Journal of Official Statistics
Vol. 6. No. 2, 1990, pp. 165–177

Retrospective Questions and Group Differences

Dirk Sikkel[1]

Abstract: Responses about actions undertaken in the past usually contain recall errors. Sikkel (1985) develops models that correct for memory effects that allow greater accuracy in the estimation of parameters that describe a series of actions. Here, these models are extended to estimate group differences in the frequency of certain actions. To do this a family of models is developed that contain (i) parameters for an underlying distribution of actions, (ii) parameters that quantify the memory effects and (iii) one parameter that characterizes the differences between subpopulations for the average frequency of certain events. Examples are taken from a health survey where respondents were asked to recall the number of contacts with the general practitioner and hospitalizations for a given period.

Key words: Stochastic process; negative binomial distribution; memory effects.

1. Introduction

We study data from retrospective questions on the number of events (e.g., visits to a doctor) that took place during a given interval. This interval always precedes the interview and the respondent is also asked to date the occurrence of these events. Responses to retrospective questions have the advantage of being inexpensive to collect, but have the disadvantage of containing memory mistakes. We use a model-based approach to analyze such data. The aim here is to analyze differences between subgroups of the population. Sikkel (1985) introduced a class of models that aim at removing memory mistakes from retrospective responses. These models consist of two parts. First, the number of events is described by a Poisson process, with a stochastic intensity parameter. Second, an event is forgotten with a certain probability, depending on the time elapsed and the number of events already reported. Here we use one model for each group, but with the assumption that all parameters but one are equal for all groups. That parameter characterizes the groups and their differences. By keeping the number of parameters small, the estimator of the essential parameter has a relatively low standard deviation. This technique works well, provided that the simplified model is a fairly good description of the data generating process. The simplification with equal parameters is testable, as shown in Section 4.

Our approach can be compared with synthetic estimators used for small groups and areas (see, e.g., Fay and Herriot 1979; Purcell and Kish 1980; or Fuller and Harter 1987). Such a model-based estimator has a relatively small standard deviation, but there is also a bias unless the model is perfect. The mean square error of a synthetic estimator may still be small in comparison

[1] Research International Nederland, P.O. Box 1257, 3000 BG Rotterdam, The Netherlands.

with that of an unbiased sample-based estimator for each group or area separately, due to small sample sizes.

The two parts of the basic model, i.e., the process of events and the forgetting, are described in Sections 2 and 3, respectively. The modeling of subpopulations is discussed in Section 4. In Section 5 the variables used to define subpopulations in the 1981 Health Survey conducted by the Netherlands Central Bureau of Statistics are discussed. In Sections 6 and 7 empirical results are presented with respect to the reported number of contacts with the general practitioner (GP) and with the number of hospitalizations, respectively. Section 8 contains our conclusions.

2. The Model for Actions

We will concentrate on the following type of questions: "How many times did you undertake action a during τ time units immediately preceding the interview and when did these actions take place?" This type of question was asked in the 1981 Dutch Health Survey to determine, e.g., the number of contacts with the general practitioner (i.e., phone calls, visits to the GP, and visits by the GP) or the number of hospitalizations. In the case of the GP, the respondents were asked to report the number of contacts during the last three months and for every contact (up to a maximum of six) respondents were asked how many weeks before the interview the contact took place. We assume that the actions of an individual can be described by a Poisson process; every individual, however, has his or her own frequency parameter. Let M_i be the number of actions of individual i in reference period T with length τ, then

$$P\{M_i = m\} = e^{-f_i\tau}\frac{(f_i\tau)^m}{m!}, \tag{1}$$

where f_i is the number of actions for individual i per time unit. This model implies that the intervals between two consecutive actions of individual i are i.i.d. random variables, exponentially distributed with parameters $1/f_i$. Of course, this model does not apply to every event or action. For example, the interval between two "teeth brushings" is certainly not exponentially distributed. On the other hand, the distribution assumption holds when the action or event is "random" enough and thus (1) can be a good approximation of reality.

