EM algorithm, which can be easily adapted to any distributional assumption on ρ_i.

The EM algorithm proceeds by iterating two steps. At the E-step we compute the conditional expected values of the frailty terms and plug them into the *complete*

conditional likelihood, that is, the likelihood we would have if we could observe the frailty terms. This likelihood can be expressed as

$$\tilde{L}_c = \prod_{i=1}^{n} \tilde{L}_i, \tag{4.7}$$

where

$$\tilde{L}_i = \phi^{N_i(\tau)-1} \gamma_i^{N_i(\tau)} \left[\prod_{j=1}^{N_i(\tau)} \omega(t_{ij}) \right] \rho_i^{N_i(\tau)} e^{-\rho_i \gamma_i \Omega_i^*(\tau)} / \left(1 - e^{-\rho_i \gamma_i \Omega(\tau)}\right) f(\rho_i), \tag{4.8}$$

where $f(\cdot)$ denotes the density or pmf of the random effects.

More precisely, at the E-step we substitute ρ_i in (4.7) with

$$\hat{\rho}_i = \int_0^\infty x \, dF\big(x | N_i(t), Z_i, \beta, \omega(t), \phi\big). \tag{4.9}$$

Note that (4.8) is not linear in ρ_i, hence, its conditional expectation does not coincide with its value at (4.9). The E-step we propose is therefore only approximate. The approximation is usually good, as we work with the logarithm of (4.7), which is approximately linear for large intervals, and due to the fact that ρ_i does not change much from one iteration to another. If a decreasing likelihood is observed, the plug-in E-step shall be substituted with a Monte Carlo E-step.

The conditional density for the frailty terms can be obtained through the Bayes theorem as follows:

$$\begin{aligned} &f\big(\rho | N_i(t), Z_i, \beta, \omega(t), \phi\big) \\ &\quad = \frac{f(N_i(t) | \rho, Z_i, \beta, \omega(t)\phi) f(\rho)}{\int_0^{+\infty} f(N_i(t) | \rho, Z_i, \beta, \omega(t)\phi) f(\rho) \, d\rho} \\ &\quad = \frac{\tilde{L}_i}{\int_0^\infty \tilde{L}_i \, d\rho}. \end{aligned} \tag{4.10}$$

When the frailty is assumed to arise from a Gamma distribution, we have a closed-form expression for $\hat{\rho}_i$. The resulting $\hat{\rho}_i$ is a ratio, whose denominator corresponds to L_i as defined in (4.6) and, with similar calculations involving (4.4), whose nominator corresponds to

$$\begin{aligned} &\phi^{N_i(\tau)-1} \gamma_i^{N_i(\tau)} \left[\prod_{j=1}^{N_i(\tau)} \omega(t_{ij}) \right] \frac{\alpha^\alpha \Gamma(\alpha + N_i(\tau) + 1)}{\Gamma(\alpha)(\gamma_i \Omega(\tau))^{\alpha + N_i(\tau) + 1}} \\ &\quad \times \zeta\left(\alpha + N_i(\tau) + 1, \frac{\gamma_i \Omega_i^*(\tau) + \alpha}{\gamma_i \Omega(\tau)}\right). \end{aligned}$$

After simplification of the terms, and noting that $\Gamma(x+1) = x\Gamma(x)$, we get

$$\hat{\rho}_i = \frac{\alpha + N_i(\tau)}{\gamma_i \Omega(\tau)} \frac{\zeta(\alpha + N_i(\tau) + 1, (\gamma_i \Omega_i^*(\tau) + \alpha)/(\gamma_i \Omega(\tau)))}{\zeta(\alpha + N_i(\tau), (\gamma_i \Omega_i^*(\tau) + \alpha)/(\gamma_i \Omega(\tau)))}. \tag{4.11}$$

Other distributional assumptions may lead to the need of numerical integration methods at the E-step or to the use of MCMC in order to sample from the posterior distribution of the random effects [therefore obtaining an MCEM algorithm, see, e.g., Both and Hobert (1999)]. In the latter case, the M-step which we will describe now can be still used, with minor adjustments. After computation of (4.9), we plug it in (4.7) to obtain an Expected Complete Conditional Likelihood (ECCL). The M-step consists in maximizing the ECCL with respect to β, $\omega(t)$, ϕ and any parameter involved in the random effect distribution $F(\rho_i)$. This step is particularly cumbersome due to the fact that there is an extremely large number of parameters involved in the ECCL when the baseline hazard function $\omega(t)$ is not known and not assumed to be constant. To tackle this issue, we note that the nonparametric MLE for $\Omega(t)$ is a step function, with jumps occurring at capture times, that is, the estimated baseline hazard can be expressed as

$$\Omega(t) = \sum_k \theta_k I(t_{(k)} \le t),$$

where $\theta_k = \omega(t_{(k)})$ measures the size of the jump at the kth capture time. Maximizing the conditional likelihood with respect to θ_k, we obtain a Nelson–Aalen type estimator [Aalen (1978), Nelson (1972)]. See also Hwang and Chao (2002) for a similar reasoning in the continuous capture–recapture context. A closed-form expression for $\hat{\theta}_k$ can be obtained by computing the first derivative of the log-ECCL with respect to θ_k and equating to zero. Straightforward algebra gives

