Survey Research Methods (2011)
Vol.5, No.1, pp. 27-37
ISSN 1864-3361
http://www.surveymethods.org

Two Measures for Sample Size Determination

Philippe Eichenberger
Swiss Federal Statistical Office

Beat Hulliger
University of Applied Sciences
Northwestern Switzerland

Jann Potterat
Swiss Federal Statistical Office

Social surveys often must estimate the sizes or the proportions of many small groups and differences among them. The discussion of the needed precision of the estimators and the corresponding sample size is difficult, in particular when lay persons are involved. Two measures are developed which help in this discussion of the precision. These measures are called precision resolutions. The first of these measures, the size resolution, is derived from approximations to the probability of not observing a group in a sample and the second measure, the difference resolution, addresses the difference of two proportions. The precision resolutions are operationalisations of the smallest group or difference which can be estimated from a sample. Since they embody elements of statistical hypothesis tests without the need of a complete test specification they are simple to specify and nevertheless contain the necessary elements for sample size determination. The precision resolutions lead to the determination of the sample size for simple random samples but extensions to more complex samples are possible with the help of the design effect. The precision resolutions were developed for the planning of the Swiss Population Survey and this survey as well as the European Social Survey serve as examples of their application.

Keywords: resolution, sample design, precision, relevance, small proportion, small group, small difference

1 Introduction

During the discussion on the sample size for a survey multiple objectives have to be taken into account. In social surveys often the main purpose is to estimate the size of particular groups for various domains. For example, the number of persons which care for an elderly person should be estimated for the districts of a canton. The traditional formula for the sample size when estimating one proportion is not sufficient when many rather small proportions and differences between proportions must be estimated. This was the case for the discussion of the sample size of the Swiss Population Survey, a large scale survey which partially replaces the Swiss Census from 2010 onwards. Other examples are surveys to estimate the prevalence of a set of diseases or surveys to estimate the size of various consumer segments.

In principle, the problem of the necessary size of a simple random sample without replacement when the sampling proportion is used to estimate an unknown population proportion is solved by applying the formula $n = \tilde{p}(1 - \tilde{p})/\tilde{V}$, where $\tilde{p}$ is an advance guess of the unknown proportion and $\tilde{V}$ is the desired variance to achieve. In addition, a finite population correction may be used. In the case where no advance guess $\tilde{p}$ is available, or when many different proportions have to be estimated, often the maximizing value $\tilde{p} = 0.5$ is used. An alternative approach is proposed by Noble et al. (2006) by assuming a Beta-binomial distribution of the guessed $\tilde{p}$ and derive an expected minimum sample size. Of course the specification of a distribution is more involved than specifying $\tilde{p}$ alone. However, the major problem seems to be that often the client cannot specify his desired precision $\tilde{V}$. For example Cochran dedicates a chapter to the determination of sample size (Cochran 1977, Chapter 4) but gives little guidance on how to elicit the desired precision from the user. In addition, often there are many clients which must reach an agreement on the desired precision and worse, there are many domains of different size for which estimates of proportions are needed. Discussing how to reach at an advance guess of the variance $S^2 = \tilde{p}(1-\tilde{p})$ when determining the sample size, Kish (1965:52) says that the "statement on the 'adequacy' or 'desired' variance is generally subject to more vagueness than is S^2 itself. This is especially true when the survey has several objectives, with conflicting demands on the desirable sample size." The discussion on sample size may then become very difficult. To assist such discussions with more intuitive measures of precision, the precision resolutions were developed during the design of the Swiss Population Survey. The precision resolutions are general measures, applicable for many surveys. The Swiss Population Survey and the European Social Survey are used as a illustrative examples here.

The specification of a precision requirement is based on the theory of statistical hypothesis tests (see, for example Bickel and Doksum 1977). The reason why clients have

Contact information: Philippe Eichenberger, Swiss Federal Statistical Office, e-mail: philippe.eichenberger@bfs.admin.ch

problems in specifying the needed precision even of a single proportion is that, implicitely, he or she must formulate a hypothesis test. The client must have in mind a null hypothesis and an alternative hypothesis which he or she would like to distinguish from the null hypothesis because he or she would decide differently in case of the alternative than the null hypothesis. In addition the client must specify the maximum risk of wrong decisions he or she is willing to take. In other words the probabilities of the error of type I and type II of a hypothesis test must be specified. In short, a statistic should be sufficiently precise to be able to detect a relevant difference with sufficient reliability. This is equivalent to saying that the power of a test for a null hypothesis should be sufficiently large at a relevant alternative. Clients are seldom trained to specify the power of a test (cf. for example Lenth 2001). In addition there is a wealth of approximations available for the calculation of the power (Sahai and Khurshid 1996) which are difficult to distinguish for the lay person but concentrate on the technical problem of approximation and not on the problem of communication with users. The precision resolutions proposed here should facilitate the discussion about the needed precision of a survey without power calculations and specification of a desired variance.

Two surveys are used to illustrate the concepts: The Swiss Population Survey and the European Social Survey 2008 in Switzerland. The Swiss Population Survey is based on a yearly sample of 200 000 persons and a sampling rate of $f = 0.0274$. The sample is stratified by municipalities and has proportional allocation with slight modifications to ensure a minimal sample for each municipality. In the discussion about the sample size the resolution has been an important support to arrive at an agreement. The European Social Survey 2008 is a stratified random sample with 1 819 respondents, which corresponds to a sampling rate of $f = 0.000283$.

The sample design considered in the article is stratified random sampling with proportional allocation. Extensions to more complex sample designs with equal probability selection are covered with the help of the design effect (Kish 1965).

In this article we concentrate on the precision of an estimator and do not consider accuracy. Accuracy is measured by the mean squared error, i.e. accuracy includes bias, while precision is measured by variance alone. Of course, in practice, once the question of precision is settled the question of bias is the next preoccupation (or even first). We treat only the variance of estimators assuming, at least for planning purposes, that the estimators are unbiased.

We do not treat estimation with quantitative variables though the variability of quantitative variables may be more important for the determination of sample size than estimators of proportion. The search for a compromise between different requirements on precision, be it for the variance of means of quantitative variables, of regression coefficients or of proportions, is rather a policy issue than a statistical problem and will not be treated here.

We propose two related measures: the size resolution and the difference resolution. The estimation of the size of a small group in a domain of the population is treated in Section 2. First, the probability of not observing any member of a group and tolerance intervals for proportions are discussed. Then Section 2.3 introduces the size resolution, which is based on a normal approximation of the probability of not observing any member of a group. The comparison of proportions and a corresponding difference resolution is discussed in Section 3. The size resolution and the difference resolution are applied to the Swiss Population Survey and to the European Social Survey of 2008 in Section 4. Section 5 draws some conclusions.

