Acta Histochemica

Bruce Wayne
Department of Computer Science
Clemson University Clemson, SC, USA
one@cs.clemson.edu

Kent Clarke
University of Someotherplace
two@someotherplace.edu

August 24, 2014

Reflections on sampling, extrapolation and interpretation

Abstract

Reflections on sampling, extrapolation and interpretation Poorly designed experimental studies are a major cause of rejection of manuscripts. If the sampling is poor or faulty, then this will lead to wrong interpretations and poor science. A recurrent problem is that many researchers do not understand the limita- tions of their study owing to the poor sampling and design of the research. Were there sufficient animals at each time interval and were appropriate controls performed? Is the study reproducible? Is there a gender issue? Were the relevant data collected and are they meaningful? Are the data suitable for rigorous statistical analysis? Are the samples taken from an identical area in each animal and are these typical or representative of the whole tissue? Does the orientation of the tissue 1matter? In practice many experimental studies involve very limited sampling, which inevitably leads to problems in analysis and interpretation. The problem of sampling becomes critical in histological and histopathological studies in view of the fact that tissue sections are essentially 2-dimensional and we need to interpret the data in a 3-dimensional format. The best advice is to plan prospective experimental studies so that the samples taken will yield data that are useful, or in other words in which the results can be quantified and subject to meaningful statistical analysis. It is recommended good laboratory practice to consult a statistician early in the planning stages and to design your experiments based on the advice given. It is no longer sufficient to provide subjective or representative images (pretty pictures), but these nowadays require quantification and if possible the morphological data should be backed up with correlated studies in biochemistry and molecular biology. The second major problem in histopathological studies lies in the interpretation of the data and understanding the limitations of the study.

1 Introduction

In practice many studies involve examination of relatively small samples (both in size and number), which are then extrapolated to the whole tissue, organ or animal. This is especially true with regard to transmission electron microscopy. 1 If the sample, for example, is from the liver, then you should ensure that all the samples come from the same lobe and from the same area of the lobules. The hepatocytes at the periphery of the classical lobules differ in structure and functionality from those nearer the central vein. A similar consideration can be applied to the kidney medulla, which superficially [7] seems morphologically fairly homogenous, but in practice exhibits regional functional differences. Despite this, many experimental studies consider the liver or kidney as homogeneous organs. This after all is the http://dx.doi.org/10.1016/j.acthis.2014.04.008 0065-1281/ 2014 Elsevier GmbH. All rights reserved. approach of much biochemistry, where the tissue is initially homogenized before analysis, or in blood studies, especially with small experimental animals such as mice where necessity dictates that the blood is pooled from several animals in order to provide enough for analysis. Histology and physiology textbooks provide estimated numer- ical values for highly complex biological systems. In some cases, these are measureable finite values. For example we can measure heart beats per minute and we know how many minutes there are in a day and how many days in a year. On this basis we can estimate with a certain degree of accuracy the absolute number of heart [8] beats between birth and old age. Though such an enormous figure would be fairly meaningless to students who find it difficult to understand or imagine very large numbers unless some form of visual analogy is supplied. [1], accountants or mathematicians. Most of the numerical examples in microanatomy or physiology have been derived from examination of relatively [5] small samples and then the data are extrapolated to include the whole organ or organ system over a specific time period. The question is how accurate or relevant are such estimations.

2 Understanding

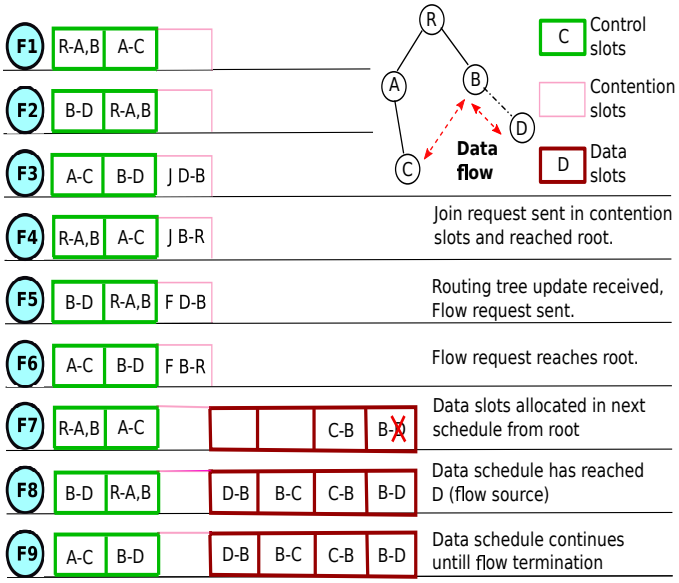

Figure 1: sampling graph

Histology textbooks indicate that there are about 300 million alveoli in the lungs providing a total surface area for gas exchange of about 80 square meters or more. Although the figure of 300 million is hard to comprehend, when expressed in square meters this can be visualized as the area of a small apartment or a tennis court. One wonders if these figures, passed

[1] In fact most individuals have diffi- culties contemplating figures above one million unless they are bankers

