

CAVEAT ON THE ERROR ANALYSIS FOR STEREOLOGICAL ESTIMATES

ZHENGWEI YANG, RENDONG ZHANG, XIAOHONG WEN, ANPEI HUANG

Morphometric Research Laboratory, North Sichuan Medical College, Nanchong, Sichuan 637007, China
(Accepted January 10, 2000)

ABSTRACT

It is frequently asked that how big a sample size, or how much measurement, is needed to achieve an accurate stereological estimate. The observed total error of a stereological estimate arises from individual difference (i.e. inter-animal / organ difference or biological variation) and intra-individual variation (or the stereological error). Statistical methods for error analysis familiar to most biological researchers are based on independent random sampling, however systematic random sampling, which is usually more efficient, is almost always performed in practice. A number of methods for error analysis were utilized in a number of model and actual studies in this paper to demonstrate from a practical point of view the pros and cons of different error analytical methods. Assumption of independence for a systematic sampling will result in overestimation of the stereological error as shown by the studies. A simple and practical approach for error analysis as recommended in this paper is to divide the systematic sample from an organ into two systematic sub-samples, regard them as two independent sub-samples and then compare the difference between the two sub-sample means.

Keywords: error, independent sampling, stereology, systematic sampling, variation.

INTRODUCTION

The prerequisite for a meaningful error analysis is that unbiased method, which normally means random sampling in stereological or morphometric practice, is used to obtain an estimate. The total error of a stereological estimate, which is usually expressed as CE (coefficient of error, equal to SEM, the standard error of the mean, divided by the mean), tells whether the overall estimate is satisfactory or not. As much effort is often made to starting an experiment and obtaining a sample for measurement, measurement with precision is often sought after. One of the most significant recent findings in stereology is considered to be the "Do more less well!" rule (Gundersen and Østerby, 1981), the fact that manual measurement with or without the help of an image system rather than automatic image analysis is often the choice in practice, or the rule of experience that the number of measurements per organ (or biological unit) does not have to be very big (e.g. no more than 100~200 point sampled intercepts or test points hitting the structure in concern to be measured or counted per organ as suggested by Gundersen *et al.*, 1988a,b) if uniform (systematic) random sampling with a proper spacing of test probes is adopted (Miles, 1987). This is amazing to potential or new users of stereology and some actual error analysis are needed for them to be

convinced. Statistical methods for error analysis familiar to most biological researchers are based on independent (simple) random sampling, however more efficient systematic sampling is almost always used in practice. A number of methods available for error analysis were applied in a number of model and actual studies in this paper to demonstrate from a practical point of view the virtues and defects of the methods.

MATERIALS AND METHODS

METHODS FOR ERROR ANALYSIS

Three methods for error analysis are mainly concerned in this paper. The error of an estimate is expressed as a percentage of the total error contributed by the intra-individual (or organ or biological unit) stereological error and the calculation was based on the ratio between squared intra- and total SEMs or CEs.

Method 1. Consider all the n_{in} measurements (e.g. intercept lengths or field measurements such as point fractions) sampled from an individual as an independent random sample, and the stereological error is calculated by

$$SEM_{in}^2 = [\Sigma(x_i - \bar{x})^2 / (n_{in} - 1)] / n_{in} \quad (i = 1, 2, \dots, n_{in}) \quad (1)$$

Method 2. Consider the intra-individual multi-stage sampling as an independent nested sampling and calculate the error according to the classical method as described by Shay (1975) and Gundersen and Østerby (1981).

Method 3. Record the systematic intra-individual measurements from an individual in order, divide them into two systematic sub-samples, regard them as two independent random sub-samples and then calculate the stereological error by

$$CE_{in}^2 = (1/2) \cdot [| \bar{x}_1 - \bar{x}_2 | / (\bar{x}_1 + \bar{x}_2)]^2 \quad (2)$$

where the \bar{x}_1 and \bar{x}_2 are the two sub-sample means. Similar method was tentatively used for the error analysis of the *fractionator* estimator by Geiser *et al* (1989).