$$0 = \frac{dN(t_{(k)})}{\theta_k} - \sum_{i=1}^{n} \hat{\rho}_i \gamma_i \left[\phi + (1-\phi) I(t_{(k)} < t_{i1}) + \frac{e^{-\hat{\rho}_i \gamma_i \Omega(\tau)}}{1 - e^{-\hat{\rho}_i \gamma_i \Omega(\tau)}} \right],$$

where $dN(t_{(k)})$ gives the number of captures occurring exactly at the kth capture time. Consequently,

$$\hat{\theta}_k = \frac{dN(t_{(k)})}{\sum_{i=1}^{n} \hat{\rho}_i \gamma_i [\phi^{I(t_{i1} < t_{(k)})} + e^{-\hat{\rho}_i \gamma_i \Omega(\tau)} / (1 - e^{-\hat{\rho}_i \gamma_i \Omega(\tau)})]}. \tag{4.12}$$

The resulting baseline hazard estimator can be seen as a Nelson–Aalen type estimator with covariates, with an exponential term correction at the denominator obtained as a consequence of conditioning to subjects with at least one event.

The expression for (4.12) shall be now plugged in the ECCL, thereby drastically reducing the number of parameters to $\Omega(\tau)$, ϕ, β and any parameter involved in the random effect distribution $F(\rho_i)$. These are estimated by solving a system of equations which are obtained by equating to zero the first derivatives of the ECCL after plug-in of $\hat{\theta}_k$, plus an additional equation due to the constraint

$$\sum_k \hat{\theta}_k = \hat{\Omega}(\tau). \tag{4.13}$$

Let now, for ease of notation,

$$A(\gamma_h, \hat{\rho}_h, \Omega(\tau)) = \frac{e^{-\hat{\rho}_h \gamma_h \Omega(\tau)}}{1 - e^{-\hat{\rho}_h \gamma_h \Omega(\tau)}} \tag{4.14}$$

and

$$B(\phi, \gamma, \hat{\rho}, \Omega(\tau), t_k) = \sum_{h=1}^{n} \hat{\rho}_h \gamma_h \left[\phi^{I(t_k > t_{h1})} + \frac{e^{-\hat{\rho}_h \gamma_h \Omega(\tau)}}{1 - e^{-\hat{\rho}_h \gamma_h \Omega(\tau)}} \right]. \tag{4.15}$$

For what concerns the derivative with respect to ϕ, we obtain

$$\begin{aligned} \sum_{i=1}^{n} \Bigg\{ & \frac{(N_i(\tau) - 1)}{\phi} \\ & - \sum_{j=1}^{N_i(\tau)} \frac{\sum_{p=1}^{n} \hat{\rho}_p \gamma_p I(t_{ij} > t_{p1})}{B(\phi, \gamma, \hat{\rho}, \Omega(\tau), t_{ij})} \\ & - \hat{\rho}_i \gamma_i \sum_k \frac{dN(t_k)}{B(\phi, \gamma, \hat{\rho}, \Omega(\tau), t_k)} \\ & \qquad \times \left[I(t_k > t_{i1}) - \phi^{I(t_k > t_{i1})} \frac{\sum_{p=1}^{n} \hat{\rho}_p \gamma_p I(t_k > t_{p1})}{B(\phi, \gamma, \hat{\rho}, \Omega(\tau), t_{ij})} \right] \Bigg\}; \end{aligned} \tag{4.16}$$

while for what concerns $\Omega(\tau)$, we have

$$\begin{aligned} \sum_{i=1}^{n} \Bigg\{ & \sum_{j=1}^{N_i(\tau)} \left[\frac{\sum_{k=1}^{n} e^{\hat{\rho}_k \gamma_k \Omega(\tau)} (\hat{\rho}_k \gamma_k A(\gamma_k, \hat{\rho}_k, \Omega(\tau)))^2}{B(\phi, \gamma, \hat{\rho}, \Omega(\tau), t_{ij})} \right] \\ & - \hat{\rho}_i \gamma_i \Bigg[\sum_k \phi^{I(t_k > t_{i1})} \frac{dN(t_k)}{B(\phi, \gamma, \hat{\rho}, \Omega(\tau), t_k)^2} \\ & \qquad \times \sum_{p=1}^{n} e^{\hat{\rho}_p \gamma_p \Omega(\tau)} \big(\hat{\rho}_p \gamma_p A(\gamma_p, \hat{\rho}_p, \Omega(\tau))\big)^2 \Bigg] \\ & \qquad + \hat{\rho}_i \gamma_i A(\gamma_i, \hat{\rho}_i, \Omega(\tau)) \Bigg\}. \end{aligned} \tag{4.17}$$

