Capture-recapture methods to size alcohol related problems in a population

Giovanni Corrao, Vincenzo Bagnardi, Giovanni Vittadini, Sergio Favilli

Abstract

Study objective—To investigate the utility of capture-recapture methods to estimate prevalence of subjects with alcohol related disorders using multiple incomplete lists.

Design—This was a cross sectional study of alcohol related disorders in a large community.

Setting—During 1997 identified cases with known alcohol related disorders were independently flagged by four sources (self help volunteering groups; psychiatric ambulatory; public alcohology service; hospital discharges).

Patients—381 records were flagged, corresponding to 349 individual cases from a target population resident in a northern Italy area.

Main results—The two sample capture-recapture estimates were clearly biased because of dependencies among sources. Estimates based on log-linear models showed prevalent counts ranged from 2297 (95% confidence intervals: 1524, 3794) to 2523 (95% confidence intervals: 1623, 4627) after adjustment for dependence among sources only or also for heterogeneity in catchability among age categories (< 50 and ⩾ 50 years), respectively.

Conclusions—The study suggests that capture-recapture is an appropriate approach for estimating prevalence of subjects with alcohol related problems who seek or need treatment and assistance when different lists of alcoholics can be obtained from different types of agencies involved with problematic use of alcohol. Critical factors are the complexity in case definition and the analysis of heterogeneity among people. Accurate estimates are needed to plan and evaluate public health interventions.

(*J Epidemiol Community Health* 2000;**54**:603–610)

Department of Statistics, Faculty of Statistical Sciences, Chair of Medical Statistics and Epidemiology, University of Milan, Bicocca, Italy
G Corrao
V Bagnardi

Division of Internal Medicine, Unit of Alcohology, Salvatore Maugeri Foundation, Pavia, Italy
G Vittadini

School of Health Statistics, Faculty of Medicine, University "La Sapienza", Rome, Italy
S Favilli

Correspondence to:
Professor Corrao, Dipartimento di Statistica; Università degli Studi di Milano, Bicocca, Via Bicocca degli Arcimboldi (angolo Viale Sarca, 202), Edificio U7, 20126 Milan, Italy (giovanni.corrao@.unimib.it)

Accepted for publication 15 February 2000

Ascertaining the size of alcohol related problems (ARP) in a community has strong implications in medical and public health issues.[1,2] Prevalence estimation of ARP using traditional methods raises serious problems.[3] Screening programmes based on self administered questionnaire or on biological markers, or both, have been used to estimate the prevalence of ARP.[4] Although screening provides important information regarding the proportion of undiagnosed cases,[1,2,5,6] the high costs,[4] the low response rates[7] and the questionable validity[3,4,8–10] make this methodology not feasible for monitoring ARP at a regional or national level.

Capture-recapture is an indirect method that generates prevalence estimates based on the degree of overlap between two or more separate samples of the population under study.[11] It was originally used in ecology to assess the size of animal population[12] and afterwards in demography to ascertain completion of census data.[13] Only more recently, capture-recapture techniques have been applied as attempts to estimate or adjust for the extent of incomplete ascertainment using information from overlapping lists of cases derived from distinct sources.[14] Capture-recapture methods have been applied to different epidemiological topics,[15–19] including the estimation of drug addiction,[20,21] of fetal alcohol syndrome[22] and of road accidents possibly caused by alcohol.[23] To our knowledge, however, no attempt to size ARP in a closed population using a capture-recapture approach has ever been made.

In this study, we estimated prevalence of subjects with ARP in a large community of northern Italy using multiple incomplete lists. The purpose of the study was to investigate the utility of capture-recapture techniques for the estimation of the size of ARP on a population basis.

Methods

CATCHMENT AREA

The study focused on all residents in the area of Voghera, a northern Italy rural area with an economy based on vine growing and wine production. The resident population aged more than 15 years was 132 618 according to the 1991 Italian Population Census (target population).

The study design was approved by the local ethical committee.

CASE DEFINITION

People resident in the catchment area who received throughout 1997: (a) diagnosis of dependence or abuse from alcohol and/or (b) diagnosis of alcohol related disease and/or (c) treatment and/or assistence for a problematic alcohol drinking behaviour and/or (d) treatment for an alcohol related disease, were considered as cases. This definition implicitly includes both persons for whom alcohol drinking is a current problem and those who, although actually abstainers, suffered in the past with dependence or misuse. The common basis that joins the captured cases and at the same time defines people with ARP is their need of treatment and assistance to achieve or to remain abstinence from alcohol over the time.

DATA COLLECTION

Four sources were used to identify patients with alcohol related disorders.

The first information source (F1) was the self help volunteering groups. In the area one Alcoholics Anonymous group and three on-treatment alcoholics groups ("CAT" from the Italian expression Club Alcolisti in Trattamento)[24] were operating. Each of them was asked to flag alcoholics attending for one or more times the group during 1997. The registry of the meetings made during this period was used to prepare the complete list. People who attended the group to be supportive of others with ARP (for example, friends or family of cases) were not included in the list.