The frequencies of actions f_i can be regarded as scores of individuals on some latent variable F. In our model we assume that the distribution of F over the population can be approximated by a gamma-distribution with parameters b and k. Its density function is

$$g(f) = \frac{e^{-f/b}f^{k-1}}{\Gamma(k)b^k}. \tag{2}$$

Here b is a scale parameter; when time is measured in weeks instead of days, b increases by a factor of 7. The parameter k is the inverse of the squared coefficient of variation of F. The larger k is, the smaller the variation between individuals. The parameters b and k determine the mean and variance of the gamma-distribution. This makes the gamma-distribution very flexible. By assuming that the f_i are gamma distributed, we do not impose serious restrictions on the model. Apart from flexibility, computational convenience is a strong argument for choosing the gamma-distribution for F. The number of actions of a randomly drawn individual denoted by M, has a negative binomial distribution (see e.g., Johnson and Kotz 1969), i.e.,

$$P\{M = m\} = \frac{k(k+1)\ldots(k+m-1)}{m!} \times \frac{(b\tau)^m}{(1+b\tau)^{m+k}}. \tag{3}$$

As we have a sample with randomly distributed scores f_i, (3) is a general equation to derive the likelihood function from. A slight complication to this likelihood function is that questionnaires sometimes leave room for only a limited number of actions, due to lack of space. We define m^+ to be the maximum number of actions that can be registered in the questionnaire. Then $\{M = m^+\}$ denotes {a randomly drawn individual has undertaken m^+ actions or more} and $P\{M = m^+\}$ is calculated by summing $P\{M = m\}$ in (3) from m^+ to infinity.

3. The Models for Memory Effects

It is well known that answers to retrospective questions are almost always affected by memory errors. For this reason, many statisticians consider such questions to be unreliable. Nevertheless retrospective questions can be used with relatively inexpensive modes of data collection. Usually, the effect of unreliable memory of the respondent is related to time (how long ago did an event take place?), see e.g., Som (1973), Sudman and Bradburn (1974) and Moss and Goldstein (1979). Conceptually, there are two different types of effects that play a role in data that are based on retrospective questions:

1. Bias of an estimator, due to underreporting of events because the respondent incorrectly remembers the number of events that took place in a given interval. This is usually caused by forgetting, but can also be the result of systematic misplacement of events on the time axis by the respondent. This is called telescoping; it may lead to underreporting as well as overreporting of events as described in Schneider (1981).
2. Increase of variance of an estimator, because the respondent does remember the events, but places them incorrectly on the time axis, without affecting the bias of the estimator. Part of the research project described here was devoted to quantification of this increase of variance. This effect, however, appeared too small to be quantified.

Memory effects may also be related to the response burden of the respondent, i.e., the number of events to be reported. When there are many detailed events to be reported, respondents may be more inclined to forget (or leave out) events, than is the case when there is only little to report. With a large response burden even the interviewer may press the respondent to hurry up and leave out events as the total interview time grows unacceptably large. In our examples, the events were recorded backward in time. During the interview the most recent event was recorded first. Thus the response burden became gradually heavier as the events occurred in more distant time periods. With the use of a stochastic process, the dependence of the memory effects on the number of previously reported actions can be taken into account. We can calculate the expected number of actions, say three to six weeks ago given the reported number of actions in the first two weeks of a randomly drawn individual. And also, this number of expected actions can be compared with the number of actually reported actions. Of course, it would be more desirable to relate the memory effects to the number of *actual* actions, but this is impossible. The number of previously *reported* actions is our only tool for estimating the response burden. Without a stochastic model for the actions, the memory effects can only be studied as a function of time (which lead to conclusions like "the longer ago, the more actions are forgotten," as in the above mentioned literature).

We can formulate our models in mathematical terms as follows. The respondent reports his or her actions in a reference period T of length τ. This total reference period T is partitioned into t intervals T_1, $T_2, \ldots, T_t$ with lengths $\tau_1, \tau_2, \ldots, \tau_t$, respectively. The intervals usually are numbered by the index u, from most recent to most distant in time. Thus T_1 is the most recent interval on which the respondent reports; T_u precedes $T_{u'}$ in time if $u > u'$. It is assumed that in T_1 no actions are forgotten. This assumption is crucial, since we need a "true" reference point. In general, T_1 should be chosen so that it is as large as possible under the condition that the memory effects in T_1 are negligible. In $T_2, \ldots, T_t$ we allow for memory effects. Such effects may be caused by underreporting or overreporting of events, although in practice the memory effects concerned only the forgetting of events. Thus we will only use the word "forgetting" to describe the memory effects. We now assume that in an interval T_u the rate of forgetting is constant, but depends on u and j, the total number of actions that are reported in the intervals T_1, $T_2, \ldots, T_{u-1}$. To be more precise, we denote the probability that a respondent reports an action that he/she undertook in T_u by $v_{u,j}$. If he/she has undertaken more than one action in T_u he/she reports each of them with probability $v_{u,j}$, independent of each other. The index u denotes the effect of time on forgetting; the index j is a proxy for response burden.