Finally, taking the first derivative of the log-ECCL with respect to β_h, $h = 1, \ldots, p$, we obtain the p equations

$$\begin{aligned} \sum_{p=1}^{n} \Bigg\{ & \frac{N_p(\tau)}{\gamma_p} - \sum_{i=1}^{n} \sum_{j=1}^{N_i(\tau)} \big([\hat{\rho}_p A(\gamma_p, \hat{\rho}_p, \Omega(\tau)) \\ & \qquad \times (1 - \hat{\rho}_p \gamma_p \Omega(\tau) e^{\hat{\rho}_p \gamma_p \Omega(\tau)} A(\gamma_p, \hat{\rho}_p, \Omega(\tau))) \\ & \qquad\qquad + \hat{\rho}_p \phi^{I(t_{ij} > t_{p1})}] \\ & \qquad / B(\phi, \gamma, \hat{\rho}, \Omega(\tau), t_{ij}) \big) \end{aligned} \tag{4.18}$$

$$
\begin{aligned}
&- \hat{\rho}_p \left[\sum_k \phi^{I(t_k > t_{p1})} \frac{dN(t_k)}{B(\gamma, \hat{\rho}, \Omega(\tau), t_k)} \right] \\
&+ \sum_{i=1}^{n} \hat{\rho}_i \gamma_i \left\{ \sum_k \phi^{I(t_k > t_{i1})} \frac{dN(t_k) \hat{\rho}_p}{(B(\phi, \gamma, \hat{\rho}, \Omega(\tau), t_k))^2} \right. \\
&\qquad \times \left[\phi^{I(t_k > t_{p1})} + A(\gamma_p, \hat{\rho}_p, \Omega(\tau)) \right. \\
&\qquad\qquad \left. \left. - \hat{\rho}_p \gamma_p \Omega(\tau) e^{\hat{\rho}_p \gamma_p \Omega(\tau)} \big(A(\gamma_p, \hat{\rho}_p, \Omega(\tau))\big)^2 \right] \right\} \\
&\qquad\qquad\qquad \left. + \hat{\rho}_p \Omega(\tau) A(\gamma_p, \hat{\rho}_p, \Omega(\tau)) \right\} \gamma_p Z_{ph},
\end{aligned}
$$

where Z_{ph} denotes the phth entry of the covariate matrix Z. The algebra involved in obtaining (4.16), (4.17) and (4.18) is given in the supplement [Farcomeni and Scacciatelli (2013)]. Note that combining (4.16) with (4.17) and (4.18) we obtain the score vector related to ϕ, $\Omega(\tau)$ and β in closed form. In order to conveniently proceed with the M-step, we exploit the Newton–Raphson (NR) algorithm. The NR algorithm only involves numerically computing the first derivative of the score vector above, augmented with (4.13). Parameters involved in the random effects distribution can be tackled separately. We can summarize our algorithm at the tth iteration with the following pseudo-code, which is iterated until convergence in the likelihood.

Update $\hat{\rho}_i^{(t)}$ as in (4.11).
Solve the system of equations given by (4.13), (4.16), (4.17) and (4.18) to update $\hat{\Omega}(\tau)$, $\hat{\beta}$ and $\hat{\phi}$.
Update $\hat{\alpha}$ as the maximizer of $\prod_i f(\hat{\rho}_i)$, where $f(\rho_i)$ is defined in (2.2).
Compute $\hat{\theta}^k$ as in (4.12).
Compute the likelihood as in (4.5).

Final comments concern computation of the information matrix, model choice and hypothesis testing. Regarding the first issue, we note that minus the derivative of the score vector corresponds to the observed information matrix. Since the score vector is available in closed form, we can obtain the corresponding information matrix as a natural by-product of our maximization strategy. The observed information matrix at the maximum likelihood estimate can be used to compute the standard errors in the usual way. A similar approach for estimation of the observed information matrix after use of the EM algorithm is proposed in Bartolucci and Farcomeni (2009), in a completely different context. A simulation study in Bartolucci and Farcomeni (2009) gives evidence of the validity of this procedure. The standard errors can then be used to build Wald statistics for testing.

4.2. *Other assumptions on the behavioral effect.* We briefly sketch in this paragraph how to modify the EM algorithm proposed for model M_{hotb} when more flexible assumptions are used for the behavioral effects, as summarized in equation (3.1). In order to estimate such a model, it suffices to substitute the indicator function arising from (2.1) with the more complex indicator function involved in (3.1). These substitutions occur directly in (4.12), (4.15), (4.16), (4.17) and (4.18). Further, straightforward algebra leads to redefine $\Omega_i^*(\tau)$ as

$$\Omega_i^*(\tau) = \Omega(\tau) + (1-\phi)\big(\Omega(t_{ic_1}) - \Omega\big(\min(t_{ic_2}, t_{ic_1} + \Delta_b)\big)\big). \tag{4.19}$$

Also, the likelihoods are modified: in (4.3), (4.6) and (4.8), $\phi^{N_i(\tau)-1}$ shall be substituted with

$$\phi^{\sum_{j=c_1}^{c_2} I(t_{ij} <= t_{ic_1} + \Delta_b)},$$

where it is understood that $t_{ij} = \infty$ if $j > N_i(\tau)$.

Recall that c_1, c_2 and Δ_b must be fixed a priori. In order to carefully choose these parameters, we can fit models for a range of values of c_1, c_2 and Δ_b and use the one corresponding to the largest log-likelihood. The selected combination (c_1, c_2, Δ_b) is important information, telling us, for instance, that a behavioral response is not expected before the c_1th capture.

4.3. *Estimation of population size.* The maximum likelihood estimates can be used in a final E-step, thus obtaining the corresponding $\hat{\rho}_i$. The latter can be used to maximize the residual likelihood [see Farcomeni and Tardella (2012), Sanathanan (1972)] through a Hortvitz–Thompson estimator of the kind

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{1 - e^{-\hat{\rho}_i \hat{\gamma}_i \hat{\Omega}(\tau)}}. \tag{4.20}$$

When N is large, as in our application, parameters involved in (4.20) are approximately multivariate normal, and their covariance matrix can be estimated as described above.