The second (F2) and the third (F3) sources were the three Psychiatric Ambulatory and the Public Alcohology Service of the area, respectively. Each of them was asked to flag subjects receiving diagnosis of alcohol dependence and/or misuse during 1997. The registry of the consulting room activities of this period was used to prepare the complete list.

The fourth source (F4) was the computerised database of the patients discharged from the General Hospital of Voghera during 1997 with a primary or secondary diagnosis of alcoholic psychoses (ICD-9 code: 291), alcohol dependence syndrome (303), non-dependent abuse of alcohol (305.0), alcoholic polyneuropathy (357.5), alcoholic cardiomyopathy (425.5), alcoholic gastritis (535.3), alcoholic fatty liver (571.0), acute alcoholic hepatitis (571.1), alcoholic cirrhosis of the liver (571.2) and unspecified alcoholic liver damage (571.3).

In addition, all the 163 family physicians of the area were asked to flag patients with ARP. As only seven physicians responded to our request and all the 74 flagged patients were present in the F1 data source, the family physicians list was not considered in the analysis.

Included in the survey were patients resident in the area and flagged from one or more of the four sources because attending (F1) or receiving care and assistance (F2, F3, F4). Patients resident outside from the catchment area and those attending or receiving care and assistance before or after the study period, but not in the study period, were excluded from the study.

Collected information included initials of name and surname (two digits), gender (one digit), date of birth (six digits), municipalities of birth (five digits) and of residence (three digits) and family physician (three digits). In this way an unique 20 digits identification code was constructed.

RECORD LINKAGE

Determination of unique persons between sources was carried out by the following four step procedure. The first step linked the 26 people who had exactly the same identification code. In a second step a 20% sample of identification codes was randomly extracted from the list of each source. By the comparison between the sample of codes and the information directly reported by the registries (F1, F2 and F3) or by the clinical records (F4) only a 5%

incorrect digits were found, without any evidences of more frequent disagreement for particular digits. Thus, in the third step, a further record linkage was applied to previously unlinked persons permitting the detection of further six cases with a high probability of being the same person, as the identification code differed by a single digit. Finally, in the fourth step, validity of the match was verified tracing back the original documentation of the matched cases. No false positive matched records were identified. Confidentiality was maintained throughout the entire process, and the identities of all persons remained unknown.

DATA ANALYSIS

Capture-recapture techniques[11][25] were applied to estimate the total number of patients with ARP in the target population.

The lists were firstly compared in pairs estimating the total number of cases (N) according to the Chapman estimator[26]:

$$N = \frac{(n_1 + 1) \cdot (n_2 + 1)}{(n_{12} + 1)} - 1$$

where n_1 is the number of cases identified by a source; n_2 is the number of cases identified by another source; and n_{12} is the number of cases identified by both the sources. An approximate unbiased estimate of the variance of N was derived by Seber[27]:

$$\mathrm{Var}(N) = \frac{(n_1 + 1) \cdot (n_2 + 1) \cdot (n_1 - n_{12}) \cdot (n_2 - n_{12})}{(n_{12} + 1)^2 \cdot (n_{12} + 2)}$$

The 95% confidence intervals (95%CI) of N were calculated using the formula:

$$95\%\mathrm{CI} = N \pm 1.96\sqrt{\mathrm{Var}(N)}$$

Validity of the two sample estimator requires the following two main assumptions: (a) the sources of ascertainment are independent; (b) for a given source each case is equally likely to be listed by that source.

The crucial assumption is the independence among lists. This means that if two sources, say F2 and F3, are considered, each case should have an equal chance to be listed by source F2, independently to be identified by source F3. Conversely, if a patient flagged by the Psychiatric Ambulatory (F2) is more probably flagged by the Alcohology Service (F3) than a patient not identified in the Psychiatric Ambulatory, the number of cases in the target population will be underestimated (positive dependence). Naturally, also a negative dependence might occur inducing an overestimation of N. Therefore, a two sample capture-recapture estimate is probably biased.

The second assumption requires that each patient of the target population has the same probability to be captured by any source.[28] Variable catchability inducing different probability of ascertainment, probably affects patients with ARP for a number of characteristics, such as gender and age.

In this study, the dependence among sources was evaluated by means of two methods.

The first one is based on the simple comparison between all the two sample Chapman estimators. Large discrepancies between pairwise estimates provide evidence that the two sources being evaluated were not independent.[29] The comparison between each pairwise estimate and the total number of ascertained cases was afterward performed with the aim to show relations among sources and to obtain indications on the direction of bias. In this way, a two source estimate much lower (or much higher) with respect to the total number of ascertained cases provide evidence of a positive (or negative) dependence among the two sources.