2 Estimation of a group size

The first problem we consider is the estimation of the size N_A of a group A in a domain D of size N_D. Thus A is a subset of the sub-population D, $A \subset D$, and $N_A < N_D$. The sample design considered is stratified random sampling and we assume for simplicity that the sample size n_D in the domain D is fixed, in other words, that the domain is a stratum.

For example, in 2000, $N_A = 158$ was the number of Italian speaking persons in the municipality Veyrier with $N_D = 8\,892$ inhabitants. The sample size of the Swiss Population Survey in Veyrier, corresponding to the planned sample size $n = 200\,000$ on the level of Switzerland, is $n_D = 244$. Are we able to estimate N_A with sufficient precision by a simple random sample of that size in Veyrier? What exactly means sufficiently precise? A variant of the size-estimation problem is the estimation of the proportion $p_A = N_A/N_D$ of a group. The proportion of Italian speaking persons in Veyrier was 1.8% in 2000. Can we specify how precise we want to estimate such a proportion?

2.1 Unobserved groups

It may happen that a group of the population is not represented at all in a sample, i.e. there is no person of the group in the sample. In that case we can say that the group has not been detected by the sample. We now look at the probability of such an event in a domain which is a stratum of the sample design. The sample size is $n_D = f_D N_D$ in D for domain sampling rate f_D. For each possible sample a different number n_A of members of the group A may be observed. Due to the simple random sampling without replacement in D the number n_A has a hypergeometric distribution. The probability that A is not detected is the probability of $n_A = 0$. This probability should be reasonably low, say maximum 5%, for a group whose size is considered relevant.

Table 1 shows for different sample sizes n_D the minimal group sizes N_A in order to achieve the requirement $P[n_A = 0] \leq \alpha$, i.e. the probability that no member of A is in the sample is at most α, with $\alpha = 0.025$ or 0.05. The size of the minimal observable group depends mainly on the sampling rate and less on the domain size N_D. For example with a sample of size $n_D = 28$ in a domain of size $N_D = 1000$ the minimal observable group size is $N_A = 121$ ($\alpha = 0.025$) while for $n_D = 274$ at $N_D = 10\,000$ it is $N_A = 123$. Thus the probability that a small group is not observed in the sample sets a lower limit to what should be considered a relevant group size. However, the direct calculation of the tail probability of the hypergeometric distribution is not suited as a simple measure for the discussion with the users of a survey.

2.2 *Tolerance interval for the sample proportion*

In order to determine a simple expression for the minimal group size, which can be expected to be observed, we use the tolerance interval of a proportion which is based on the variance of the estimator of a proportion. The estimation of the proportion $p_A = N_A/N_D$ of a group $A \subset D$ based on a simple random sample of size n_D is a standard problem in survey sampling (Cochran 1977:50). The number n_A of persons from sup-population A in the sample can be written as a sum over an indicator variable $1\{i \in A\}$ which is 1 for persons from A and 0 otherwise, i.e. $n_A = \sum_{i=1}^{n_D} 1\{i \in A\}$. The classical estimator is the sample proportion $\hat{p}_A = n_A/n_D$, which is the sample mean of the indicator variable and thus the Horvitz-Thompson estimator. Under simple random sampling $\hat{p}_A$ is an unbiased estimator of p_A.

The variance of $\hat{p}_A$ under simple random sampling without replacement is

$$\sigma^2(\hat{p}_A) = (1 - f_D)\,\frac{p_A(1 - p_A)}{n_D}\,\frac{N_D}{N_D - 1}. \tag{1}$$

The variance of the sample proportion is the same whether one estimates p or $(1 - p)$ and we may switch to $(1 - p)$ when $p > 0.5$. Therefore we do not treat explicitly proportions larger then 0.5. The size of the group A is estimated by $\hat{N}_A = N_D\hat{p}_A$ and the variance of $\hat{N}_A$ is $N_D^2\sigma^2(\hat{p}_A)$.

We may try to elicit requirements on the precision of $\hat{p}_A$ from a user by directly asking him or her how small the variance $\sigma^2(\hat{p}_A)$ or the standard deviation $\sigma(\hat{p}_A)$ should be. The answer of the user would depend on the problem he or she has in mind, in particular on the group A of interest and the domain or municipality D and their respective sizes. A user may have many situations in mind and give many different answers. There is no simple answer to the requirement of a single user and less so for the many different users of multi-purpose multi-user surveys.

Often the question on the needed precision of a sample is posed in the form of a confidence interval. Then the user must decide what is a relevant width of a confidence interval (see, for example Cochran 1977:75). Here we are planning the survey and we may use a tolerance interval instead of the confidence interval to simplify the treatment. A tolerance interval is a central region of the distribution of $\hat{p}_A$ which contains a specified proportion, for example 95% of the outcomes. More generally a tolerance interval for a random variable X is an interval I which fulfils $P[X \in I] = 1 - \alpha$ for a small positive proportion α. Note that the tolerance interval is a statement on the distribution of X, not on the parameters of the distribution. Often a tolerance interval must be estimated from the sample but here we know that the distribution is hypergeometric and we assume known the proportion p_A. Values which are not included in the tolerance interval are very rarely observed. A trained user may state at which distance from the true value he would need the values to have only a small probability of being observed. We then may say that an estimator is sufficiently accurate when the sample yields a tolerance interval which respects this relevant distance.

An approximate tolerance interval for $\hat{p}_A$ based on the assumption of a normal distribution is

$$[p_A - z\,\sigma(\hat{p}_A), p_A + z\,\sigma(\hat{p}_A)], \tag{2}$$

where the constant z is a quantile of the standard normal distribution, i.e. $z = \Phi^{-1}(1 - \alpha/2)$ for some small $\alpha > 0$. The width of the tolerance interval depends on the sample size through $\sigma(\hat{p}_A)$. For a 95% tolerance interval we have $\alpha = 0.05$ and thus $z = 1.96$. In other words the numbers between $p_A - 1.96\,\sigma(\hat{p}_A)$ and $p_A + 1.96\,\sigma(\hat{p}_A)$ make up a 95% tolerance interval. Of course we could use another constant than $z = 1.96$ to obtain a different tolerance level. The lower limit of the tolerance limit (2) is set to 0 at least and the upper limit to 1 at most to prevent absurd results for extreme cases.

The true distribution of $\hat{p}_A$ is (scaled) hypergeometric and the normal distribution is an approximation only. The quality of the approximation of the hypergeometric distribution of n_A by the normal distribution depends on the population and sample sizes and on the proportion p_A to be estimated. For small sample sizes n_D and small p_A the hypergeometric distribution is strongly asymmetric, giving a high probability to a value of 0. For confidence intervals many proposals to deal with the asymmetry when the number of observed members of a group is small exist (see, for example Korn and Graubard 1998). Fortunately the coverage property for the tolerance interval (2) is better than for the corresponding confidence interval because the parameters p_A and $\sigma(\hat{p}_A)$ are known. We assume that the domain is not too small here, say $N_D > 100$. Then the coverage of the 95% tolerance interval (2) is above 90% when the expected sample in A is larger than 1, i.e. when $n_D\ p_A > 1$, and for moderate sample size and group size the coverage is close to 95%. It seems reasonable to assume that the expected sample of a group should be at least 1 to consider its estimation. Phrased differently we assume that we only consider group proportions p_A which are larger than $1/n_D$.