In order to obtain the standard error of $\hat{N}$, we note that n is random as well and that (4.20) should be actually expressed as

$$\hat{N} = \sum_{i=1}^{N} \frac{\delta_i}{1 - e^{-\hat{\rho}_i \hat{\gamma}_i \hat{\Omega}(\tau)}},$$

where the random variable δ_i is defined as $\delta_i = I(N_i(\tau) \geq 1)$, $i = 1, \dots, N$. Therefore [Böhning (2008), van der Heijden et al. (2003)], by conditioning,

$$\operatorname{Var}(\hat{N}) = \operatorname{Var}_n\big(E(\hat{N}|n)\big) + E_n\big[\operatorname{Var}(\hat{N}|n)\big]. \tag{4.21}$$

The first term on the right-hand side can be estimated by $\sum_{i=1}^{n}(1-w_i)/w_i^2$, where $w_i = 1 - e^{-\hat{\rho}_i \hat{\gamma}_i \hat{\Omega}(\tau)}$. The second term can be computed approximately by means of the delta method, where we have

$$E_n\big[\widehat{\operatorname{Var}(\hat{N}|n)}\big] \to \nabla g(\hat{\theta})' J(\hat{\theta})^{-1} \nabla g(\hat{\theta}) \tag{4.22}$$

with $\hat{N} = g(\hat{\theta})$, $\hat{\theta}$ being the maximum likelihood estimate of the parameter vector and $J(\hat{\theta})$ its information matrix. An expression for $\nabla g(\hat{\theta})$ is cumbersome, but the latter can be easily derived using numerical differentiation methods. For a similar strategy refer also to Böhning and van der Heijden (2009).

5. Data description and analysis. In Italy use or possession of a small amount of drugs may be punished by administrative sanctions. If it is the first or second offense, the Prefect may only issue a warning (Art. 75 modified by law 49/2006). According to the Central Statistics Office of the Interior Ministry (DCDS), about 30,000 interviews are conducted in the presence of Italian Prefects each year [European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (2009)]. Following these interviews, around 25,000 individuals (aged from 10 to 64 years) are formally warned to refrain from further use of narcotic substances. During our observation period, starting in 2006 immediately after enforcement of the new law, the DCDS collected a database with information on subjects reported for breaking the drug law (Art. 75 on personal use). Individuals are identified with their fiscal code (FC), which is the Italian equivalent of a social security number in the USA. The database includes a record for each capture, with date and time of detected abuse, and some information about the subject including the FC for identification. We restrict to the adult population, that is, subjects aged 18 or more. In this work we use information related to the gender, district of residence (South, Central, North–West or North–East Italy, where subjects captured on the islands separated from the Italian mainland are included as customary in the Southern district), age at the start of the observation period and its square.

A preliminary issue regards closure of the population of interest, which is achieved here by restricting the observation period to 2 full years. Given that the duration of cannabis use is often several years [Antidrug Policies Department (2009)], and that frequent users who are most at risk of being captured are also less likely to cessate [Chen and Kandel (1998)], our study period may be a reasonable choice. We have also performed a simple sensitivity analysis, by repeatedly estimating the population size after truncation of the observation period at shorter time horizons (6, 12 and 18 months). All resulting estimates are rather close to those reported below.

Table 2 shows the counting distributions over the categorical covariates. The mean age is 25.01, with a standard deviation of 6.91. In Table 3 we report the log-likelihood and corresponding population size estimate for general M_{hotb} models with all predictors and different choices of c_1 and c_2. On the basis of results in Table 3, we end up choosing $c_1 = 2$ and $c_2 = \infty$.

In order to confirm the presence of all four sources of heterogeneity with these data, we compare in Table 4 the chosen M_{hotb} model to submodels. Each likelihood ratio test of the full versus nested models has $p < 0.0001$. There are a few features that we should notice: first of all, the likelihood of model M_{hot} is very low when compared to models including a behavioral effect. This allows us to conclude that

TABLE 2
Marginal and conditional counting distributions for cannabis users

		Count			
Covariate	Description	1	2	3	4
Substance	Cannabis	50,785	1124	60	4
Gender	Male	47,181	1035	55	4
	Female	3604	89	5	0
District	South	12,684	239	17	1
	Center	18,427	411	22	0
	North–East	7850	193	6	0
	North–West	11,824	281	15	3

TABLE 3
Complete M_{hotb} models fit on the Italian cannabis users data. We report the difference in log-likelihood with respect to the first model, which has log-likelihood −453,196.3

c_1	c_2	log-lik	$\hat{N}/10^6$
1	∞	0	3.199
2	∞	195.2	3.265
3	∞	161.4	2.984
4	∞	148.4	2.903
1	2	33.9	3.079
1	3	20.6	2.911
2	3	171.6	3.150
2	4	192.1	3.151
3	4	187.4	2.934

TABLE 4
Comparison of M_{hotb} model with $c_1 = 2$, $c_2 = \Delta_b = \infty$ with nested models for the Italian cannabis users data. We report the difference in log-likelihood with respect to the full model, which has log-likelihood −453,001.1. *All likelihood ratio tests would lead to* $p < 0.0001$