The second method is based upon a variant of the log-linear model for complete data.[30 31] Data were arranged in the 2^4 incomplete multi-way contingency table with one missing cell corresponding to absence in all sources. Log-linear models were fitted to this contingency table to estimate the number of missing cases, taking into account the pattern of association between sources. As the table has 15 (= 2^4–1) observations (corresponding to the presence in one, two, three or four sources), any model may contain at most 14 terms. These are four independence terms (each corresponding to the main effect of a source), six first order interaction terms (each corresponding to the interaction effect between two sources) and four second order interaction terms (corresponding to the interaction effect between three of the considered sources). Interaction terms represent different aspect of the dynamics of the population or of its response to the sampling procedure.[31] This means that if the four independence terms and a first order interaction term are considered in the model, the corresponding estimate of the number of missing cases results adjusted for dependence among the two sources. Dependence may be inferred by the estimates obtained from the model with only independence terms (say estimate A) and from the model including also a first order interaction term (say estimate B). If estimate A has a higher (or lower) value than estimate B, we may deduce a negative (or positive) dependence between the sources considered in the interaction term. In the current application, among all the possible models, the best fitting one was selected starting from the model with only independence terms and adding hierarchically, step by step, interaction terms till when no significant improvement of the goodness of fit of the model was obtained (forward selection strategy). The model that fitted the observed cells with the fewest possible parameters, allowing for various dependencies among the sources, and with the smallest variance was chosen to cover the missing cell, thus yielding an estimate of the total population size.[19] The frequency for the missing cell was estimated as antilog of a linear combination of the parameters of the chosen model.[31] The corresponding 95% CI were obtained from the variance-covariance matrix evaluate at convergence.[32] A more complex alternative approach based on the goodness of fit of the models was also used.[33] However, as

similar intervals were obtained from the two approaches, we chose to report only the variance based confidence intervals because of their major general accessibility.

By considering now both, dependence among sources and heterogeneous catchability among persons, we may hypothesise that each list tends to preferentially cover different subsets of the population and to interact with each other. It should be noticed that source dependence and variable catchability are, in fact, intertwined concepts because the heterogeneous probability of capture might be the underlying explanation for an observed dependence among sources. Thus, the estimation methods should take into account both dependence among sources and heterogeneous catchability among people in a unique model. In the current application, several log-linear models were again fitted to estimate missing cells taking into account at the same time dependence and heterogeneity. The above described 2^4 contingency table was stratified by gender and age. Because of the small number of flagged cases, only two age classes were used (< 50 and $\geqslant$ 50 years). The resulting stratified contingency table has four missing cells (each corresponding to absence in all sources for an age and gender stratum) and 60 (= 2^6–4) observations. The main effects of gender and age specify the probability of capture of persons on the basis of their demographic characteristics. The interaction term between gender (or age) and a variable coding a source, specifies the tendency of that source to cover persons of a gender (or age) category. Obviously, dependence among sources and heterogeneity among people may at the same time be taken into account by means of second order interaction terms among gender (or age) and first order interaction terms among capture occasions. Other more sophisticated methods suggested from the statistical literature[34 35] as well as the inclusion of more detailed age classes and/or other stratifying variables able to explain different capture probabilities among people, were not used in the current application because the small sample size.

For all the considered models the parameters were estimated by maximising the likelihood function. The goodness of fit of each model was assessed by the residual deviance (D statistics). The D statistics have asymptotic χ^2 distribution under the null hypothesis with df obtained by the difference between observed cells and parameters estimated. The comparison between two models, when feasible, was tested by the difference between the values of the D statistics for the two models (likelihood ratio test: LRT). Again, LRT statistics have asymptotic χ^2 distribution with df obtained by the difference between the df values of the D statistics of the two models compared.[36]

The corresponding calculations have been carried out using the GLIM package.[37]

For all hypothesis tests p values of less than 0.05 were considered as significant.

Table 1 Data from prevalent cases of subjects with alcohol related problems organised according to source of flagging (Voghera, northern Italy, 1997)

Flagging sources	Men		Women		
	<50 years	≥50 years	<50 years	≥50 years	All
Patients flagged by one source					
F1	30	40	6	12	88
F2	31	12	7	7	57
F3	12	5	3	2	22
F4	46	81	12	16	155
Patients flagged by two sources					
F1 - F2	2	1	0	0	3
F1 - F3	1	0	0	0	1
F1 - F4	1	2	0	0	3
F2 - F3	2	2	0	1	5
F2 - F4	3	3	1	0	7
F3 - F4	3	1	0	0	4
Patients flagged by three source					
F1 - F2 - F3	0	0	0	0	0
F1 - F2 - F4	0	0	0	0	0
F1 - F3 - F4	1	0	0	0	1
F2 - F3 - F4	1	0	1	0	2
Patients flagged by four sources					
F1 - F2 - F3 - F4	1	0	0	0	1
Total	134	147	30	38	349

Flagging sources: F1: self help volunteering groups (anonymous alcoholics and on treatment alcoholic groups); F2: Psychiatric Ambulatory; F3: Public Alcohology Service; F4: hospital discharges. Among the 74 cases flagged by seven responder family physicians (unreported in the table) all were also flagged by F1, 2 by both F1 and F2, 8 by both F1 and F4, 1 by F1, F2 and F4.