The tolerance interval (2) depends on p_A, n_D and f_D. Table 2 shows some tolerance intervals. A lower limit of 0 usually indicates that the formula (2) gives, in fact, a negative lower limit, which is then set to 0 explicitly as mentioned above. For example, with a domain of size $N_D = 500$ and a proportion of $p_A = 0.2$, which corresponds to a group of size $N_A = 100$, we would still obtain a lower limit of the tolerance interval of -3, which is set to 0. In other words we cannot exclude with 95% of confidence that nobody from group A would be in the sample. This aligns well with the probability of 4% for this event under the hypergeometric distribution. Note that in this case the expected number of elements from group A in the sample is $n_A = f_DN_Dp_A = 2.74$ and the approximate 95% tolerance interval has a coverage of 0.958.

Table 1: Minimal group size N_A for which the probability of observing 0 is below $\alpha = 0.025$ or 0.05.

	$\alpha = 0.025$		$\alpha = 0.05$	
n_D	$N_D = 1000$	$N_D = 10000$	$N_D = 1000$	$N_D = 10000$
28	121	1232	101	1014
50	70	710	57	581
100	35	361	29	294
274	12	123	10	108
500	6	72	5	59

Table 2: 95% tolerance intervals for $\hat{N}_A = N_D\hat{p}_A$ with $f_D = 0.0274$.

N_D	$p_A = 0.005$	$p_A = 0.01$	$p_A = 0.05$	$p_A = 0.20$	$p_A = 0.50$
500	[0 , 21]	[0 , 31]	[0 , 81]	[0 , 203]	[121 , 379]
1000	[0 , 31]	[0 , 47]	[0 , 131]	[51 , 349]	[314 , 686]
3000	[0 , 60]	[0 , 94]	[10 , 290]	[344 , 856]	[1180 , 1820]
10000	[0 , 132]	[0 , 216]	[245 , 755]	[1533 , 2467]	[4416 , 5584]
15000	[0 , 176]	[8 , 292]	[438 , 1062]	[2428 , 3572]	[6785 , 8215]
100000	[240 , 760]	[633 ,1367]	[4195 , 5805]	[18523 ,21477]	[48154 , 51846]
363273	[1320,2313]	[2932,4333]	[16630 ,19698]	[69839 ,75470]	[178117 ,185156]

2.3 Size resolution

The existence of a natural inferior bound for the detection of a group (see Subsection 2.1) stimulates the search for a description of precision which is directly linked to the estimation of the size of a small group. We look for an alternative way to express the probability of not observing a group, which is based on the tolerance interval and should be simpler. In other words instead of limiting $P[\hat{p}_A = 0]$ under the hypergeometric distribution we limit $P[\hat{p}_A \leq 0]$ under a normal distribution.

How large must a sample be to ensure that $P[\hat{p}_A \leq 0] \leq \alpha/2$? Phrased as a requirement for the size of a group given a particular sample size and with a tolerance interval: How large must a group be such that a $p_A \pm z\sigma(\hat{p}_A)$ tolerance interval does not include 0? The constant z can be chosen similar to a confidence interval: The probability of an estimate $\hat{p}_A = 0$ should be sufficiently low, for example below 2.5%. This choice leads to $z = 1.96 = \Phi^{-1}(0.975)$. At the same time as we give a low probability for the observed proportion $\hat{p}_A$ to be 0 we ensure a low probability that $\hat{p}_A$ is larger than a reasonable upper bound, namely roughly twice the theoretical proportion. For example if $p_A = 2\%$ we will have a low probability of obtaining $\hat{p}_A = 0$ or $\hat{p}_A > 4\%$. We think that this is a reasonable band for probabilities up to about 10% or even up to 25%.

The condition $0 \notin [p_A \pm z\sigma(\hat{p}_A)]$ implies more than avoiding an empty group in the sample because the variability $\sigma(\hat{p}_A)$ is taken into account explicitly. Thus in addition to limiting the probability of not observing a member of the group also the variability of the estimator is limited. This means that the condition $0 \notin [p_A \pm z\sigma(\hat{p}_A)]$ describes what group sizes N_A may be estimated with a certain minimal precision.

The condition $0 \notin [p_A \pm z\sigma(\hat{p}_A)]$ has a close relation to hypothesis testing. Fixing the confidence level of the tolerance interval is equivalent to fixing the level of a test, i.e. the probability of an error of type I. The test would have null hypothesis $p = p_A$ and alternative $p = 0$ and we implicitely state that $p_A - 0$ is a relevant difference. Under the alternative hypothesis $p = 0$ the sampling proportion would be 0 always and thus has a degenerate distribution. This implies that the test for $p = p_A$ would have infinite power at the alternative $p = 0$. Therefore, in a way, the requirement of the tolerance interval not including 0 also specifies the power of a test.

Thus given the sample size we try to determine a size or a proportion of a group which is just large enough to be estimated with sufficient precision from the sample. Sufficient precision is embodied by the limitation on $\sigma(\hat{p}_A)$ through the condition $0 \notin [p_A \pm z\sigma(\hat{p}_A)]$. We call this smallest estimable size the *size resolution* of the sample in analogy to the capability of optical devices to distinguish objects which are very small. We will derive the *difference resolution* in Section 3.1 where resolution is used in analogy to the capability of optical devices to distinguish objects which are close together. When it is necessary to distinguish the resolution measures we develop here from other resolutions, for example from sensor resolution, we call them precision resolutions.

We now proceed to derive a formula for the resolution and approximations to it. We may assume that $p_A < 0.5$ and therefore only the lower limit of the tolerance interval must be considered. The inequality that n_D must fulfil is

$$p_A - z\sigma(\hat{p}_A) = p_A - z\sqrt{(1-f_D)\frac{p_A(1-p_A)}{n_D}\frac{N_D}{N_D-1}} > 0. \quad (3)$$

Solving the inequality for p_A we obtain

$$p_A > \frac{z^2(1-f_D)\frac{N_D}{N_D-1}}{n_D + z^2(1-f_D)\frac{N_D}{N_D-1}} = \frac{a'}{n_D + a'}, \quad (4)$$

where $a' = z^2(1-f_D)\frac{N_D}{N_D-1}$. Multiplying both sides of inequality (4) by N_D yields an inequality for the size $N_A = N_D \ p_A$:

$$N_A > N_D \frac{a'}{n_D + a'}. \tag{5}$$

For practical purposes the approximation $N_D/(N_D-1) = 1$ is rarely problematic and we assumed that the domain sizes of interest to us are larger then 100. Thus we use $a = z^2(1 - f_D)$ only instead of a'.