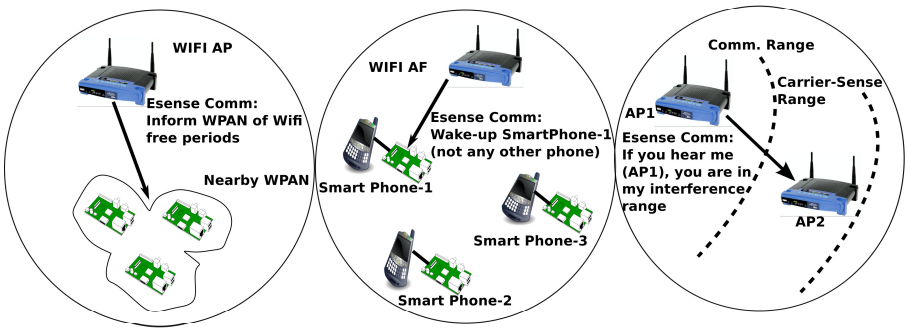

Figure 2: sampling graph

on for generations, are really accurate or miss the mark by several orders of magnitude. How was the study done? Are there gender differences? It is doubtful if these figures are similar for age-matched males and females or for the young adult and the elderly. We often cannot find the reference for the original study, but repeat the numerical values and analogies for generations of students and their textbooks. The total number of endothelial cells (lining blood vessels) in the adult is estimated at 6 10 23 and they cover an enormous surface area estimated between 200 and 1000 square meters and with a total weight of about 1.5 kg (more than the largest internal998 Editorial / Acta Histochemica [4] 116 (2014) 997999 Start with a block of ssu e or organ (liver?) approx. 5 cm 3 (=5,000 mm 3). Cut into cubes no larger than 1mm 3 (i. e. less than 0.0002 [1/5000] of original bloc k). Here sampling interval n is defined as,

$$n = \frac{{\zeta^2}_{\alpha/2} * \pi * (1 - \lg \phi)}{\psi^2 + \sqrt{\sin \beta^2 + \cos \theta^2}}$$

Mount block and trim to fa ce of ap prox. 0.5 x 0.2 mm (500 m x 200 m)sec-intro [1 mm = 1000 m]. Cut secons on ultramicrotome about 50 nm [1 m = 1,000 nm]. A cell 25 m thic k will cut into 500 secon s for EM viewing . Volume of secon = 500 m x 200 m x 50 nm = 5,000 m 3 (1mm 3 = 1,000 x 1,000 x 1000 m 3 = 10 9 m 3). As original organ was about 5000 mm 3 and secon volum e is 5,000 m 3 , therefo re each secon conta ins on ly about 1/1,000,000 of the original ssue. $\forall \rho, \exists \mu, \sum_{\epsilon=1}^{N} (\rho_\epsilon - \mu)^2 = \sum_{\epsilon=1}^{N} ({\rho_\epsilon}^2 - 2\overline{\rho_\epsilon}\mu + \mu^2)$ At lowes t useabl e magnificaon in th e EM (x 1,50 0) th e electron micros cope field of view is less than 50 m x 50 m (i.e. on ly 1/40 area of a block face 500 m x 200 m.). The final electron 2 microscope image ca nnot repre sent more than 2 x 10 -11 of original spec imen of 5 cm 3 (1/40 x 10 -11). As most ss ues show great heterogeneity of cells , sampling and int erpr eta on of EM pic tures is very problemac and the EM pi ctures in most sub mission s have limited valu e owing to this samp ling pr oblem.

It is recommended good laboratory practice to consult a statistician early in the planning stages and to design your experiments based on the advice given. It is no longer sufficient to provide subjective or representative images (pretty pictures), but these nowadays require quantification and if possible the morphological data should be backed up with correlated studies in biochemistry and molecular biology. The second major problem in histopathological studies lies in the interpretation of the data and understanding the limitations of the study. Fig. 1. Problems of sampling in transmission electron microscopy. organ, the liver) representing an estimated 100 million kilometers of blood capil-

No.	Instrument	sampling depth	start gmt time	end gmt time
1	RCM7(C1S)	36.5	22:40/06/03	22:40/07/03
2	RCM4(C1B)	70.5	22:40/06/03	22:40/07/03
3	RCM7(C2S)	23.0	16:20/06/03	16:20/07/03
4	RCM4(C2B)	44.5	16:20/06/03	16:20/07/03

Table 1: Sampling mesurements. *Instrument types and sampling depths.*

laries (I do not know if this should be Imperial millions or USA millions), though I can possibly imagine this in terms of several adjacent tennis courts. It is recommended good laboratory practice to consult a statistician early in the planning stages and to design your experiments based on the advice given. It is no longer sufficient to provide subjective or representative images (pretty pictures), but these nowadays require quantification and if possible the morphological data should be backed up with correlated studies in biochemistry and molecular biology.It is recommended good laboratory practice to consult a statistician early in the planning stages and to design your experiments based on the advice given. It is no longer sufficient to provide subjective or representative images (pretty pictures), but these nowadays require quantification and if possible the morphological data should be backed up with correlated studies in biochemistry and molecular biology. Sampling rate is given as,

$$\delta = \frac{[\sqrt{(\epsilon_1 - \epsilon_2)} - \sqrt{(\eta_1 - \eta_2)}]}{\overline{S}_{\iota_1 - \iota_2}}$$

Should we even bother to present such data in our histology or anatomy lectures or leave the figures for trivia quizzes? Fig. 1. Problems of sampling in transmission electron microscopy.