Results

FROM FLAGGING ARCHIVE TO CASE ARCHIVE

Patients were flagged from the self help volunteering groups (97 patients of whom 74 from Alcoholics Anonymous and 23 from CAT), from the Psychiatric Ambulatory (75), from the Public Alcohology Service (36) and from hospital discharge list (173). Primary or secondary diagnoses reported on hospital discharge records were: alcoholic psychoses (8 inpatients), alcohol dependence syndrome (45), non-dependent abuse of alcohol (40), alcoholic polyneuropathy (11), alcoholic fatty liver (7), acute alcoholic hepatitis (28), alcoholic cirrhosis of the liver (39), and unspecified alcoholic liver damage (4).

Among the 381 flagged records, an archive of cases of 349 patients was constructed, with an average per patient flagging frequency of 1.09. There were 281 men and mean age was 51.7 years (SD = 15.4).

Data from prevalent cases organised according to source of flagging are shown in table 1. Among the 349 patients, 322 were flagged from one source, 23 from two sources, three from three sources and one from all the considered sources. Corresponding frequencies were 257, 21, 2 and 1 and 65, 2, 1 and 0 in men and women respectively, so that the 8.5% of men but only the 4.4% of women were flagged by more than one source. By stratifying the patients in two age classes (< 50 and ≥ 50 years), 147, 13, 3 and 1 (younger) and 175, 10, 0 and 0 (older) were flagged by one, two, three and four sources respectively, so that the 10.4% of the younger but only the 5.4% of the older patients were flagged by more than one source.

Although men were flagged with major frequency with respect to women (80.5% *v* 19.5%), relevant differences among sources in the proportion of captured men were not

Table 2 Estimated prevalent count of subjects with alcohol related problems resulting from combining the six couples of flagging sources (Voghera, northern Italy, 1997)

Sources	Number*	(95% CI)*
F1 - F2	1489	(366, 2611)
F1 - F3	906	(171, 1640)
F1 - F4	2841	(837, 4845)
F2 - F3	311	(153, 470)
F2 - F4	1201	(592, 1810)
F3 - F4	714	(339, 1090)

*Number of subjects with alcohol related problems, and corresponding 95% confidence intervals. Flagging sources: F1: self help volunteering groups (anonymous alcoholics and on treatment alcoholic groups); F2: Psychiatric Ambulatory; F3: Public Alcohology Service; F4: hospital discharges.

observed (81.4%, 77.3%, 81.0% and 82.6% of men listed by F1, F2, F3 and F4, respectively). By using the capture-recapture terminology, we may therefore deduce a homogeneous catchability of sources among gender. Conversely, younger and older subjects were flagged with about the same frequency (50.4% *v* 49.6%). This is an effect of the design, as the age classes were chosen with the aim to obtain balanced categories. However, relevant differences among sources in the proportion of captured younger subjects were in this case observed (45.3%, 65.3%, 81.0% and 40.4% of younger patients listed by F1, F2, F3 and F4, respectively). We may therefore deduce a heterogeneous catchability of sources among the considered age classes.

TWO SAMPLE CAPTURE-RECAPTURE ESTIMATES

Table 2 shows the estimated numbers of patients with ARP by applying the two sample capture-recapture approach to all the possible combinations of two source of ascertainment. Only one of the six estimates was lower than the aggregate prevalent count (349) suggesting a positive dependence between F2 (psychiatric ambulatory list) and F3 (public alcohology service list). Conversely, the highest prevalent count (2.841) suggests a negative dependence between F1 (self help groups list) and F4 (hospital discharges list).

LOG-LINEAR MODEL BASED ESTIMATES

Table 3 shows the estimated prevalent count of subjects with ARP resulting from several log-linear models. The first model that entailed only the main effects of the four sources did not fit the data (tabulated $\chi^2_{0.05;10}$ = 18.3). Conversely, models 2–5 (that considered also any interaction terms) had lower residual deviance and showed significant goodness of fit (tabulated $\chi^2_{0.05;9}$ = 16.9; $\chi^2_{0.05;8}$ = 15.5; $\chi^2_{0.05;7}$ = 14.1). In particular, the model 3 that considered two first order interaction terms (F2*F3 and F3*F4) showed a significant improvement of the goodness of fit with respect to the model 2 that considered only one first order interaction term (LRT: D_2-D_3 = 6.38 with 1 df, p < 0.05). Adding to the model 3, one first (F1*F4: fourth model) or second (F2*F3*F4: fifty model) order interaction term, significant improvements of the goodness of fit were not obtained (D_3-D_4 = 1.51 with 1 df, p > 0.05; LRT: D_3-D_5 = 1.37 with 1 df, p > 0.05). Thus, choosing the more parsimonious model (model 3) an

Table 3   Log-linear based estimates of prevalent count of subjects with alcohol related problems. Models include interaction terms involved to the adjustment of estimates for dependence among sources. Summary of forward selection strategy of the best model (Voghera, northern Italy, 1997)