Now we define the size resolution:

Definition: *The* $100 \cdot (1-\alpha)\%$ *size resolution of a simple random sample in a domain is*

$$R_s(1-\alpha, f_D, N_D) = N_D \frac{z^2(1-f_D)}{f_D N_D + z^2(1-f_D)} \tag{6}$$

where $z = \Phi^{-1}(1-\alpha/2)$ *is a standard normal quantile,* f_D *is the sampling rate and* N_D *is the size of the domain.*

We write the size resolution R_s with a subscript s to make the distinction with the difference resolution R_d introduced in Section 3. Of course $f_D N_D$ in the denominator is the sample size (or expected sample size) in domain D. We use $f_D N_D$ to make clear that there are, in fact, only two parameters involved. Thus the resolution does not depend on the proportion to estimate! This is a desirable feature because it simplifies the determination of the sample size.

The size resolution is the smallest size of a group which is estimable with a simple random sample. Here estimable means that the probability of a sample estimate of the size being larger than 0 and less than the double of R_s is approximately $1-\alpha$ and that the standard deviation of the estimator is less than R_s/z.

Solving (6) for f_D yields the sampling fraction necessary to estimate a group of size R_s:

$$f_D = \frac{z^2(1-R_s/N_D)}{R_s + z^2(1-R_s/N_D)}. \tag{7}$$

The corresponding sample size is, of course, $n_D = f_D N_D$. For a uniform sampling fraction we readily have the necessary total sample size as $n = f\ N$. The fraction R_s/N_D is the proportion of the minimal estimable group in the domain. For small groups, however, this dependence is weak. Also the dependence of the size resolution R_s on the domain size is weak and we will derive an approximate size resolution in the following that gets rid of this weak dependence.

The size resolution still depends on the size N_D of the domain. Using the inequality $1 + a/n_d > 1$ we arrive at

$$N_A > N_D \frac{a}{n_D + a} = \frac{N_D}{n_D} \frac{a}{1 + a/n_D} > \frac{a}{f_D} = \frac{z^2\ (1-f_D)}{f_D}. \tag{8}$$

If N_A fulfills (8) it also fulfills (3). Since $a = z^2(1-f_D)$ usually is small compared to n_D, for example $z^2(1-f_D) = 3.75$ for $z = 1.96$ and $f_D = 0.0274$, the relaxation of inequality (4) by (8) is often small. In inequality (8) neither the proportion p_A nor the size of the domain N_D is directly involved anymore. Knowing the sampling fraction f_D and fixing z, for example $z = 1.96$, we can calculate the minimal N_A.

We may also drop the finite population correction for small sampling rates f_D. Assuming $1 - f_D \approx 1$ the size resolution becomes

$$\tilde{R}_s = z^2/f_D. \tag{9}$$

We call $\tilde{R}_s$ the approximate size resolution when it must be distinguished from R_s.

Using the approximate size resolution (9) we can derive the necessary sampling fraction or sample size which is needed to estimate a group of size $\tilde{R}_s$:

$$f_D > \frac{z^2}{\tilde{R}_s}. \tag{10}$$

and the corresponding sample size is

$$n_D > \frac{z^2}{\tilde{R}_s/N_D}. \tag{11}$$

Thus the simplest form to use the size resolution is to calculate the desired proportion to estimate $p = R_s/N_D$ and to use z^2/p as the sample size.

Table 3 shows the 95% size resolutions at N_D = 1 000, 10 000 and 100 000 for selected sampling rates. Sampling rate $f = 0.0003$ corresponds to the European Social Survey and $f = 0.0274$ to the Swiss Population Survey. For low sampling fractions the size of the domain matters while for moderate and large sampling fractions it is of minor importance. For very low sampling fractions and low domain size the sample is set to 1 at least resulting in a resolution of $N_D z^2/(1+z^2)$.

Setting, in addition, $z^2 = 4$ we find an approximate 95% size resolution

$$\tilde{R}_s = 4/f_D = 4N_D/n_D. \tag{12}$$

Solving for the sampling fraction we obtain that an approximate 95% size resolution needs a sampling fraction $f_D > 4/\tilde{R}_s$, which corresponds to a sample size of

$$n_D = 4N_D/\tilde{R}_s. \tag{13}$$

This rough rule of thumb for the sample size yields $n_D = 400$ for a proportion of $R_s/N_D = 1\%$. In other words with a sample of size 400 any group of relative size 1% is estimable. This sample size is a customary result when requiring a 95% confidence interval of half length 5% for a population proportion of 50% (see, for example Cochran 1977:76). Note that the derivation used here is different from this customary derivation.

3 Estimation of a difference of proportions

In this Section we treat the estimation of the difference of the proportion of a group A in two disjoint domains D_1 and D_2 of size N_{D1} and N_{D2}. Thus we have two groups $A_1 = A \cap D_1$ and $A_2 = A \cap D_2$. We denote the size of the groups

Table 3: 95% size resolution R_s and approximate 95% resolution $\tilde{R}_s$.

	R_s for N_D =						
f_D	500	1000	3000	10000	15000	100000	$\tilde{R}_s$
0.0003	397	794	2381	5615	6517	11349	12805
0.0010	397	794	1684	2774	3056	3696	3842
0.0100	217	276	338	367	371	379	385
0.0274	106	118	130	135	136	137	141
0.0500	64	69	72	73	73	73	77
0.1000	33	34	35	35	35	35	39
0.5000	4	4	4	4	4	4	8

in the two domains by N_{A1} and N_{A2} and the corresponding proportions are $p_{A1} = N_{A1}/N_{D1}$ and $p_{A2} = N_{A2}/N_{D2}$. We are interested in the difference of proportions $\delta_A = p_{A2} - p_{A1}$. For example, we may want to know the difference of the proportion of Italian speaking persons in Veyrier and Lausanne, two municipalities of very different size. The difference in proportions in the same domain but at two time points is analogue if the sampling at the two time points is independent. Thus for Veyrier the proportions of Italian speaking persons for 1990 and 2000 are $p_{A1} = 172/7039 = 2.44\%$ and $p_{A2} = 158/8892 = 1.78\%$. The difference is $\delta_A = -0.66\%$. How accurate do we have to estimate such a change by a particular Survey? Alternatively we may want to estimate the absolute change $\Delta_A = N_{A2} - N_{A1}$. But since the sizes of the domain may be different usually the difference of proportions is the interesting characteristic.

We assume that we have two independent random samples either at the same time in two distinct domains or at two different times for the same domain. This differs from what we assumed for estimating one proportion where only one random sample was considered and a deviation from a hypothetical parameter, i.e. 0, was the focus. Now the variability of both samples must be considered.