Model Study 1 (an independent random sampling model for particle size estimation)

Suppose there is a population of ORGAN bags in each of which there are billions of particles with sizes uniformly distributed between 1 and 99999 (u). And suppose that each ORGAN consists of a great number of SECTION bags and each SECTION consists of a great number of FIELD bags in each of which there are a great number of particles. To estimate the mean particle size, nested sampling is presumed: 5 ORGANS are sampled from the population, and then 5 SECTIONS from each ORGAN, 5 FIELDS from each SECTION and 5 particles from each FIELD are sampled step by step, all in an independent random manner. That is, a total of 125 particles from each ORGAN (625 from the population) are sampled. 625 5-digit random numbers are chosen from a random number table to represent the particle sizes sampled from the 5 ORGANS in order.

Model Study 2 (a systematic random sampling model for volume fraction estimation)

Suppose there is a population of spherical cells and in the center of each cell there is one spherical nucleus. 5 cells are randomly sampled and a set of 4 systematic random parallel sections through each cell is obtained. The diameters of these sampled cells are arbitrarily presumed to be 11, 12, 13, 12 and 11 (u) with their nuclear diameters being 7, 8, 9, 8 and 7 (u), respectively. The distance between the parallel sections is exactly 1/4 of each cell's diameter and the distances from the cell end to the first section are determined to be 1.04, 2.23, 2.41, 2.89 and 0.94 (u) for the 5 cells, respectively, using a random number table.

The areas of the nuclear and cell profiles on the sections are calculated according to their diameters.

Thus, (i) a consistent estimate of the nuclear volume fraction for each of the 5 cells is estimated by dividing the total area of the cell profiles by the total area of the nuclear profiles on the 4 sections (Mayhew and Cruz-Orive, 1974). (ii) Consider the 4 sections through each cell as independent and the nuclear volume fraction for each cell is also estimated by averaging the 4 nuclear area fractions (area of nuclear profile / area of cell profile). This is an inconsistent and biased estimator for volume fraction (Mayhew and Cruz-Orive, 1974). (iii) The Matheron's transitive method for one-dimensional systematic sampling as described by Gundersen and Jensen (1987) and Cruz-Orive (1989) is also tentatively used for error analysis in this model.

Actual Study 3 (estimating the numerical density of spermatozoa in the rat epididymis)

An epididymis was removed from each of 6 normal adult male SD rats and three systematic sections orthogonal to the long axis of each organ were cut. The sections were methacrylate-embedded 25 μm -thick sections, stained with hematoxylin, and observed on a video screen with a 100 \times oil lens at a final magnification of 3286. Fields were systematically sampled with a motorized stage, the space between fields being 0.75~1.00 mm along X or Y-axis. On each field was superimposed a set of 12 regularly spaced counting frames each with area 48 μm^2 . Section was optically sectioned along Z-axis with a distance of 0.25 μm between the focusing planes (optical sections) using a computerized stage. Elongated and curved spermatozoa were counted in 10 μm of section in depth according to the optical *disector* principle (Gundersen *et al.*, 1988b), thus the number of spermatozoa per volume of epididymal fluid filled with spermatozoa was estimated. The upper left corners (i.e. test points) of the counting frames hitting the spermatozoal fluid in the epididymal tubule lumen on the first focusing plane were counted to represent the number of *disectors* used for spermatozoal counting. The average number of *disectors* and the number of spermatozoa counted per animal were 138 and 148, respectively. To analyze inter-*disector* variation, those *disectors* with the test points not hitting the spermatozoal fluid but there were spermatozoa counted in them were also included. As a result, an average of 154 data (spermatozoal numbers per "*disector*") per organ and 7 to 112 data per section were collected.

Error analysis was also performed using the equation (7) for the error analysis of numerical density estimate in the paper by Braendgaard *et al.* (1990).

Actual Study 4 (estimating the volume fraction of the inter-villus space in placenta)

5 placenta were obtained from 5 full-term Chinese women with pregnancy anemia (maternal venous hemoglobin levels 80~90 g/L). 7-8 (average 7.6) *vertical sections* (methacrylate-embedded, 5 μm -thick, stained with hematoxylin and eosin) orthogonal to the fetal side of the placenta and of similar sizes were cut from each placenta (Baddeley *et al.*, 1986). Sections were observed on a video screen at a final magnification of 631 and fields were systematically sampled with a motorized stage, the space between fields being 800 μm along X or Y-axis. A test system with 20 regularly spaced test points was superimposed on each field and test points hitting the inter-villus space and the whole section were counted to estimate the volume fraction of the inter-villus space in placenta. 9~20 fields (average 17.6) were measured per section.