Model	df†	D†	Number (95% CI)‡
(1) Main effects only $(u_1+u_2+u_3+u_4)$	10†	31.133★	1492 (1121, 2082)
(2) Main effects + 1 two factor interaction $(u_1+u_2+u_3+u_4+u_{23})$	9	16.718★	1818 (1306, 2703)
(3) Main effects + 2 two factor interactions $(u_1+u_2+u_3+u_4+u_{23}+u_{34})$	8	10.343★	2297 (1524, 3794)
(4) Main effects + 3 two factor interactions $(u_1+u_2+u_3+u_4+u_{23}+u_{34}+u_{14})$	7	8.830★	1845 (1164, 3367)
(5) Main effects + 2 two factor interactions + 1 three-factor interaction $(u_1+u_2+u_3+u_4+u_{23}+u_{34}+u_{234})$	7	8.976★	2813 (1487, 4564)

†Degrees of freedom of the model and corresponding residual deviance. ‡Number of subjects with alcohol related problems estimated according to the parameters of the log-linear model and corresponding 95% confidence intervals. ★p<0.05; the tested null hypothesis is that the model fits the data; alternative hypothesis is a significant departure of the model from the data. u_i: parameter of the main effect of the i-th flagging source (i = 1,..,4); u_{ij}: parameter of the two factor interaction between the i-th and the j-th flagging sources (i = 1,..,4; j = 1,..,4); u_{ijk}: parameter of the three factor interaction between the i-th, the j-th and the k-th flagging sources (i = 1,..,4; j = 1,..,4; k = 1,..,4); Flagging sources: F1: self help volunteering groups (anonymous alcoholics and on treatment alcoholic groups); F2: Psychiatric Ambulatory; F3: Public Alcohology Service; F4: hospital discharges.

estimate of about 2300 patients with ARP was forecast in the target population. As this estimate is much higher with respect to that obtained from the main effects model (2297 *v* 1492), a positive dependence between F2 (psychiatric ambulatory list) and F3 (public alcohology service list), and between F3 and F4 (hospital discharges list) may be deduced.

Table 4 shows the estimates of prevalent count based on models that include interaction terms involved the adjustment of estimates for dependence among sources and/or for heterogeneity among gender or age. From the first three models, as well as from model 1 of table 3, the same estimate of prevalent count was obtained. This is explained by the absence of interaction terms in the four compared models. An improvement of the goodness of fit were obtained adding to the model 1 the term of the main effect of gender $(D_1-D_2 = 139.6$ with 1 df, p < 0.05). This simply indicates a significant difference between the number of flagged men and women. It should be noticed, however, that no terms involving the interaction between gender and one or more capture occasions were selected. Conversely, age did not show a main effect $(D_1-D_3 = 1.3$ with 1 df, p > 0.05), but a significant tendency of both, F2 and F3 to cover different age classes $(D_1-D_4 = 12.3$ with 2 df, p < 0.05) was observed. These results suggest, as already noticed as comment to table 1, that the considered age classes, but not gender, were responsible for heterogeneous catchability among sources. Models 1–4, however, did not fit the data (tabulated $\chi^2_{0.05;55} = 73.3$; $\chi^2_{0.05;54} = 72.2$; and $\chi^2_{0.05;53} = 71.0$). Conversely, model 5 that entailed either main effects, dependencies and heterogeneity, showed a significant goodness of fit (tabulated $\chi^2_{0.05;49} = 66.3$). On the basis of this model, an estimate of about 2500 patients with ARP was forecast in the target population.

Discussion

The results provide the first estimate of the size of ARP in a general population based on capture-recapture methodology. Although 349 persons of the target population were captured from the four local sources because they were attending or receiving care and assistance during 1997, we estimated that the number of persons with ARP was about 2500 with a prevalence of 19 every 1000 inhabitants aged more than 15 years.

It is not possible to compare these estimates with those of other studies directly. However, a 11.6 % prevalence of positive answers to one or more questions of the CAGE questionnaire has been recently reported[38] in the Italian population.[39] Literature reports that CAGE scale, when used with one or more positive answers, achieves a sensitivity of 86% and specificity of 93% in detecting ARP.[40] By applying these values at the observed prevalence of positive to the questionnaire, a true prevalence of alcohol related disorders of 5.8% is expected. These considerations suggest that the administration of a validated instrument of screening to a sample of population is an inconsistent way to size ARP in a target population.