The difference of proportions $\delta_A = p_{A2} - p_{A1}$ may be estimated by the difference of estimates of proportions $\hat{\delta}_A = \hat{p}_{A2} - \hat{p}_{A1}$. Since the two samples are assumed independent the variance of $\hat{\delta}_A$ is the sum of the variances of the estimates of proportion, i.e. $\sigma^2(\hat{\delta}_A) = \sigma^2(\hat{p}_{A2}) + \sigma^2(\hat{p}_{A1})$. An approximate tolerance interval based on the assumption of normal distribution of the two estimates of proportion is

$$[\delta_A \pm z\,\sigma(\hat{\delta}_A)]. \tag{14}$$

For a $(1-\alpha)\%$ tolerance $z = \Phi^{-1}(1-\alpha/2)$, for example $z = 1.96$ for $\alpha = 0.05$.

Here we assume that $p_{A1} < p_{A2}$, i.e. δ_A is assumed positive. In analogy to equation (3) for the resolution when estimating a single proportion we require that

$$\delta_A - z\sigma(\hat{\delta}_A) > 0. \tag{15}$$

This equation assumes an alternative hypothesis of $\delta_A = 0$ but in contrast to requirement (3) the distribution at the alternative is not degenerate. In other words, we cannot expect an infinite power at $\delta_A = 0$. In fact, the power at the alternative is approximately 0.5 since for $\delta_A = 0$ and approximately symmetric distributions we can expect that about half of the estimates result in $\hat{\delta}_A < 0$. Nevertheless the simple requirement in (15) contains a specification of a relevant difference, a level and a power of a test.

3.1 Difference resolution

We would like to obtain a simple expression for δ_A from inequality (15). A convenient parameterization is to use $p_{A1} = p - \delta_A/2$ and $p_{A2} = p + \delta_A/2$. In other words we express the requirement with the help of the mean of the involved proportions $p = (p_{A1} + p_{A2})/2$.

In the following we drop the subscript D to simplify the notation. For example N_{D1} will be written N_1 only.

Setting

$$a_i' = z^2 \frac{1-f_i}{f_i N_i} \frac{N_i}{N_i - 1}, i = 1, 2$$

and after some algebra, the following quadratic inequality is obtained:

$$\delta_A^2 > p(1-p)(a_1'+a_2') + 2\delta_A(1-2p)(a_2'-a_1')/4 - \delta_A^2(a_1'+a_2')/4.$$

The positive solution of this quadratic equation is

$$\delta_A > \left(p(1-p)b' + (1-2p)^2 c'^2\right)^{1/2} + (1-2p)c' = r_d'(p, N_1, N_2, f_1, f_2), \tag{16}$$

where

$$b' = \frac{a_1' + a_2'}{1 + (a_1' + a_2')/4} \text{ and } c' = \frac{(a_2' - a_1')/4}{1 + (a_1' + a_2')/4}.$$

The solution r_d' depends on 5 parameters: p, N_1, N_2, f_1, f_2. The last two parameters are the sampling fractions in D_1 and D_2, which might be different. The sample sizes $n_1 = f_1 N_1$ and $n_2 = f_2 N_2$ depend on these parameters. Note that the term under the square root is always positive. Obviously we must let depend the bound on the involved probabilities through $p = (p_{A1} + p_{A2})/2$.

As a function of p, r_d' has a positive value or is 0 at $p = 0$, increases until a maximum around 0.5 and then decreases again to reach 0 or a positive value at $p = 1$. Thus, as expected, the most difficult differences are the ones with proportions around 0.5. As a function of N_1 or N_2, r'_d is decreasing as expected. Figure 1 shows r'_d for different configurations of domain sizes and sampling fractions. Note that for p close to 0 or 1 the value for r'_d may be impossible to reach because $p - \delta_A/2 < 0$ or $p + \delta_A/2 > 1$. Increasing the domain sizes yields a lower curve. Different sampling fractions or different domain sizes make the curve asymmetric. The difference in r'_d when assuming equal domain size and equal sampling fraction to the true r'_d depends on the particular configuration but may be considerable.

The expression for r'_d is too complex for a simple communication with users in the planning stage of a survey. To simplify the expression we assume $N_1 = N_2 = N_D$ and $f_1 = f_2 = f_D$. This is at least approximately the situation when we consider the same domain under the same sample design at two near time-points where the change in domain size is negligible. In that case $a'_1 = a'_2$ and therefore $c' = 0$ and as a consequence the terms involving the factor c' in r'_d drop.

Assuming $N_D > 100$ the factor $N_D/(N_D - 1)$ in b'_i and a'_i is of minor importance and dropping it leads to $b_1 = b_2 = b$ and $a_1 = a_2 = a$ and therefore the function r'_d becomes

$$\begin{aligned} r_d(p, N_D, f_D) &= \sqrt{b\, p(1-p)} = 2\sqrt{\frac{a}{2+a}\, p(1-p)} \\ &= 2\sqrt{p(1-p)}\sqrt{\frac{z^2(1-f_D)}{2\, f_D N_D + z^2(1-f_D)}}, \end{aligned} \tag{17}$$

where $a = z^2(1-f_D)/(f_D N_D)$. This expression still depends on p. As for the well-known majorization for sample size calculation we may use the maximum of $p(1-p)$ at $p = 0.5$ to obtain a conservative bound. Setting $p = 0.5$ we obtain the expression

$$r_d(0.5, N_D, f_D) = \sqrt{\frac{z^2(1-f_D)}{2 f_D N_D + z^2(1-f_D)}}. \tag{18}$$

Of course the value $r_d(p, N_D, f_D)$ for a particular p may be much lower than $r_d(0.5, N_D, f_D)$. But we are interested in a simple general bound which covers all possible situations. In Section 2 we aimed at small proportions because the small groups are difficult to estimate reliably. For the difference resolution we have to take into account a small difference between any two probabilities between 0 and 1. Therefore, taking the worst case, $p = 0.5$ seems reasonable here.

Note that for equal domain sizes $\Delta_A = N_{A2} - N_{A1} = N_D\delta_A$ holds. Using $\Delta_A = N_D\delta_A$ we can also establish a bound for the absolute size Δ_A of a difference in group sizes in two domains of the same size N_D, which will always be estimated reliably by a simple random sample.

Definition: *The $100 \cdot (1-\alpha)\%$ difference resolution $R_d(1-\alpha, N_D, f_D)$ of two independent simple random samples with equal sampling rate f_D in two domains of equal size N_D is*

$$R_d(1-\alpha, N_D, f_D) = N_D\, r_d(0.5, N_D, f_D) = N_D\sqrt{\frac{z^2(1-f_D)}{2 f_D N_D + z^2(1-f_D)}} \tag{19}$$

where $z = \Phi^{-1}(1-\alpha/2)$ is a standard normal quantile.