Our model for memory effects can be understood as a model for changing time-scale parameters. A respondent with a large memory effect in T_u reports as if time goes slower in T_u. His/her number of reported actions is Poisson-distributed with the parameter $\tau_u(1 - v_{u,j})f_i$, given that he/she has reported j actions up to T_u. Thus he/she answers *as if* the T_u does not contain τ_u but

$$\tau'_{u,j} = \tau_u(1 - v_{u,j}) \tag{4}$$

time units. In Sikkel (1985) models of this type were investigated for medical consumption. The models were not tested on the individual level, i.e., assumptions like the independence of actions (going to the doctor) or forgetting were not examined for every respondent separately. Instead, the complicated multivariate analogue of (3), taking into account (4), was calculated and the models were tested by maximum likelihood. The expression of the likelihood function, although derived with elementary means, is very complicated and not presented here. There appears to be a good fit for three variables, namely "contacts with the general practitioner (GP)," "contacts with the specialist," and "hospitalizations." This is confirmed in Sikkel and Jelierse (1987), where model predictions for correlation coefficients are studied. Another validation can be found in Sikkel (1986) where model predictions of the number of individuals that have no contacts with the GP (or specialists) during a year is compared with empirical data, resulting in a satisfactory correspondence. The model was also tested for contacts with the dentist which usually take place every six months. This appeared far too regular for obtaining a reasonable fit.

Equation (4) implies that we have $(t - 1)^* m^+ + 2$ model parameters (including b and k). When t and m^+ are moderately large, the total number of parameters can be very large. For this reason it is desirable to reduce, if possible, the number of parameters by making more restrictive models. A more restrictive model also makes it easier to estimate b and k efficiently. Sikkel (1985) investigates a range of such models. Here we give two examples that we will use later.

The first example is the "0-1 linear time-number model." This model states that the memory effects depend on the time elapsed and on the number of previously reported actions. The model has the following functional form

$$v_{u,j} = \beta(\tau_1 + \cdots + \tau_{u-1} + \tfrac{1}{2}\tau_u) + \gamma \cdot \text{sign}(j). \quad (5)$$

Thus, the memory effect depends linearly on time with regression coefficient β. The dependence on the number of previously reported actions is a threshold relationship: only the distinction between "0 reported actions" (sign $(j) = 0$) and "more than 0 reported actions" (sign $(j) = 1$) is made. When there is at least one previously reported action, an effect γ is added to the linear time effect.

A second example is the "0-1 number model." Here the only relevant distinction is whether or not the respondent has previously reported any actions. Its functional form is

$$v_{u,j} = \gamma \cdot \text{sign}(j). \quad (6)$$

Thus, in every interval T_u before the interval in which the last action is reported, there is a memory effect γ.

4. The Models for Group Differences

The goal of this paper is to use the underlying Poisson-gamma process and the model for memory effects to estimate differences in medical consumption between different subpopulations. In principle there are two different ways to achieve this. The first is to use the models of the previous section separately for every subpopulation. Then we obtain for every subpopulation s an estimate $\bar{f}_s$ for the average frequency of, e.g., contacts with the GP. With models that require only a few parameters this method could be more efficient than making no model assumptions at all. In order to estimate differences between subpopulations s_1 and s_2 we simply subtract $\bar{f}_{s_1}$ from $\bar{f}_{s_2}$. There is, however, a second method that goes further and entails an extension of the model. Briefly, this method assumes that the memory effects in the different subpopulations are equal. The idea behind this assumption is that forgetting may be a general human flaw. This leaves only the parameters b and k to be different among subpopulations. Since we want to use only one parameter to express the differences between subgroups, it seems natural to assume that the coefficient of variation of the medical consumption is constant over the subpopulations (associated with the parameter k of the underlying gamma distribution). Then the differences between the subpopulations are expressed by b_s only, the scale parameter of the underlying gamma distribution for subpopulation s. In this way we are left with few enough parameters to expect a further increase in efficiency. Of course, if we allow more parameters, either the $v_{u,j}$ or k, to be different in the subpopulations, the models will fit the data better, but then the standard error will be larger than with a single different parameter.