This study shows another, but not alternative, approach to size ARP based on capture-recapture methods. Routinely collected incomplete lists are easily available and multiple lists with variable degree of completeness (hospital discharges, alcoholic and psychiatric services, self help associations, etc) can be obtained cheaply in most geographical areas. After

Table 4   Log-linear based estimates of prevalent count of subjects with alcohol related problems. Models include only the main effects of sources, gender and age (two classes: <50 and ≥50 years) (models 1–3) or also several interaction terms among these variables (models 4–5). Summary of forward selection strategy of the best model (Voghera, northern Italy, 1997)

Model	Main effects	Interaction effects	df	D†	Number (95% CI)‡
(1)	All sources $(u_1+u_2+u_3+u_4)$	—	55	215.8★	1492 (1901, 2560)
(2)	All sources and gender $(u_1+u_2+u_3+u_4+u_g)$	—	54	76.2★	1492 (1147, 2414)
(3)	All sources and age $(u_1+u_2+u_3+u_4+u_a)$	—	54	214.5★	1492 (1927, 2502)
(4)	All sources $(u_1+u_2+u_3+u_4+u_{a2}+u_{a3})$	F2 and age, F3 and age	53	203.5★	1499 (1998, 2536)
(5)	All sources, gender and age $(u_1+u_2+u_3+u_4+u_{23}+ u_{34}+ u_g+ u_a+u_{a2}+u_{a3})$	F2 and F3, F3 and F4, F2 and age, F3 and age	49	32.7★	2523 (1623, 4627)

†Degrees of freedom of the model and corresponding residual deviance. ‡Number of subjects with alcohol related problems estimated according to the parameters of the log-linear model and corresponding 95% confidence intervals. ★p<0.05; the tested null hypothesis is that the model fits the data; alternative hypothesis is a significant departure of the model from the data. u_i: parameter of the main effect of the i-th flagging source (i = 1,..,4); u_{ij}: parameter of the two factor interaction between the i-th and the j-th flagging sources (i = 1,..,4; j = 1,..,4); u_g: parameter of the gender effect (g = 0,1); u_a: parameter of the age class effect (a = 0,1); u_{ai}: parameter of the two factor interaction between age and the i-th flagging source (a = 0,1; i = 1,..,4). Flagging sources: F1: self help volunteering groups (anonymous alcoholics and on treatment alcoholic groups); F2: Psychiatric Ambulatory; F3: Public Alcohology Service; F4: hospital discharges.

record linkage of all available lists, the methods described in this paper can be easily applied at a very low cost by researchers and the public health administration.[19]

A number of methodological issues that are potential limitations to the capture-recapture methodology should be however considered.

The most important problem is the difficulty of delimiting appropriately the population under study.

Because of our strict definition of case, our estimates referred to alcoholics who seek or need treatment, so that only agencies deputed to the treatment of alcoholics were considered in this study. Agency lists include people with problems related to more severe forms of alcohol related disorders.[21] This necessarily leads to underestimate the size of ARP in the target population and might explain the lower prevalence obtained by applying the capture-recapture methodology with respect to that obtained from the CAGE administration (see above).

In almost all the epidemiological applications, capture-recapture methods point out the estimate of the total population actually affected by a disease or condition. In the current application, we captured also subjects characterised by problematic use of alcohol in the past, but actually abstainers. For example, several people who have not drunk for many years attend self help groups or alcoholic services with the aim to receive assistance in maintaining abstinence. Analogously, long term alcohol related diseases (for example, alcoholic cirrhosis) may be acquired because of heavy drinking pattern in the past but with patients actually abstainers. Also for these people, however, the abstinence keeping is required to avoid the disease worsening. Thus, in the field of the ARP, it is very important to count all the subjects who need treatment and assistance to achieve or to keep abstinence from alcohol, independently by their current alcoholic behaviour.

As Hook and Regal[41] discussed, a major limitation of the capture-recapture methods is that, in virtually all the epidemiological applications, you cannot formally establish whether any estimate is in fact unbiased. The unique way to take precautions against the risk of obtaining biased estimates, is the attempt to verify that the underlying assumptions of the methods are at least plausible in any particular application. This is that has been made in the current study considering the four major assumptions that must be satisfied to produce reliable counts from capture-recapture methods.

The first assumption is that the target population is closed. In reality, no population is closed, thus this assumption can only be satisfied to a reasonable degree. If the alcoholic population is very dynamic, with members entering and leaving with great frequency, the recapture will be less likely to include people identified in the first capture. This leads to an overestimate of the population of alcoholics.[42] To allow for the assumption that the alcoholic population was reasonably stable during the

study period, we limited the duration of the observation to one year.

The second assumption is that false-positive subjects should not be present on any list.[43] In our study, it is assumed that all the listed people have ARP. In the current application, people who attended self help groups to be supportive of others with ARP (for example, friends or family of alcoholics), patients who received care and assistance from the psychiatric ambulatory or the public alcoholic service with diagnosis different from alcohol dependence and/or abuse (for example, dependence from other substances), and patients discharged with a diagnosis not directly indicating the alcohol aetiology (for example, liver cirrhosis without mention of alcohol), were not included in the lists.

The third assumption is that the capture sources are independent. In reality, this assumption is not required if more than two captures are used and an algorithm is used to account for dependence among captures.[29 31] Despite this, you should be aware that not all the possible interactions are appreciable by any algorithm, especially when sample size is small. In the current application, although the two sample capture-recapture estimates suggested a negative dependence between F1 (self help groups lists) and F4 (hospital discharges list), the corresponding interaction term did not significantly improve the goodness of fit of the log-linear model accepted. This might lead to an overestimate of the prevalent count of subjects with ARP.