Model	log-lik	$\hat{N}/10^6$
M_{hotb}	0	3.265
M_{otb}	−13.8	2.417
M_{htb}	−28.4	3.254
M_{hob}	−12.0	1.558
M_{hot}	−28.1	2.368

there is strong evidence in favor of the behavioral effect. The peculiar behavioral effect detected can be explained as a purely random variation, that is, a feature of the counting distribution, or by considering that more serious legal consequences are expected after the second time a subject is reported. In the first case, we note from Table 2 that there is some sort of *one-inflation* in the counting distribution, that is, subjects are mostly captured only once. This may be a consequence of the fact that the probability of repeatedly meeting the same person by chance alone is low, at least in large cities and in general within our short observation period. In the second case, it can be noted that after the second capture under the new Italian law, subjects may be submitted to mandatory psychotherapy, they may be revoked their driver's licence and entry visa when foreigners, and they may have to pay a fine and/or participate in treatment programs. On the other hand, the first and second time a subject is identified as a user have most likely the same consequence of a warning by the judge. Identifications for breaking other laws or other articles of the same law do not count toward legal consequences arising at the third capture. It is not surprising then that subjects may be "trap-shy" after the second capture (compare with estimate $\hat{\phi}$ in Table 5 below). What we cannot say with available information is if this is a feature of the counting distribution which is due only to chance or if it is a feature of the subjects being captured. In the second case, we also cannot say if repression works, that is, subjects actually quit using cannabis after the second time they are reported or if they only start using it at home and have friends buy and carry it for them.

We also have evidence in favor of both observed and unobserved heterogeneity, and also the assumption of time-homogeneity shall be rejected. Even if we obtain maximum likelihood estimates differently, most of the models in Table 4 correspond to a generalization of models in Hwang and Chao (2002) (e.g., M_{otb} and submodels) and Xi, Yip and Watson (2007) (e.g., M_{htb} and submodels), allowing

TABLE 5
Results of M_{hotb} model with $c_1 = 2$, $c_2 = \Delta_b = \infty$

Parameter	**Estimate**	**Std. Err.**
N	3,265,071	91,247.39
α	129.69	0.56
ϕ	0.20	0.01
$\Omega(\tau)$	0.007	0.001
β(Age)	−0.04	0.04
β(Age^2)	−0.03	0.05
β(Female)	−0.65	0.11
β(North–East)	−2.13	0.08
β(North–West)	−2.20	0.07
β(South)	−2.03	0.07

for a delayed onset of the behavioral response. For our selected model M_{hotb} with $c_1 = 2$ we report in Table 5 the parameter estimates and standard errors.

Our final estimate for the population size of our catchable population is slightly less than 3.3 million, which accounts for approximately 8.9% of Italians aged 18 to 64, and 7% of the entire adult population. In order to obtain an indirect estimate for all cannabis users, we can use results in Rey, Rossi and Zuliani (2011), who claim that around two-thirds of the population of cannabis users is catchable under Art. 75. Consequently, an indirect estimate for the population of adults who have used cannabis at least once in Italy is given by 5 million, which accounts for 13.5% of Italians aged 18 to 64, or 10.6% of the entire adult population.

Unobserved heterogeneity is relatively weak, even if present. For what concerns observed heterogeneity, we find that the risk of capture decreases with age and its square. Females are at a lower risk of capture than men, and the subjects in the central district of Italy are at the highest risk of being reported. Note that we can interpret these findings as a differential in prevalence only under the assumption that time-heterogeneity is independent of predictors; otherwise the difference in prevalence could be explained by different ability of the officers in different Italian regions. Under this assumption, as age grows use of cannabis in the street is likely more limited, females tend to use cannabis less than men, and the largest number of users can be found in the central region.

These findings are reasonable, as in the southern regions of Italy the use of drugs is more limited due to sociocultural differences as compared to the central and northern parts. In the North–East and North–West we have both a slightly smaller population at risk and possibly a larger prevalence of cocaine rather than cannabis use. This assumption could be easily verified/relaxed by stratifying our Cox-type model to have a different baseline for each category of an observed covariate combination.

We conclude with a comparison with other estimators of the population size, which can be found in Table 6. We use the independence model M_0, the Chao

TABLE 6
Estimates of population size for cannabis data for some time-homogeneous population size estimators

Estimator	$\hat{N}/10^6$	**Std. Err.**$/10^5$	**log-lik**
M_0	1.11	0.31	−53.66
Chao [Chao (1987)]	1.20	0.36	−15.37
M_h Poisson2 [Rivest and Baillargeon (2007)]	1.85	1.06	−19.54
M_h Darroch [Darroch et al. (1993)]	3.70	4.43	−16.16
M_h Gamma3.5 [Rivest and Baillargeon (2007)]	7.32	13.54	−15.60
M_{ho} [Böhning and van der Heijden (2009)]	1.14	0.05	−5660