We can formulate the extension of the model more accurately in the following way. At the individual level, M_i still has a Poisson distribution with parameter $f_i\tau$. In group s, however, the distribution of F over all individuals is approximated by a gamma distribution with parameters b_s and k. Because we focus on inference on differences between subpopulations (do women go to the GP more often than men?), we will always consider differences with respect to subpopulation 1. To this end we define $b_1 = b$

and $b_s = b(1 + \alpha_s)$, with $\alpha_1 \equiv 0$. This implies

$$P\{M_i = m|i \text{ in group } s\} = \frac{k(k+1)\ldots(k+m-1)}{m!} \times \frac{\{b(1+\alpha_s)\tau\}^m}{\{1+b(1+\alpha_s)\tau\}^{m+k}}. \quad (7)$$

We also have $\bar{f}_s = \bar{f}(1 + \alpha_s)$ with $\bar{f} = \bar{f}_1$.

The behaviour of an individual in the random sample is characterized by a vector $\mathbf{M}_i = (M_i(1), M_i(2), \ldots, M_i(t))'$, where $M_i(u)$ is the number of reported actions of individual i in $T_1 \cup T_2 \cup \ldots \cup T_u$. The vector of actions of a randomly selected individual is in the same way given by $\mathbf{M} = (M(1), M(2), \ldots, M(t))'$. Such a vector, which we will call a profile, consists of t nondescending integers between 0 and m^+, inclusive. The probability distribution of $\mathbf{M}_i|i$ in group s can be derived from the multivariate negative binomial distribution, the multivariate version of (5) using $\tau'_{u,j}$ instead of τ, along the same lines as in Sikkel (1985). From this probability distribution, the likelihood function can be maximized by standard methods.

Keeping k constant over the subpopulations is something like the assumption of homoscedasticity in linear regression. When it is not completely satisfied, we hope that it will not damage the estimators of the location parameters α_s or $\alpha_s \bar{f}$ too much. The assumption that the memory effects are identical in structure and have the same parameter values for all subpopulations is more risky. When, for the case of a dichotomous variable, there are hardly any memory effects in group 1 and large memory effects in group 2, these effects will be spread over the two groups in the estimation process. This leads to a considerable bias in the estimator of $\alpha_2 \bar{f}$, the difference in frequency between the groups.

Fortunately it is possible to do some testing to find out whether the model assumptions are reasonable. To this end we write the probability of $\mathbf{m}$ in group s as $p^{(s)}_{\mathbf{m}}$. Then without any extra assumptions $p^{(s)}_{\mathbf{m}}$ depends on b_s, k_s and $v^{(s)}_{u,j}$, $u = 2, \ldots, t$, $j = 0, \ldots, m^+ - 1$. Here we choose only one model for $v^{(s)}_{u,j}$ (for example the 0-1 linear time-number model). When we make the differences between groups part of the model, then we let $k_s = k$ and $v^{(s)}_{u,j} = v_{u,j}$ be constant over the groups. Hence this is a special case of the model where the groups are considered separately. Since we can estimate with the method of maximum likelihood, we can also test the dissimilarity between the two models by the likelihood ratio test. The test statistic is G^2 (see, e.g., Bishop, Fienberg, and Holland, 1975, p. 125). Under the null hypothesis of a constant k_s and a constant $v^{(s)}_{u,j}$ (in s) G^2 has a χ^2-distribution with a known number of degrees of freedom.

We now have three different types of assumption for the model. First, there is the assumption of the underlying Poisson distribution, of which the parameters are gamma distributed over the individuals. Second, memory effects are accounted for in the model. These memory effects may have a variety of structures as described in the previous section. Third and finally, it is assumed that the memory effects and the parameter k are identical for the different subpopulations and that the differences between the groups can be expressed by one scale parameter. Here this model is called the "common" model in contrast to the "separate" model, where k and the memory effects are estimated separately for the different groups.

5. Variables

The Health Survey conducted by the Netherlands Central Bureau of Statistics is a continuing survey in which many retrospective questions are asked with different reference periods. The sample is a cluster sample, stratified with respect to region and degree of urbanization. All individuals have equal probability of being selected for the sample. Here the data are treated as if they formed a random sample. Seasonal effects cancel out, as the data were collected during the entire year of 1981. Such effects, may, however, affect the estimation of the parameter k of the gamma distribution of the individual frequencies of events since they are a source of variation.