Assuming a binomial distribution of the test points in space (i.e. regarding the test points independent), the intra-organ error was also evaluated according to the equation (3.26) in the book by Weibel (1979):

$$CE_{in}^2 = (1/P) - (1/P_0) \quad (3)$$

where P and P₀ were the total numbers of test points hitting the inter-villus space and the placental section, respectively.

Actual Study 5 (estimating intercept lengths in the placental membrane)

The same materials described above were used in this study. A straight test line with length 200 μm was also superimposed on each field. The test line was rotated after each field was measured so that the directions of test lines were isotropic in distribution in placenta (sine-weighted on the *vertical sections* according to Baddeley *et al.*, 1986). Intercept lengths were measured along the direction of the test line, from the intersection between the test line and the boundary of the capillary vessels in the terminal villus to the nearest boundary of the terminal villus, to estimate the

mean thickness of the placental membrane. Those intercepts without completely inside the placental membrane (e.g. those crossing the other side of the capillary vessel or another vessel) were not measured. 2~65 (average 22) intercepts were measured per section.

RESULTS

The main results of the error analysis are shown in Table 1.

In model study 1, the true (theoretical) inter-ORGAN mean particle size should be 50000 (u) and was estimated to be 47734 (u). The true inter-ORGAN variation should be 0, i.e. the observed total error would be all contributed by the intra-ORGAN sampling. The error contribution of intra-ORGAN sampling as estimated by Method 2 appeared to be more consistent with the true value (100%) than by Method 3 in this model (Table 1).

In model study 2, from the true volume fractions of the 5 cells as calculated from the nuclear and cell (3D) diameters, the mean nuclear volume fraction estimate for the cell population is 28.8% with a CE of 4.85% which is contributed by inter-cell variation. From the consistent estimates of the 5 cells' volume fractions, the mean volume fraction estimate is 28.88% with a total CE of 6.34% which is contributed by both inter- and intra-cell variations. The intra-cell variation would therefore account for 41% [(0.06342 - 0.04852) / 0.06342] of the total error. When the 4 systematic sections through each cell were regarded as independent (i.e. an estimate was calculated from each section), the mean volume fraction estimate was 21.41% (CE 6.88%).

In actual study 3, the spermatozoal number per unit volume (480 μm^3) of spermatozoal fluid in epididymis was estimated to be 1.16 (the total spermatozoal number counted per organ, divided by the total volume of the optical *disectors* used for the counting), with a total CE of 10.38%. When the number of *disectors* was specially handled as described for estimating the inter-*disector* variation, the spermatozoal number per "*disector*" was 0.97 (CE 6.49%).

Table 1. *Results of error analysis.*

| | Total error (CE, %) | Error (% of total error) contributed by intra-individual sampling | | | |
|----------------|------------------------|---|----------|----------|------------------------------------|
| | | Method 1 | Method 2 | Method 3 | Additional |
| Model study 1 | 1.8 | 176 | 128 | 46 | 100 ^a |
| Model study 2 | 6.3 | 985 | 985 | 99 | 41 ^b , 179 ^c |
| Actual study 3 | 10.4 | 31 | 30 | 8 | 53 ^d |
| Actual study 4 | 5.6 | 27 | 169 | 3 | 5 ^e |
| Actual study 5 | 6.6 | 22 | 68 | 0.3 | - |

a: the true (theoretical) value; b: according to the true and consistently estimated volume fractions (see the second paragraph on next page); c: according to the Matheron's transitive method as described by Gundersen and Jensen (1987) and Cruz-Orive (1989); d: according to Braendgaard *et al* (1990); e: according to Eq. 3.

DISCUSSION

It has been well recognized that systematic sampling is usually more efficient than independent sampling and the stereological error will be overestimated when a systematic sample is treated as an independent one (Gundersen and Jensen, 1987; Cruz-Orive, 1989; Mattfeldt, 1989). Methods 1 to 3 based on assumption of independent sampling will, therefore, tend to overestimate the stereological error when used in a study with a systematic sampling scheme. However the magnitude of the overestimation by Method 3 would not be as large as by Methods 1 and 2 as the two sub-samples used in Method 3 are still systematic. Consistent results were obtained in the studies of this paper: (i) the errors estimated by Methods 1 and 2 were about 4 to 200 times larger than that by Method 3 in studies 2~5, and (ii) bias of the error estimate by Methods 1 and 2 was relatively comparable with Method 3 in the model study 1 where the intra-individual sampling was indeed independent random (Table 1).