[Chao (1987)] lower bound estimator and other M_h models, with the assumptions for the mixing distribution available in the function `closedp.0` in the `R` package `Rcapture`. In Table 6 we obtain rather different estimates of the population size, particularly relative to the estimated standard errors. We believe this happens when there are biases in population size and standard error estimates. In this case, any model not including the four sources of heterogeneity may be misspecified for the data at hand. Among the M_h models, the largest likelihood arises with the mixing distribution Gamma3.5. The latter seems to grossly overestimate the population size. It is not surprising that Gamma3.5 leads to overestimating the population size, as with this mixing distribution the capture probabilities are not bounded from below. See, for instance, Baillargeon and Rivest (2007) on this point and also for a detailed description of the three mixing distributions used in Table 6. The M_h Darroch model gives the closest estimate to our final estimate of 3.3 million, but with a standard error that is almost five times larger than ours. Further, the latter would not be used in practice, as M_h Gamma3.5 yields a larger likelihood with the same number of parameters. We also include the Böhning and van der Heijden (2009) model, an M_{ho} model generalizing that of Zelterman (1988) and therefore providing an efficient lower bound for the population size. The resulting estimate is comparable to that obtained with the Chao lower bound, with a smaller standard error due to the inclusion of covariates.

Finally, we compare with two indirect estimates. The first is given in Santoro, Triolo and Rossi (2013) using a dealer/consumer ratio as proposed in Bouchard and Tremblay (2005), and estimates as 3.5 million the entire population of cannabis users. The second is the official estimate of 5.5 million for those who have used cannabis at least once in Italy, given by the Antidrug Policies Department (2009) and European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (2010). As it is estimated that about 85% of all marijuana users are aged 18 or more [European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (2010)], it can be said that from Santoro, Triolo and Rossi (2013) we have an estimate of about 3 million for the adult population and that the official estimate is slightly under 4.7 million. We believe these numbers to be comparable with our indirect estimate of 5 million, and our results confirm the general idea that official estimates may be slightly underestimating the population size [e.g., Santoro, Triolo and Rossi (2013)].

6. Discussion. We have proposed a general framework for continuous time capture–recapture. Our model includes time-heterogeneity, unobserved subject-specific heterogeneity, observed subject-specific heterogeneity and behavioral response to capture. Classical behavioral response has been generalized, allowing the user to specify a delayed onset and a finite time memory.

In our application we have found predictors that are able to explain some heterogeneity, but we also have unobserved heterogeneity. A model including both

observed and unobserved heterogeneity is therefore needed for these data. Unfortunately, inferential approaches already available cannot be directly extended to this case. Our EM-type estimation approach can be easily adapted to any of the submodels of M_{hotb}. Under the assumption of Gamma distributed random effects we derived closed-form expressions for some of the quantities involved, which greatly speeds up our algorithm. We also found the score in closed form and have made use of numerical differentiation only to compute its first derivative. The accuracy of numerical first derivatives is much better than the accuracy of numerical second derivatives, which we do not need in this paper.

We have developed a challenging application, with a small sampling fraction and many sources of heterogeneity. Available estimates of the number of cannabis users in Italy are mostly based on chemical analyses of waste waters or on consumer/dealers ratios. In our application we developed a direct estimate of the population size of adults (≥ 18) who buy, carry and/or use cannabis in the streets in Italy. It shall be noted that our population includes only subjects found in possession of less than 500 mg of cannabis. Possession of more than 500 mg or for the purpose of trafficking, selling, trafficking and cultivation are offences under different articles of the Italian law, and offenders are not included in our sample. Our final estimate for the population size is about 3.3 million, with a standard error of about 91 thousands. Our final indirect estimate for the entire population of adult cannabis users is 5 million. It shall be furthermore noted that subjects in possession of more than 500 mg, trafficking, selling and/or cultivating are commonly estimated to be around 200 thousand [Santoro, Triolo and Rossi (2013)]. This number is smaller than the width of our estimated confidence interval, hence, even including these subjects, our final estimates can be thought to be approximately the same.

With our age restriction we exclude an important subpopulation, given that younger people are often the target of prevention policies and delays in age of use may be associated with lower risks [e.g., Fergusson and Horwood (2006)]. Nevertheless, patterns of abuse, officer policies and prefect behavior are different for underage users, hence, it is important that these two populations are separately investigated. See also Kandel and Logan (1984) and Ellickson, Martino and Collins (2004) for a discussion on patterns of drug use. Investigation of the under-18 population should in our opinion be performed with open population models, which are beyond the scope of this paper. Open population models, furthermore, need a precise assessment of date of first and very last use at least for some subjects, which may difficult to measure without bias.

There are further limitations in our data set. Given our catchable population is made only of subjects who use, buy or carry cannabis outside their apartment, there may be concerns about the interpretation of the estimated behavioral effect. We have found that the risk of being captured decreases abruptly at the second capture time. Even if this feature is not purely random, it may be that after the second capture most users do not actually quit, but start using drugs at home instead of outside and have someone else buy for them. We, furthermore, as often

happens with capture–recapture experiments, cannot guarantee that the population is closed. A final limitation regards the limited information we have: we could not take into account covariates related to the officers, and we do not have information about time of day of each event. Therefore, we cannot distinguish between different habits (e.g., day and night users). We also do not have information regarding intensity of use and arrests for other crimes, which directly affect the risk of capture.

We conclude giving a brief account of possibilities for further work. Our strategy for computation of standard errors may yield estimates that are somewhat biased downward, given that we are not taking into account uncertainty brought about by model search. This is a primary issue for further work. Our model could also be extended to include nonparametric baseline functions specific to known or even unknown blocks of subjects. Finally, it could also be extended to take into account spatial dependence.