In Sikkel (1985) "contacts with the GP," "contacts with specialists," and "contacts with the dentist" were the subject of analysis. Because it appeared that the Poisson-gamma model did not apply to "contacts with the dentist" this variable was left out. "Contacts with specialists" is replaced here by "hospitalizations." This variable performs well with respect to the Poisson-gamma model (see Sikkel, 1986). In the case of contacts with the GP, the reference period was partitioned into six intervals, with $\tau_1 = 1.5$ weeks and $\tau_2 = \ldots = \tau_6 = 2$ weeks. In the case of hospitalizations there were also six intervals with $\tau_1 = 1.5$ months and $\tau_2 = \ldots = \tau_6 = 2$ months.

From the 1981 Health Survey six categorical variables have been chosen for analysis. These variables were not chosen because of their subject matter content. It seemed more important to choose variables that would lead to deviations of the model assumptions and provide empirical evidence on how this effects the estimations. The following variables have been used in the analysis:

1. age: 0–14 years, 15–64 years, 65+;
2. sex: male, female;
3. marital status: married, divorced, widow(er), never married;
4. urbanization: rural, smaller cities, three big cities;
5. proxy: interview with respondent, proxy interview;
6. kind of insurance: private, sickness-fund (obligatory health insurance).

With age, marital status, and proxy, differences in memory effects cannot be excluded *a priori* on the ground that older people may have a bad memory. Furthermore, marital status is strongly related to age and proxy reporting (proxy = "yes") is more difficult than reporting about oneself. With marital status there is also the problem that the groups "divorced" and "widow(er)" are relatively small. These groups have little influence upon the estimation of parameters over the groups. The power of the likelihood ratio test is relatively low with respect to the parameters of these small groups.

6. Contacts With the General Practitioner

From Sikkel (1985) it appeared that the linear 0–1 time-number model, as described by (5), gives the best description of the reported contacts with a GP. Here we consider only this model. In Table 1 different estimation procedures are compared. This has been carried out for three different methods. First, the averages, differences, and standard errors have been simply estimated without a model on the basis of a reference period of the first 1.5 weeks. According to our assumptions there are no memory effects in this period. In fact, it is the obvious method of estimation when no assumptions are made about the underlying stochastic process and the structure of memory effects. Second, estimates are given that are based on the separate model. These

Table 1. Contacts with the GP and background variables estimated with the 0–1 linear time model. Comparison of estimates, standard errors and quality of the fit. For each variable the average frequency in the first category is given and the differences from this average frequency in the subsequent categories.

	1,5 Weeks No Model		11,5 Weeks Sep. Model		11,5 Weeks Common Model		n	Quality Fit	
	$\hat{f}$	$\hat{\sigma}$	$\hat{f}$	$\hat{\sigma}$	$\hat{f}$	$\hat{\sigma}$		G^2	df
Age								231.1	6
0–14 years	0.0452	0.0039	0.0470	0.0030	0.0383	0.0011	2418		
15–64 years (diff.)	0.0168	0.0047	0.0161	0.0036	0.0266	0.0021	6801		
65 + (diff.)	0.0529	0.0094	0.0528	0.0064	0.0601	0.0057	999		
Sex								39.6	3
man	0.0537	0.0029	0.0537	0.0021	0.0506	0.0014	5144		
woman (diff.)	0.0159	0.0043	0.0166	0.0031	0.0228	0.0024	5074		
Marital status								55.6	9
married	0.0685	0.0032	0.0693	0.0023	0.0730	0.0020	5169		
divorced (diff.)	0.0517	0.0206	0.0464	0.0192	0.0423	0.0139	172		
widow(er) (diff.)	0.0472	0.0149	0.0369	0.0091	0.0261	0.0080	432		
never married (diff.)	−0.0225	0.0043	−0.0223	0.0031	−0.0299	0.0021	4445		
Urbanization								34.5	6
rural	0.0601	0.0062	0.0603	0.0045	0.0608	0.0019	1353		
smaller cities (diff.)	−0.0032	0.0067	−0.0035	0.0049	0.0004	0.0024	6911		
3 big cities (diff.)	0.0191	0.0083	0.0124	0.0058	0.0040	0.0034	1954		
Proxy								27.9	3
self	0.0762	0.0032	0.0752	0.0024	0.0782	0.0021	5502		
proxy (diff.)	−0.0317	0.0042	−0.0289	0.0031	−0.0354	0.0022	4716		
Kind of insurance								24.3	3
sickness-fund	0.0673	0.0027	0.0671	0.0019	0.0657	0.0017	7028		
private (diff.)	−0.0184	0.0045	−0.0162	0.0032	−0.0123	0.0023	3190		