Regarding the classical Method 2 for nested sampling, the sample size at each sampling level should be constant, or the same average sample size should be used in calculation, otherwise inconsistent results would be obtained: the total intra-individual error would be quite different when different levels of intra-sampling are concerned. But such a constant sample size at each sampling level may not be guaranteed in a systematic sampling practice (see the actual studies 3~5).

As a matter of fact, if the sample size is arbitrarily made to be constant, e.g. the same number of fields are always sampled from each systematic section no matter how big the section is, the stereological estimate is not unbiased. The importance of systematic random sampling, or uniform random sampling, can never be over emphasized to obtain an unbiased estimate (Cruz-Orive and Weibel, 1981; Gundersen, 1991).

Treating a systematic sample as independent for error analysis by Methods 1 and 2 may be an awkward procedure as well. In actual study 3, the inter-*disector* variation was hard to be evaluated because not all *disectors* would be completely inside the measuring space: the spermatozoal fluid in epididymis. And it may also induce bias for the stereological estimate. In model study 2, for example, calculating an "independent" volume fraction estimate from each section through the cell resulted in an underestimation of ~26%.

In summary, systematic sampling rather than independent sampling is often used in practice and unbiased error estimator for systematic sampling is not

available. Assuming independence for a systematic sample will overestimate the stereological error. Divide a systematic sample into two systematic sub-samples and evaluate error by comparing the two sub-sample means (see Eq. 2). This appeared to be a reasonably simple and practical good method for error analysis.

A preliminary report of some of the data (Yang *et al.*, 1999) has been presented at the Xth International Congress for Stereology, Melbourne, Australia, 1-4 November 1999.

REFERENCES

- Baddeley AJ, Gundersen HJG, Cruz-Orive LM (1986). Estimation of surface area from vertical sections. *J Microsc* 142:259-76.
- Braendgaard H, Evans SM, Howard CV, Gundersen HJG (1990). The total number of neurons in the human neocortex unbiasedly estimated using optical disectors. *J Microsc* 157:285-304.
- Cruz-Orive LM (1989). On the precision of systematic sampling: a review of Matheron's transitive methods. *J Microsc* 315-33.
- Cruz-Orive LM, Weibel ER (1981). Sampling designs for stereology. *J Microsc* 122:235-57.
- Geiser M, Cruz-Orive LM, Hof VI, Gehr P (1989). Counting particles retained in the conducting airways of Hamster lungs with the fractionator. *Acta Stereol* 8:419-24.
- Gundersen HJG (1991). New stereology. Abstract of the 8th International Congress for Stereology in California, USA, 71.
- Gundersen HJG, Bendtsen TF, Korbo L, Marcussen N, Møller A, Nielsen K, Nyengaard JR, Pakkenberg B, Sørensen FB, Vesterby A, West MJ (1988a). Some new, simple and efficient stereological methods and their use in pathological research and diagnosis. *APMIS* 96:379-94.
- Gundersen HJG, Bagger P, Bendtsen TF, Evans SM, Korbo L, Marcussen N, Møller A, Nielsen K, Nyengaard JR, Pakkenberg B, Sørensen FB, Vesterby A, West MJ (1988b). The new stereological tools: disector, fractionator, nucleator and point sampled intercepts and their use in pathological research and diagnosis. *APMIS* 96:857-81.
- Gundersen HJG, Jensen EB (1987). The efficiency of systematic sampling in stereology and its prediction. *J Microsc* 147:229-63.
- Gundersen HJG, Østerby B (1981). Optimizing sampling efficiency of stereological studies in biology: or "Do more less well!". *J Microsc* 121:65-73.
- Mattfeldt T (1989). The accuracy of one-dimensional systematic sampling. *J Microsc* 153:301-13.
- Mayhew TM, Cruz-Orive LM (1974). Caveat on the use of

the Delesse principle of the areal analysis for estimating component volume densities. *J Microsc* 102:195-207.

Miles RE (1987). Preface. *Acta Stereol* 6/II (ISS Commemorative-Memorial Volume), 5-10.

Shay J (1975). Economy of effort in electron microscope morphometry. *Am J Pathol* 81:503-12.

Weibel ER (1979). *Stereological Methods. Vol. Practical Methods for Biological Morphometry*. London: Academic Press, 97.