Note the subscript d in R_d to distinguish the difference resolution from the size resolution R_s. The similarity of the size resolution R_s and the difference resolution R_d is obvious. R_d uses twice the sample size in the denominator and the square root. Again the difference resolution does not depend on the involved proportions, though the price in terms of quality of approximation now is higher than for the resolution R_s. In addition the dependence on the domain size N_D is marked for the difference resolution and will remain in the approximate difference resolution we derive next.

Given a desired difference resolution R_d the sampling fraction to achieve it is

$$f_D = \frac{z^2(1-r_d^2)}{2 r_d^2 N_D + z^2(1-r_d^2)}, \tag{20}$$

where $r_d = R_d/N_D$ is the relative difference of the size of a group A in two equal domains. The necessary sample size to reach R_d is

$$n_D = f_D N_D = \frac{z^2(1-r_d^2)}{2 r_d^2 + z^2(1-r_d^2)/N_D} \tag{21}$$

It is obvious that the direct dependence on the domain saize N_D is weak, i.e. the main driver for the necessary sample size is the proportion $r_d = R_d/N_D$. For the communication with users it is therefore advisable to discuss the difference of the proportions. Then the sample size is readily derived when assuming that the domains are of equal size and the actual proportions are symmetric around $p = 0.5$. If this assumption is not useful then one way to determine sample size is to find a root of $r_d(p, f_1, f_2, N_1, N_2)$ numerically.

Alternatively, if the domain sizes, where we plan to estimate a small difference of proportions, are in fact different and also the sampling fractions may differ then a suitable definition of the difference resolution for different domain sizes and sampling rates is:

$$R_d(1-\alpha, N_1, N_2, f_1, f_2) = \max(N_1, N_2)\sqrt{\frac{z^2(1-\min(f_1, f_2))}{2\min(f_1 N_1, f_2 N_2) + z^2(1-\min(f_1, f_2))}} \tag{22}$$

This may be a conservative bound, of course. But the bound may remain useful for planning purposes.

Further simplifications of the difference resolution are possible. If $2 f_D N_D = 2 n_D$ is much larger than $z^2(1-f_D)$ we may neglect the second summand $z^2(1-f_D)$ in the denominator of the difference resolution. If in addition we neglect the

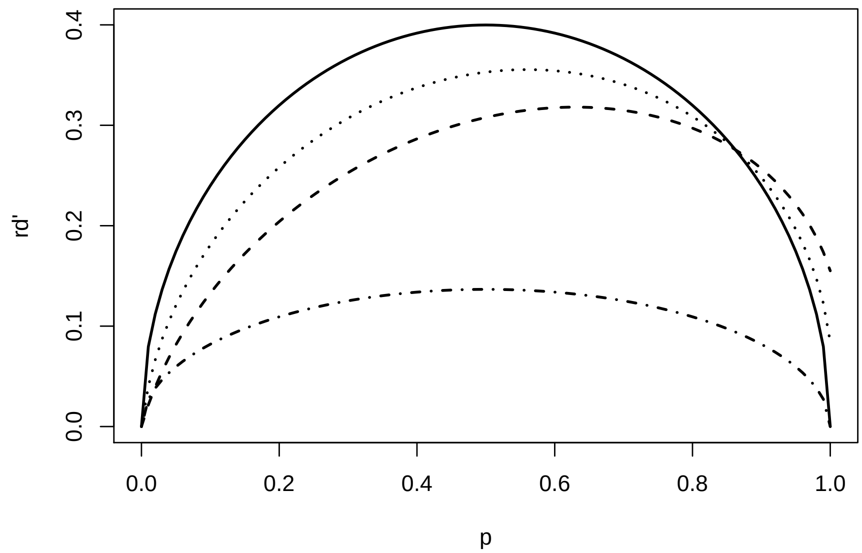

Figure 1. Function r'_d with argument $p = (p_{A1} + p_{A2})/2$ at $f_1 = f_2 = 0.01$ and $N_1 = N_2 = 1\,000$ (solid line), $N_1 = 1\,000, N_2 = 10\,000$ (dashed line), $N_1 = N_2 = 10\,000$ (dash-dotted line) and $f_1 = 0.01, f_2 = 0.02, N_1 = N_2 = 1\,000$ (dotted line).

finite sample correction $1 - f_D$ we arrive at the approximate difference resolution

$$\tilde{R}_d = N_D \sqrt{\frac{z^2}{2 f_D N_D}} = N_D \sqrt{\frac{z^2}{2 n_D}}. \tag{23}$$

To reach a desired approximate difference resolution $\tilde{R}_d$ the sample size must be

$$n_D > \frac{N_D^2\, z^2}{2\tilde{R}_d^2} = \frac{z^2}{2\tilde{r}_d^2}, \tag{24}$$

where $\tilde{r}_d = \tilde{R}_d / N_D$ is the resolution expressed as a proportion. This possible approximation could also have been directly derived from (21).

Setting in addition $z^2 = 4$ we obtain the following approximate 95% difference resolution

$$\tilde{R}_d = N_D \sqrt{2/n_D}. \tag{25}$$

This expression is a useful shortcut for quick calculations. Obsiously it leads to a sample size which is needed to achieve a specific approximate difference resolution $\tilde{R}_d$:

$$n_D = 2N_D^2 \Big/ \tilde{R}_d^2 = \frac{2}{\tilde{r}_d^2}. \tag{26}$$

Thus the simplest way of using the difference resolution is to decide the difference $\tilde{r}_d$ of proportions to be estimated and to determine $2/\tilde{r}_d^2$ as the sample size. For example, if a difference of proportions of 0.1 should be estimated, a sample size of $2/0.1^2 = 200$ is needed in both domains and we arrive at a total sample of size 400 again.

In the case of different domain sizes and sampling fractions the approximate 95%-difference resolution is

$$\tilde{R}_d(0.95, N_1, N_2, f_1, f_2) = \max(N_1, N_2) \sqrt{\frac{z^2}{2 \min(f_1 N_1, f_2 N_2)}} \tag{27}$$

Tabel 4 shows the difference resolution for the same domain sizes and sampling fractions as Table 3. The dependence on N_D is marked. Obviously in all cases $R_d > R_s$, which is clear from the formulae, too.

The difference resolution has a close relation to a Z-test for a particular difference of two means with known variance in independent samples. This is obvious from the form of the condition for deriving the difference resolution (15). In fact, a Z-test for the null-hypothesis $\delta_A = p_{A2} - p_{A1} = r_d(p)$ has a power of approximately 50% at the alternative hypothesis $\delta_A = 0$. Thus the difference resolution may be interpreted as the smallest difference at which the Z-test reaches approximately 50% power for the alternative hypothesis $\delta_A = 0$ given $(p_{A1} + p_{A2})/2 = 0.5$. If the involved proportion is different from 0.5 or if the domain sizes and sample sizes are not equal then the power would be larger. Note that contrary to this test, the test that will be applied when analysing the survey has a null hypothesis $\delta_A = 0$ and the power at $\delta_A = r_d(p)$ will be different.