Acknowledgments. The authors are grateful to five referees and the Editor (Professor T. Gneiting) for very kind and detailed comments which helped improve the presentation; and to Professor Carla Rossi for suggestions on an earlier draft. This work has been developed within the framework of the EU project JUST/2010/DPIP/AG/1410: "New methodological tools for policy and programme evaluation," with the financial support of the Prevention and Information Programme of the European Commission. The contents of this publication are the sole responsibility of the authors and can in no way be taken to reflect the views of the European Commission.

SUPPLEMENTARY MATERIAL

Derivation of (4.16), (4.17) and (4.18). (DOI: 10.1214/13-AOAS672SUPP; .pdf). Derivatives of the log ECCL.

REFERENCES

AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726. MR0491547

AMSTRUP, S. C., MCDONALD, T. L. and MANLY, B. F. J., eds. (2003). *Handbook of Capture–Recapture Analysis*. Wiley, London.

ANTIDRUG POLICIES DEPARTMENT (2009). National report to the EMCDDA by the reitox Italian focal point. New development, trends and in-depth information on selected issues. Presidency of the Italian Council of Ministers, Rome.

BAILLARGEON, S. and RIVEST, L. P. (2007). `Rcapture`: Loglinear models for capture–recapture in R. *Journal of Statistical Software* **19**.

BARTOLUCCI, F. and FARCOMENI, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *J. Amer. Statist. Assoc.* **104** 816–831. MR2751454

BÖHNING, D. (2008). A simple variance formula for population size estimators by conditioning. *Stat. Methodol.* **5** 410–423. MR2528565

BÖHNING, D. and VAN DER HEIJDEN, P. G. M. (2009). A covariate adjustment for zero-truncated approaches to estimating the size of hidden and elusive populations. *Ann. Appl. Stat.* **3** 595–610. MR2750674

BÖHNING, D., SUPPAWATTANABODEE, B., KUSOLVISITKUL, W. and VIWATWONGKASEM, C. (2004). Estimating the number of drug users in Bangkok 2001: A capture–recapture approach using repeated entries in one list. *European Journal of Epidemiology* **19** 1075–1083.

BOTH, J. G. and HOBERT, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 265–285.

BOUCHARD, M. and TREMBLAY, P. (2005). Risks of arrests across drug markets: A capture–recapture analysis of hidden dealer and user populations. *Journal of Drug Issues* **35** 733–754.

CHAO, A. (1987). Estimating the population size for capture–recapture data with unequal catchability. *Biometrics* **43** 783–791. MR0920467

CHEN, K. and KANDEL, D. B. (1998). Predictors of cessation of Marijuana use: An event history analysis. *Drug and Alcohol Dependence* **50** 109–121.

CHIANG, S.-C., CHEN, C.-Y., CHANG, Y.-Y., SUN, H.-J. and CHEN, W. (2007). Prevalence of heroin and methamphetamine male users in the northern Taiwan, 1999–2002: Capture–recapture estimates. *BMC Public Health* **7** 292.

COX, D. R. (1972). Regression models and life-tables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **34** 187–220. MR0341758

CROWDER, M. J., KIMBER, A. C., SMITH, R. L. and SWEETING, T. J. (1991). *Statistical Analysis of Reliability Data*. Chapman & Hall, London. MR1122148

DARROCH, J. N., FIENBERG, S. E., GLONEK, G. F. V. and JUNKER, B. W. (1993). A three sample multiple capture–recapture approach to census population estimation with heterogeneous catchability. *J. Amer. Statist. Assoc.* **88** 1137–1148.

DE JONG, W. and WEBER, U. (1999). The professional acceptance of drug use: A closer look at drug consumption rooms in the Netherlands, Germany and Switzerland. *International Journal of Drug Policy* **10** 99–108.

ELLICKSON, P. L., MARTINO, S. C. and COLLINS, R. L. (2004). Marijuana use from adolescence to young adulthood: Multiple developmental trajectories and their associated outcomes. *Health Psychol.* **23** 299–307.

EUROPEAN MONITORING CENTRE FOR DRUGS AND DRUG ADDICTION (EMCDDA) (2009). *Drug Offences*: *Sentencing and Other Outcomes*. EMCDDA, Lisbon.

EUROPEAN MONITORING CENTRE FOR DRUGS AND DRUG ADDICTION (EMCDDA) (2010). *The EMCDDA Annual Report 2010*: *The State of the Drugs Problem in Europe*. Publications Office of the European Union, Luxembourg.

FARCOMENI, A. (2011). Recapture models under equality constraints for the conditional capture probabilities. *Biometrika* **98** 237–242. MR2804224

FARCOMENI, A. and SCACCIATELLI, D. (2013). Supplement to "Heterogeneity and behavioral response in continuous time capture–recapture, with application to street cannabis use in Italy." DOI:10.1214/13-AOAS672SUPP.

FARCOMENI, A. and TARDELLA, L. (2010). Reference Bayesian methods for recapture models with heterogeneity. *TEST* **19** 187–208. MR2610924

FARCOMENI, A. and TARDELLA, L. (2012). Identifiability and inferential issues in capture–recapture experiments with heterogeneous detection probabilities. *Electron. J. Stat.* **6** 2602–2626. MR3020278

FERGUSSON, D. M. and HORWOOD, L. (2006). Early onset cannabis use and psychosocial adjustment in young adults. *Addiction* **92** 279–296.