estimates are based on a reference period of 11.5 weeks. Third, the averages, differences, and standard errors that are based on the common model are presented. For each variable the average frequency is estimated for the first category. For the other categories, the difference from the first category is estimated. These figures are given in the columns $\hat{f}$. Thus, the average frequency of the age group 0–14 years is estimated by 0.0452 without model, by 0.0470 with the separate model and by 0.0383 with the common model. The average frequency of the age group 15–64 years is estimated 0.0168 higher than 0.0452 without model, 0.0161 higher than 0.0470 with the separate model and 0.0266 higher than 0.0383 with the common model. These differences between the subpopulations are our main concern here. Besides the estimates, the standard errors of the estimates are also given. These are given in the column $\hat{\sigma}$. Note that, except for the first categories of each variable, these standard errors refer to the differences between the categories. The common model is tested by G^2 with the appropriate number of degrees of freedom against the alternative, the separate model. Note that when we have g groups, the common model has parameter b, k, β, and γ and $(g - 1)$ parameters α_s to indicate the differences between the groups, which makes a total of $(g + 3)$ parameters. The corresponding separate model has $4g$ parameters. Hence there are $(3g - 3)$ degrees of freedom in the likelihood ratio test.

From Table 1 it appears that the goal, lower standard errors by using fewer parameters, is reached. From the values of G^2, however, it appears that the common model does not fit the data perfectly. The clearest example of this is the variable age. Given the very large sample size, this is not a good reason to reject the estimation procedure. Any model is bound to fail the test when there are enough observations; with 1000 respondents there would have been a reasonable fit. It is remarkable that the differences in estimates between the model free approach and the separate model are relatively small. The differences between the separate and the common model are much larger, especially with age and marital status. This does not necessarily imply that the outcomes of the model free approach and the separate model are better (their standard errors are relatively high). Nevertheless, given the lack of fit, we have no definite evidence that the common model is better than the separate model.

Now we come to the question of identifying the exact differences between separate and common estimation of the parameters for the different groups. This question may be answered using Table 2. In this table the different model parameters are given. In the rows with the variable names the estimated parameters of the common model are given; in the rows with the category names the estimated parameters of the separate model are presented. Under the null hypothesis we would have: $\beta = \beta_1 = \beta_2 = \ldots, \gamma = \gamma_1 = \gamma_2 = \ldots$ and $k = k_1 = k_2 = \ldots$. It may be expected that the categorical background variable explains some variance. The total variance of F is equal to kb^2. Thus it seems reasonable to suppose that the variances within groups $(k_s b_1^2 (1 + \alpha_s)^2$ for $s = 1, 2, \ldots)$ are smaller than the overall variance as estimated by the common model (kb^2). This appears clearly not to be the case. The values of k_s have large fluctuations, especially with the age variable. Also, there are fluctuations in the memory effects, but large deviations are found only in the smaller groups (divorced, widow(er)). The bizarre estimates of β and γ for these groups decrease one's faith in the common model. The groups "divorced" and

Table 2. Estimates of model parameters and zero percentages on the basis of different models to compare groups for contacts with the GP

	b	k	β	γ	Percent no contact last year		n
					separate	common	
Age	0.061	0.633	2.4	23.5			
0–14 years	0.161	0.292	2.3	45.6	52.0	40.5	2418
15–64 years	0.103	0.614	2.2	23.5	32.2	31.0	6801
65+	0.057	1.161	2.7	9.9	11.1	24.7	999
Sex	0.084	0.604	2.4	24.1			
man	0.131	0.431	2.0	33.2	43.0	36.0	5144
woman	0.088	0.795	2.5	19.1	25.4	30.0	5074
Marital status	0.113	0.649	2.4	23.2			
married	0.098	0.706	2.0	22.5	27.9	28.7	5169
divorced	0.180	0.642	−1.1	56.6	22.3	22.1	172
widow(er)	0.039	2.752	5.5	10.8	4.8	24.1	432
never married	0.114	0.413	2.5	32.8	45.0	38.0	4445
Urbanization	0.105	0.058	2.3	24.6			
rural	0.135	0.445	3.0	9.9	39.5	33.9	1353
smaller cities	0.104	0.565	1.7	29.7	35.0	33.8	6911
3 big cities	0.091	0.797	3.7	17.0	24.8	32.8	1954
Proxy	0.121	0.647	2.4	23.4			
self	0.097	0.772	2.4	19.4	24.8	27.7	5502
proxy	0.111	0.416	2.2	36.2	45.1	38.2	4716
Kind of insurance	0.112	0.589	2.3	24.6			
sickness-fund	0.098	0.682	2.8	21.1	29.1	32.3	7028
private	0.121	0.417	1.0	32.1	43.8	35.7	3190

"widow(er)" seem to be too small to estimate such parameters.