4 Application of the precision resolutions

4.1 Swiss Population Survey

The approximate 95% size resolution for the Swiss Population Survey with $f = 0.0274$ is $\tilde{R}_s = 141$. This means that we can estimate such a group with a sample size of $n_D = f \cdot N_D$ for any(!) municipality or domain. Table 5

Table 4: 95% difference resolution R_d and approximate 95% difference resolution $\tilde{R}_d$.

	R_d for N_D =						$\tilde{R}_d$ for N_D =
f_D	500	1000	3000	10000	15000	100000	10000
0.0003	482	930	2476	6248	8204	24527	8002
0.0010	446	811	1874	4013	5052	13722	4383
0.0100	263	400	733	1367	1679	4357	1386
0.0274	174	253	448	823	1009	2611	838
0.0500	131	188	329	604	739	1910	620
0.1000	92	131	228	416	509	1315	439
0.5000	31	44	76	139	170	439	196

shows the size of the language groups of the municipality Veyrier in 1990 and in 2000. All language groups except Spanish are above $\tilde{R}_s$. In fact the proportion of the Spanish language group in Veyrier in 2000 was $p_A = 0.0148$ and the standard error of the sample proportion with sampling rate $f = 0.0274$ would be $\sigma(\hat{p}_A) = 0.0076$. Thus a rough 95% tolerance interval would reach down to 0.

To better judge the change of the size of language groups the sizes of 1990 have been updated proportionally to the increase of the size of the municipality (column 1990 update in Table 5). The change Δ'_A of the size of this update of 1990 compared with 2000 is relatively small. For example the german language decreased by 120. The difference resolution would be $R_d = 776$, which is larger than any of the updated changes Δ'_A of language group sizes of Veyrier. Thus from the difference resolution we must say that these differences are not estimable. Similarly a χ^2-test for the equality of proportions yields a p-value of 0.27 for the French language and even larger p-values for the rest of the languages. Thus there is no significantly different language group proportion as predicted by the difference resolution.

Assuming, that the same language distribution holds in a city five times as large as Veyrier, the χ^2-test for the change of proportions becomes significant for French but for none of the other languages. In this situation the difference resolution is $R_d = 1740$ and thus the difference resolution would suggest that the change in the size $\Delta'_A = 5 * 395 = 1975$ of the French language group would be estimable while the other changes not, completely analogue to the test result. This is similar to the situation where we can pool together five independent yearly samples to form an estimate of the change of the language group sizes for Veyrier over five years.

Table 5: Language groups of municipality Veyrier in 1990 and 2000 (Source SFSO Census 2000).

Language	1990	2000	1990 update	Δ'_A
French	5443	7271	6876	395
German	504	517	637	-120
English	386	404	488	-84
Italian	172	158	217	-59
Portuguese	150	143	189	-46
Spanish	119	132	150	-18
Other	265	267	335	-68
Total	7039	8892	8892	

4.2 European Social Survey

The Swiss Sample of the European Social Survey of 2008 (round 4 of the ESS) consists of $n = 1819$ interviews. The sample is designed as a stratified random sample of households with stratification according to regions (NUTS2) and proportional allocation. Within a household a member above age 15 is selected randomly for the interview. The design effect due to the clustering and unequal probability sampling can be expected to be larger than 1 if households are more homogenous than the general population (cf. Ganninger 2006). Note however, that the design effect due to clustering (deff_c, cf. Ganninger 2010) may vary from item to item and sufficient margins must be allowed for when determining sample size. The population size is $N = 6\,416\,728$ and thus the overall sampling rate is $f \approx 0.000283$.

The approximate 95% precision resolutions for a simple random sample with the sampling rate of the ESS are $\tilde{R}_s = 13\,552$ and $\tilde{R}_d = \sqrt{N_D}\,83$. The design effect due to unequal probability sampling (deff_p) for ESS4 of 2008 is estimated as 1.24 (European Social Survey 2010:259). In the formulae for the size resolution and the difference resolution the sample size $n_D = f_D N_D$ must be adjusted to the effective sample size (Lynn et al. 2007). For the approximate precision resolutions this adaptation amounts to a multiplication by the design effect or its square root respectively. The size resolution becomes $\tilde{R}'_s = \text{deff}\tilde{R}_s$ and the difference resolution becomes $\tilde{R}'_d = \sqrt{\text{deff}}\tilde{R}_d$. The design effect due to clustering deff_c of the ESS is not reported for 2008 since it depends on the variable considered. We use the rough approximation $\text{deff} = \text{deff}_p$ here to show the adjustments to the precision resolutions. Using as the design effect $\text{deff}_p = 1.24$ we would have $\tilde{R}'_s = 1.24 \cdot 13\,552 = 16\,804$ and $\tilde{R}'_d = \sqrt{1.24}\,\sqrt{N_D} \cdot 83 = \sqrt{N_D}\,93$.

Thus the ESS4 survey should be able to estimate any group of size 16 804 or larger on any level of disaggregation. As an example we may look at the group of persons in Switzerland which have not completed primary education and are of age 65 or more (European Social Survey 2010, Appendix 1). According to the Swiss Census in 2000 there were 12 820 men and 29 884 women of age 65 or more which have not completed primary education. According to the approximate size resolution $\tilde{R}'_s = 16\,804$ only the group of women would be estimated reliably from the ESS survey if these sizes were the same in 2008. Actually ESS4 contains no male person but 4 female persons of age 65 or more with incomplete primary school. For the male persons we thus have an unobserved group. Under the hypergeometric distribution, and assuming the numbers of 2000 hold, the probability of this event is 0.67 and therefore is no surprise. Of course, such distortions arise in practice much more due to particular non-response problems than because of sampling variation. The actual estimate of the number of women over 65 with incomplete primary school from ESS4 is 9 141 with a standard deviation of 4 687. The usual rough 95% confidence interval would include 0 and therefore the estimate may be classified as too variable in the light of the survey results. The count for 2000 of 29 884 would not be included in the confidence interval either and thus we would conclude that the number of women with incomplete primary school had dropped in the mean time.

The difference resolution is studied when estimating the gender difference in education. Data from the 2000 Census shows that there were 459 761 university academics living in Switzerland in 2000 (first column of Table 6). The difference between men and women was $\Delta_A = 128\,857$.