GILL, R. D. (1992). Marginal partial likelihood. *Scand. J. Stat.* **19** 133–137. MR1173595

GRADSHTEYN, I. S. and RYZHIK, I. M. (2007). *Table of Integrals*, *Series*, *and Products*, 7th ed. Elsevier, Amsterdam. Translated from the Russian, Translation edited and with a preface

by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX). MR2360010

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

HWANG, W.-H. and CHAO, A. (2002). Continuous-time capture–recapture models with covariates. *Statist. Sinica* **12** 1115–1131. MR1947066

INTERNATIONAL WORKING GROUP FOR DISEASE MONITORING AND FORECASTING (1995a). Capture–recapture and multiple record systems estimation I: History and theoretical development. *American Journal of Epidemiology* **142** 1047–1058.

INTERNATIONAL WORKING GROUP FOR DISEASE MONITORING AND FORECASTING (1995b). Capture–recapture and multiple record systems estimation II: Applications in human diseases. *American Journal of Epidemiology* **142** 1059–1068.

KANDEL, D. B. and LOGAN, J. A. (1984). Patterns of drug use from adolescence to young adulthood: I. Periods of risk for initiation, continued use, and discontinuation. *American Journal of Public Health* **74** 660–666.

KHAZAEI, S., POOROLAJAL, J., MAHJUB, H., ESMAILNASAB, N. and MIRZAEI, M. (2012). Estimation of the frequency of intravenous drug users in Hamadan city, Iran, using the capture-recapture method. *Epidemiol. Health* **34** e2012006.

LINK, W. A. (2003). Nonidentifiability of population size from capture–recapture data with heterogeneous detection probabilities. *Biometrics* **59** 1123–1130. MR2019822

MAGNUS, W., OBERHETTINGER, F. and SONI, R. P. (1966). *Formulas and Theorems for the Special Functions of Mathematical Physics*. Springer, New York.

MIROWSKY, J. and ROSS, C. (2003). *Social Causes of Psychological Distress*. de Gruyer, New York.

NELSON, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14** 945–965.

NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. and SØRENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Stat.* **19** 25–43. MR1172965

RAMSEY, F. L. and SEVERNS, P. M. (2010). Persistence models for mark-recapture. *Environ. Ecol. Stat.* **17** 97–109. MR2594937

REUTER, P. and KLEIMAN, M. A. (1986). Risks and prices: An economic analysis of drug enforcement. *Crime and Justice* **7** 289–340.

REY, G. M., ROSSI, C. and ZULIANI, A. (2011). *Il Mercato delle Droghe*: *Dimensione*, *Protagonisti*, *Politiche*. Marsilio, Venezia.

RIVEST, L.-P. and BAILLARGEON, S. (2007). Applications and extensions of Chao's moment estimator for the size of a closed population. *Biometrics* **63** 999–1006, 1309. MR2414575

SANATHANAN, L. (1972). Estimating the size of a multinomial population. *Ann. Math. Statist.* **43** 142–152. MR0298815

SANTORO, M., TRIOLO, L. and ROSSI, C. (2013). Drug user dynamics: A compartmental model of drug users for scenario analyses. *Drugs*: *Education*, *Prevention & Policy* **20** 184–194.

SEMPLE, D. M., MCINTOSH, A. M. and LAWRIE, S. M. (2005). Cannabis as a risk factor for psychosis: Systematic review. *J. Psychopharmacol.* (*Oxford*) **19** 187–194.

VAISSADE, L. and LEGLEYE, S. (2009). Capture–recapture estimates of the local prevalence of problem drug use in six French cities. *Eur. J. Public Health* **19** 32–37.

VAN DER HEIJDEN, P. G. M., BUSTAMI, R., CRUYFF, M. J. L. F., ENGBERSEN, G. and VAN HOUWELINGEN, H. C. (2003). Point and interval estimation of the population size using the truncated Poisson regression model. *Stat. Model.* **3** 305–322. MR2012155

XI, L., YIP, P. S. F. and WATSON, R. (2007). A unified likelihood-based approach for estimating population size in continuous-time capture–recapture experiments with frailty. *Biometrics* **63** 228–236, 315. MR2345593

YANG, H.-C. and CHAO, A. (2005). Modeling animals' behavioral response by Markov chain models for capture–recapture experiments. *Biometrics* **61** 1010–1017. MR2216194

ZELTERMAN, D. (1988). Robust estimation in truncated discrete distributions with application to capture–recapture experiments. *J. Statist. Plann. Inference* **18** 225–237. MR0922210

DEPARTMENT OF PUBLIC HEALTH
AND INFECTIOUS DISEASES
SAPIENZA—UNIVERSITY OF ROME
PIAZZALE ALDO MORO, 5
00185 ROME
ITALY
E-MAIL: alessio.farcomeni@uniroma1.it

CENTRE FOR BIOSTATISTICS AND BIOINFORMATICS
UNIVERSITY OF ROME—TOR VERGATA
VIA ORAZIO RAIMONDO 18
00173 ROME
ITALY
E-MAIL: daria.scacciatelli@gmail.com