In order to make the differences even more clear, Table 2 also contains the estimated percentages of respondents who have no contacts with the GP during an entire year. These percentages are calculated for the common model as well as the separate model. Taking a longer interval is one way of magnifying the effects of different parameter estimates. The percentage of respondents without contacts with the GP can be calculated simply by taking $m = 0$ and $\tau = 52$ and substituting of the estimated parameter values into (1). It appears that generally the commonly estimated percentages vary less between the groups than the separately estimated percentages. This is understandable, since such a percentage depends strongly on the variance of the distribution. It is the left tail of the distribution, and the larger the variance, the larger the tails. By assuming that the coefficient of variation of the distribution of F is constant over the groups, the distribution of contacts over the groups (artificially) obtains some common variance. This leads to more equal tails within each group.

7. Hospitalizations

Sikkel (1986) showed that the yearly number of hospitalizations is so small that it did not

Table 3. Hospitalizations and background variables in the number model. Comparison of the estimates and standard errors and quality of the fit. For each variable the average frequency in the first category is given and the differences from this average frequency in the subsequent categories.

	1,5 Weeks No Model		11,5 Weeks Sep. Model		11,5 Weeks Common Model		n	Quality Fit	
	$\hat{f}$	$\hat{\sigma}$	$\hat{f}$	$\hat{\sigma}$	$\hat{f}$	$\hat{\sigma}$		G^2	df
Age								3.0	4
0–14 years	0.0055	0.0013	0.0047	0.0005	0.0046	0.0003	2418		
15–64 years (diff.)	0.0019	0.0016	0.0023	0.0007	0.0023	0.0005	6801		
65+ (diff.)	0.0038	0.0028	0.0112	0.0016	0.0077	0.0017	999		
Sex								1.1	2
man	0.0057	0.0009	0.0058	0.0004	0.0057	0.0004	5144		
woman (diff.)	0.0026	0.0014	0.0020	0.0007	0.0022	0.0006	5074		
Marital status								5.8	4
married	0.0088	0.0011	0.0083	0.0006	0.0079	0.0005	5169		
widow(er) (diff.)	−0.0057	0.0024	−0.0019	0.0019	0.0069	0.0028	432		
never married (diff.)	−0.0030	0.0015	−0.0036	0.0007	−0.0032	0.0006	4445		
Urbanization								5.7	4
rural	0.0059	0.0017	0.0059	0.0008	0.0059	0.0005	1353		
smaller cities (diff.)	0.0005	0.0019	0.0007	0.0004	0.0008	0.0006	6911		
3 big cities (diff.)	0.0047	0.0027	0.0023	0.0009	0.0020	0.0009	1954		
Proxy								4.1	2
self	0.0084	0.0010	0.0084	0.0006	0.0083	0.0005	5502		
proxy (diff.)	−0.0027	0.0014	−0.0034	0.0007	−0.0032	0.0006	4716		
Kind of insurance								0.2	2
sickness-fund	0.0073	0.0009	0.0071	0.0004	0.0072	0.0004	7028		
private (diff.)	−0.0006	0.0015	0.0009	0.0007	−0.0011	0.0006	3190		

make sense to distinguish between separate groups of 3, 4, 5, and 6 hospitalizations. For this reason only the categories 0, 1, and "2 and more" have been used in the estimation process. The number model fitted best for the whole population. This implies that the memory effect does not depend on the time that has passed since the hospitalization but does depend on the number of events that have occurred since.

From Table 3 it appears that in no case is there a significant deviation (significance level 0.05) of the common model from the separate model. Compared to the model-free estimations the common model's standard errors are considerably smaller, and also compared to the separate model the common model shows an increase in precision. Exceptions are the relatively small groups "65+" and "widow(er)." For "65+" the separate model has the smallest standard deviation. The group "widow(er)" also has a small standard deviation; there the standard error in the model free estimate is smaller than in the common model.