Would we be able to estimate the difference of 128 857 between the number of men and women university academics with the ESS? Assuming equal size of the domains of men and women of $N_D = 6\,416\,728/2$ we obtain a difference resolution of $\tilde{R}'_d = 164\,083$. Therefore we would not expect to estimate the difference of male and female university academics with the sample size of the ESS in any case. Refining the difference resolution by taking into account the proportion of university academics of $p = 0.076$, i.e. using (17), we arrive at a difference resolution of $R'_d = 86\,963$, which now is lower than 128 857 and thus indicating that it should be possible to estimate the difference reliably. A χ^2-test for the equality of proportions with the data of ESS4, taking into account the design, reveals a p-value of 0.0014 and thus a highly significant difference. Obviously the resolution at $p = 0.5$ is too conservative in that case because the involved probability is much lower.

In the Census 2000 among the persons aged 35 to 44 there were 281 271 men and 335 358 women with highest educational level Secondary II (fourth column in Table 6). Thus in 2000 the difference between men and women with highest educational level Secondary II is $\Delta_A = 54\,087$. Would we be able to estimate the difference of 54 087 between the number of men and women with Secondary II education and of age 35 to 44 based on the ESS? Assuming equal size of the domains of men and women, i.e. setting $N_D = 1\,256\,584/2$ we obtain a difference resolution (corrected for a design effect of 1.24) of $\tilde{R}'_d = 83\,548$, indicating that we cannot estimate the above difference. Note that in this case the proportions are close to 50% and therefore the difference resolution should not be too conservative. The actual estimate from the ESS is $\hat{\Delta}_A = 99\,690$, which is larger than the difference resolution and thus should be estimable. Nevertheless, the χ^2-test for the equality of proportions gives a p-value of 0.31 and thus does not indicate a significant difference. This may happen because even if the true difference is above the difference resolution the power of a test corresponding to the difference resolution is not particularly high and the fact that in the test corresponding to the resolution the usual roles of null and alternative hypothesis are exchanged also has an effect. To avoid such a situation it may be necessary to increase the level of the tolerance interval. For example for $1 - \alpha = 0.999$ we obtain an approximate difference Resolution of $\tilde{R}'_d = 122\,082$.

5 Conclusion

The precision resolutions help in the discussion of the precision of a survey at the planning stage when many small groups and small differences in proportions should be estimated. The precision resolutions set a lower limit to a group size or a difference of proportions and restrict the variability of the involved estimators. They describe whether a group size or a difference of proportions may be estimated by the planned survey.

Given a desired (approximate) size resolution $\tilde{R}_s$ we can easily calculate the size of a sample that is necessary to achieve this size resolution. In other words we may choose a limit above which a group is considered to have a relevant size and the size should be estimable. Setting the size resolution to this relevant size leads to the necessary sample size.

The difference resolution R_d is more complex since it depends on the domain size and the proportions involved. The approximation error when neglecting the proportions involved, i.e. replacing the true mean proportion by 0.5, may be large. Thus the difference resolution may be conservative. Often this is what is needed to prevent too high expectations.

When discussing the precision of a planned sample the precision resolutions help to clarify the understanding of the implications of sample size. A user may well be able to state an approximate size of a group that is interesting to him or her. The size resolution is able to lead to the sample size needed. Often the user is not interested in the size of a group alone but in a change of the size or in a comparison with another domain. He or she may still give a rough measure of size of the groups that are involved or of the difference in proportion he or she would like to be able to estimate. Specifying the difference resolution seems possible and captures the main ingredients of a hypothesis test.

The sample size corresponding to a desired difference resolution does not guarantee that a test will be significant once the data is available and the difference is actually of the size assumed. This is due to the relatively low power of 0.5 of the test corresponding to the difference resolution. Setting the z-quantile in the difference resolution to a higher value, for example $z = 2.58$, would increase the power.

The precision resolutions are not intended for quantitative variables and in surveys where quantitative variables are the key characteristics to estimate they may just help to simplify the discussion on the precision of estimators of proportions without determining the sample size. In addition,

Table 6: Persons with highest education level tertiary in Switzerland.[a]

	University Tertiary			Secondary II, age 35-44			
	Census 2000	ESS 2008	SE	Census 2000	ESS 2008	SE	$\hat{p}_{Ai}$
Men	294 309	351 648	39 253	281 271	286 987	35 214	0.504
Women	165 452	237 876	30 967	335 358	386 677	40 047	0.562
Δ_A	128 857	113 772		54 087	99 690		
Total-CH	6 043 350	6 416 728		1 195 364	1 256 584	66 618	

[a]SFSO, Census 2000 according to documentation for ESS-CH.

the requirements of different users still may diverge and the precision resolutions cannot solve the problem of the multi-dimensionality of a survey. They merely intend to help lay persons to participate in this discussion.

Some of the results on resolutions can be extended to other sampling designs with the help of the design effect (Kish 1965). Thus the sample size in the formulae for the resolution is to be understood as effective sample size. However, since design effects may vary a lot for different groups and domains even for the same design more caution is needed as soon as complex sample designs are involved.

For a sub-population or domain which is not a stratum with a fixed sample size, additional variability is induced by the hypergeometric distribution of the size of the part of the sample falling into the sub-population. The results on the resolution may be impaired if the variability of n_D is too large. The issue is even more complex for non-proportional sampling when a domain cuts across strata. Then the question on the precision is mixed with the question of the allocation of the sample and the precision resolutions may not be able to capture the situation. However, the example of the European Social Survey shows, that for mild deviations from stratified simple random sampling the precision resolutions may still be useful.

Acknowledgements

We are very grateful for the valuable comments of two anonymous referees and of the editors on drafts of the paper.

References

Bickel, P. J., & Doksum, K. A. (1977). *Statistics, Basic Ideas and Selected Topics*. Oakland: Holden-Day.

Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). New York: Wiley.

European Social Survey. (2010). *ESS4 – 2008 documentation report, edition 3.01.* Technical report, The ESS Data Archive.

Ganninger, M. (2006). *Estimation of Design Effects for ESS Round II.* Technical report, European Social Survey.

Ganninger, M. (2010). *Design effects: Model-based versus design-based approach.* GESIS-Schriftenreihe Band 3. GESIS.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Korn, E., & Graubard, B. (1998). Confidence intervals for proportions with small expected number of positive counts estimated from survey data. *Survey Methodology*, *24*(2), 193-201.

Lenth, R. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, *55*(3), 187-193.

Lynn, P., Häder, S., Gabler, S., & Laaksonen, S. (2007). Methods for achieving equivalence of samples in cross-national surveys: The European Social Survey experience. *Journal of Official Statistics*, *23*, 107-124.

Noble, R. B., Bailer, A. J., Kunkel, S. R., & Straker, J. K. (2006). Sample size requirements for studying small populations in gerontology research. *Health Services and Outcomes Research Methodology*, *6*, 59-67.

Sahai, H., & Khurshid, A. (1996). Formulae and tables for the determination of sample sizes and power in clinical trials for testing differences in proportions for the two-sample design: A review. *Statistics in Medicine*, *15*, 1-21.