The fourth assumption is that for any single source each case in the population has the same probability of ascertainment, although any two sources may differ in this probability.[14] In contrast, we can reasonably hypothesise that each list tends to preferentially cover different subsets of the population. It is very difficult, in an epidemiological observational framework, to control every putative source of heterogeneity among people. For example, we observed that considering age as a heterogeneity source in the log-linear model, improvements of the goodness of fit of the model were obtained. In particular, the psychiatric ambulatory and the

public alcoholic service tended to cover preferentially younger subjects. Although an analysis stratified by gender and age has permitted us to correct the estimate, we cannot exclude that other important strata remain unidentified and may contribute to a biased estimate. Confidence intervals of our counts, already very wide when only dependence among sources was considered (1524, 3794), were even wider when also gender and age effects were taken into account (1623, 4627).

From a health service point of view, it is very important to know the proportion of the total estimated population who have attended different services. Our data suggest that of every 1000 subjects who seek treatment for alcohol related disorders resident in Voghera, only 14 attended the Public Alcohology Service, 30 the Psychiatric Ambulatory, 38 the self help volunteering groups and 68 were discharged with a primary or secondary alcohol related diagnosis. Moreover, only 7.7% of the cases were flagged by more than one source, suggesting a scarce collaboration among services and groups in the management of alcoholics. These data may change over time and space depending on the type and availability of treatment, on the ability of physicians to identify patients with ARP and on the management of patients discharged with an alcohol related diagnosis.

It might be speculated that, in the specific issue of the treatment of alcoholics, overlapping among sources might be considered as a proxy indicator of the functioning of the network among public services and self help volunteering groups supporting care and assistance of alcoholics and facilitating their abstinence from alcohol.[44] Two main positive dependencies among sources were observed in our application. Firstly, people flagged by the Psychiatric Ambulatory (F2) were more pobably also flagged by the Alcohology Service (F3). Secondly, people discharged for alcohol related diseases (F4) were more probably flagged by the Alcohology Service. These findings suggest a good functioning of the network between the public services of the area involved in the treatment of alcoholics. On the other hand, as the two sample estimate considering intersection between hospital and self help groups was much higher with respect to the other two sample estimates, then the patients discharged for alcohol related diseases were not preferentially sent to the self help groups of the area. These findings suggest that the proposed method, involving the comparison between multiple information sources, is potentially useful for the health planners to verify the functioning of the network among the public services and self help volunteering groups.

Our study suggests that capture-recapture is an appropriate approach for estimating the prevalence of ARP when different lists of alcoholics can be obtained from different types of agencies involved with problematic use of alcohol. However, the use of this method requires careful attention to the underlying assumptions discussed above and cautious interpretations.