The difference between the parameter estimates of models are shown in Table 4. There appears to be large fluctuations in the memory effects. Since the number of hospitalizations is relatively small, this does not lead to a rejection of the hypothesis of the common model, as seen in Table 3. The

Table 4. Estimates of model parameters and zero percentages on the basis of different models to compare groups for hospitalizations

	b	k	v_1	Percentage no contact last 5 years		n
				separate	common	
Age	0.025	0.183	38.5			
0–14 years	0.019	0.243	−13.2	83.0	84.5	2418
15–64 years	0.044	0.158	46.3	81.5	80.5	6801
65+	0.045	0.253	34.5	72.0	74.4	999
Sex	0.033	0.176	37.8			
man	0.028	0.210	36.5	81.4	82.6	5144
woman	0.049	0.162	37.1	80.2	79.4	5074
Marital status	0.044	0.181	39.4			
married	0.050	0.167	51.0	79.4	79.2	5169
widow(er)	0.023	0.451	−72.6	67.9	72.5	432
never married	0.030	0.157	32.4	85.1	84.3	4445
Urbanization	0.033	0.176	37.2			
rural	0.016	0.359	−65.2	78.2	82.4	1353
smaller cities	0.041	0.160	38.6	81.9	81.2	6911
3 big cities	0.044	0.185	60.1	78.7	79.5	1954
Proxy	0.044	0.187	36.9			
self	0.042	0.201	44.0	77.7	78.5	5502
proxy	0.032	0.153	21.3	84.8	83.5	4716
Kind of insurance	0.040	0.177	36.8			
sickness-fund	0.038	0.187	33.2	80.1	80.4	7028
private	0.041	0.152	46.4	82.8	82.1	3190

percentages of "no hospitalizations in the past five years" are close together. The largest deviation appears in the group "widow(er)."

8. Conclusion

The most radical method to model and estimate differences between subpopulations is to parameterize the model in such a way that one parameter indicates the differences, while the other parameters are kept constant over the subpopulations. In our study this method led to a considerable decrease in standard errors and the avoidance of implausible parameter estimates. Since we do not have the population values of the parameters there is no way to be sure that the estimates in the common model are superior to other estimates. However, the estimates are wrong only when the models are misspecified and the estimates are not robust against misspecification. If the within group variation (k) is misspecified we have the analogy with linear regression where ordinary least squares is applied instead of generalized least squares. The estimators of averages are robust against moderate misspecifications of this kind. More serious are misspecifications in the memory effects since they are approximately proportional to the estimators of the true averages. When the subpopulations are large, the possibility of misspecification can easily be tested and its consequences can be investigated.

9. References

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975). Discrete Multivariate Analysis. Cambrige (USA): MIT Press.

Fay, R.E. and Herriot, R.A. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. Journal of the American Statistical Association, 74, 47–65.

Fuller, W.A. and Harter, R.M. (1987). The Multivariate Components of Variance Model for Small Area Estimation, in Small Area Statistics. An International Symposium, eds. R. Platek, J.N.K. Rao, C.E. Särndal, and M.P. Singh, New York: Wiley.

Johnson, N.L. and Kotz, S. (1969). Discrete Distributions. New York: Wiley.

Moss, L. and Goldstein, H. eds. (1979). The Recall Method in Social Surveys. London: University of London Press.

Purcell, N.J. and Kish, L. (1980). Postcensal Estimates for Local Areas. International Statistical Review, 48, 3–18.

Schneider, A.L. (1981). Methodological Problems in Victim Surveys and Their Implication for Research In Victimology. Journal of Criminal Law and Criminology, 72, 818–838.

Sikkel, D. (1985). Models for Memory Effects. Journal of the American Statistical Association, 80, 835–841.

Sikkel, D. (1986). Methods of Analysis of Memory Effects Applied to the Health Survey 1981. Statistical Essays M27, Staatsuitgeverij, Den Haag. (In Dutch.)

Sikkel, D. and Jelierse, G. (1987). Retrospective Questions and Correlations, Psychometrika, 52, 251–261.

Som, R.K. (1973). Recall Lapse in Demographic Enquiries. New York: Asia Publishing House.

Sudman, S. and Bradburn, N.M. (1974). Response Effects in Surveys. Chicago: Aldine.

Received April 1987
Revised December 1988