1 Fleming MF, Barry KL, Manwell LB, *et al*. Brief physician advice for problem alcohol drinkers: a randomized controlled trial in community-based primary care practices. *JAMA* 1997;**277**:1039–45.
2 Fleming MF, Manwell LB, Barry KL, *et al*. At-risk drinking in an HMO primary care sample: prevalence and health policy implications. *Am J Publ Health* 1988;**88**:90–3.
3 Midanik LT. Validity of self-reported alcohol use. *Br J Addict* 1988;**83**:1019–30.
4 Litten RZ, Allen JP. *Measuring alcohol consumption: psychosocial and biochemical methods*. Totowa, NJ: Humana Press, 1992.
5 Wallace P, Haines A. Use of a questionnaire in general practice to increase the recognition of patients with excessive alcohol consumption. *BMJ Clin Res Ed* 1985;**290**:1949–53.
6 Conigliaro J, Lofgren RP, Hanusa BH. Screening for problem drinking: impact on physician behavior and patient drinking habits. *J Gen Intern Med* 1998;**13**:251–6.
7 Hill A, Roberts J, Ewings P, *et al*. Non-response bias in a lifestyle survey. *J Public Health Med* 1997;**19**:203–7.
8 Babor TF, Stephens RS, Marlatt GA. Verbal report methods in clinical research on alcoholism: response bias and its minimization. *J Stud Alcohol* 1987;**48**:410–24.
9 Fuller RK, Lee KK, Gordis E. Validity of self-report in alcoholism research: results of a veterans administration cooperative study. *Alcohol Clin Exp Res* 1988;**12**:201–5.
10 Bradley KA, Boyd-Wickizer J, Powell SH, *et al*. Alcohol screening questionnaires in women: a critical review. *JAMA* 1998;**280**:166–71.
11 International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation. I. History and theoretical development. *Am J Epidemiol* 1995;**142**:1047–58.
12 Chapman DG. The estimation of biological populations. *Ann Math Statist* 1954;**25**:1–15.
13 Chandrasekar C, Deming WE. On a method of estimating birth and death rates and the extent of registration. *J Am Stat Assoc* 1949;**44**:101–15.
14 Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* 1995;**17**:243–64.
15 Cates W JR, Smith JC, Rochat RW. Assessment of surveillance and vital statistics data for monitoring abortion mortality, United States, 1972–1975. *Am J Epidemiol* 1078;**108**:200–6.
16 Huether CA, Gummere GR. Influence of demographic factors on annual Down's syndrome births in Ohio, 1970–1979, and the United States, 1920–1979. *Am J Epidemiol* 1982;**115**:846–60.
17 Robles SC, Marrett LD, Clarke EA, *et al*. An application of capture-recapture methods to the estimation of completeness of cancer registration. *J Clin Epidemiol* 1988;**41**:495–501.
18 Cochi SL, Edmonds LE, Dyer K, *et al*. Congenital rubella syndrome in the United States, 1970–1985. On the verge of elimination. *Am J Epidemiol* 1989;**129**:349–61.
19 Bruno G, LaPorte RE, Merletti F, *et al*. National diabetes programs. Application of capture-recapture to count diabetes? *Diabetes Care* 1994;**17**:548–56.
20 Frischer M, Bloor M, Finlay A, *et al*. A new method of estimating prevalence of injecting drug use in an urban population: result from a Scottish city. *Int J Epidemiol* 1991;**20**:997–1000.
21 Domingo-Salvany A, Hartnoll RL, Maguire A, *et al*. Use of capture-recapture to estimate the prevalence of opiate addiction in Barcelona, Spain, 1989. *Am J Epidemiol* 1995;**141**:567–74.
22 Egeland GM, Perham-Hester KA, Hook EB. Use of capture-recapture analyses in fetal alcohol syndrome surveillance in Alaska. *Am J Epidemiol* 1995 **141**:335–41.
23 Aptel I, Salmi LR, Masson F, *et al*. Road accident statistics: discrepancies between police and hospital data in a French island. *Accident Analysis and Prevention* 1999;**31**:101–8.
24 Aricò S, Zannero A, Galatola G, *et al*. Family Compliance to a treatment programme for alcoholics: a prospective study of prognostic factors. *Alcohol Alcohol* 1994;**29**:679–85.
25 International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation. II. Applications in human diseases. *Am J Epidemiol* 1995;**142**:1059–68.
26 Chapman DG. Some proprieties of the hyper-geometric distribution with applications to zoological sample census. *Berkeley, University of California Publications in Statistics* 1951;**1**:131–60.
27 Seber GAF. The effect of trap response on tag recapture estimates. *Biometrics* 1970;**26**:13–22.
28 Hook EB, Regal RR. Effect of variation in probability of ascertainment by sources ("variable catchability") upon capture-recapture estimates of prevalence. *Am J Epidemiol* 1993;**137**:1148–66.
29 Wittes JT, Colton T, Sidel VW. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *J Chron Dis* 1974;**27**:25–36.
30 Fienberg SE. The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika* 1972;**59**:591–603.

31 Cormack RM. Log-linear models for capture-recapture. *Biometrics* 1989;**45**:395–413.

32 Bishop YMM, Fienberg SE, Holland PW. *Discrete multivariate analysis: theory and practice*. Cambridge, MA: MIT Press, 1975:229–56.

33 Hook EB, Regal RR. Goodness-of-fit based confidence intervals for estimates of the size of a closed population. *Stat Med* 1984;**3**:287–91.

34 Darroch JN, Fienberg SE, Glonek GF, *et al*. A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *J Am Stat Ass* 1993;**88**:1137–48.

35 Agresti A. Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics* 1994;**50**:494–500.

36 McCullagh RJ, Nelder JA. *Generalized linear models*. Cambridge: Chapman and Hall University Press, 1988.

37 Baker RJ, Nelder JA. *The GLIM System. Release 3.77*. Oxford: Numerical Algorithms Group, 1985.

38 Ewing JA. Detecting alcoholism: the CAGE questionnaire: a critical review. *JAMA* 1984;**252**:1905–7.

39 Osservatorio Permanente sui Giovani e l'Alcool. *Gli italiani e l'alcool. Consumi, tendenze ed atteggiamenti*. Rome: Vignola Editore, 1998.

40 Liskow B, Campbell J, Nickel EJ, *et al*. Validity of the CAGE questionnaire in screening for alcohol dependence in a walk-in (triage) clinic. *J Stud Alcohol* 1995;**56**:277–81.

41 Hook EB, Regal RR. Internal validity analysis: a method for adjusting capture-recapture estimates of prevalence. *Am J Epidemiol* 1995;**142**:S48–52.

42 Mastro TD, Kitayaporn D, Weniger BG, *et al*. Estimating the number of HIV-infected injection drug users in Bangkok: a capture–recapture method. *Am J Public Health* 1994;**84**:1094–9.

43 Brenner H. Use and limitations of the capture-recapture method in disease monitoring with two dependent sources. *Epidemiology* 1995;**6**:42–8.

44 Tam TW, Schmidt L, Weisner C. Patterns in the institutional encounters of problem drinkers in a community human services network. *Addiction* 1996;**91**:657–69.