3 Related work

1The number of nephrons per kidney is estimated at about one million. Each nephron if unraveled is about 50 mm long and if all the nephrons were attached to each other it is estimated they would stretch about 100 km. Again it is hard to imagine one million, but I think most people can visualize a distance of 100 km. The kidney is regarded rightly as an incredible organ filtering waste metabolic products throughout our life. It is relatively easy to visualize the 2 l or so of urine produced per day, but much harder to visualize the 180 l of ultrafiltrate that pass the glomerular filter every day before being largely reabsorbed.[2] There are about 5.5 million erythrocytes in every microliter of blood of an adult male. The total blood volume in an adult male is about 5 l. The lifespan of an erythrocyte is about 120 days. In other words, the numbers of new erythrocytes produced in the bone marrow every day is absolutely enormous and beyond human comprehension or experience. However, we determine and quantify accurately the hematocrit and a wide range of other blood components on a daily basis from a small blood sample taken in the clinic. For centuries we have been able to examine and quantify the various formed elements in the lab or office using a hemocytometer.

- In the adult human brain the total number of neocortical neurons is estimated to be

[2](about the volume of water in a standard domestic bath).

about 19.3 billion (10 9) in females and 22.8 billion in males.

- Some more recent studies have increased these estimates up to 35 billion in normal humans.
  - Some studies that include both the neocortex and cerebellum estimate a total number of about 125 billion neurons.
- Each neuron in the neocortex has many synapses (up to 200,000) and it is estimated that the total number of synapses in the human neocortex is a mind boggling 0.15 quadrillion (10 15).

Wow! (I do not think it makes much dif- ference if you use Imperial billions (million million) or US billions (100,000 million). These figures were reached by extrapolation from small [6] brain samples from a very limited number of specimens from selected areas of the brain. Should I be 2 worried in my old age if I lose tens of thousands [7] or more synapses per day (joking of course)? It is estimated that the total length of the myelinated nerve fibers in a young adult is about 180,000 km (about half the distance to the moon). This is quite useless information, but fascinating to consider. It just emphasizes the complexity of the nervous system and how humble we should be in our limited understanding of the brain and its functioning. 4 Spermatozoa are produced in the testes in the spermatogenic epithelium of the seminiferous tubules. An individual seminiferous tubule, if unraveled, is some 3050 cm long. It has been calculated that each testis has about 250 m of seminiferous tubules busy producing the male gametes. In a healthy adult male, the testes produce on a daily basis several million spermatozoa for several decades. These typically take about 70 days to reach maturity and appear in seminal ejaculate. A few ml of semen taken for fertility analysis will contain many millions of the gametes, and a thin smear spread on a microscope slide confirms the enormous number, yet the average male will in a lifetime father only 24 children for which a single sperm is needed in each case for fertilization. All this productivity and redundancy of sperm is a biological necessity to provide the genetic variability for the gene pool, but the figures are really mind boggling. The female fetus has about 23 million primordial follicles in the ovary. Each follicle has the potential to develop after puberty, however, atresia begins well before birth and after puberty typi- cally only a single egg [9] will reach maturity (Graafian follicle) and be ovulated each menstrual cycle. In practice this means a total of about 400450 ovulated eggs until menopause. Again, this redundancy is associated with the need for genetic variability. These are just some examples of mega-numbers derived from extrapolation from small samples. What about the total number of cells in the body. Why should we even try and assess such data? It has been estimated that the adult body has some 50 trillion cells (with a margin of error of many orders of magnitude). My cell biology colleagues inform me that each cell nucleus has densely and intricately packed DNA extending for 2 m. In other words I have about 1 100 billion km DNA in my body! Wow, I find this impossible to visualize, almost like estimating the grains of sand on the beach or of stars visible in the night sky. We are amazed to learn the estimated enormous number of microorganisms living in harmony on each square mm of our skin or the number of dead keratinocytes sloughed off our skin every day and their concentrations in the air we breathe.

Sample no.	Hct0	Hctt	V(ml)	P(ml)
1	40	20	7.5	+7.5
2	40	30	10	+5.0
3	40	50	18	-3.0
4	40	60	20	-5.0

Table 2: Sampling mesurements. *Determination of plasma levels to be added to achieve target Hematocrit levels.*

4 conclusion

1. In fact the best extrapolations for many of these studies were performed using stereological [1] techniques on serial sections.

2. We need to accept these figures at face value. I find it hard to compre- hend why anybody would undertake [3]such studies which demand an enormous amount of time and tedious work and involveEditorial / Acta Histochemica 116 (2014) 997999 analysis of relatively few individuals.

3. Although it is impossible to comprehend enormous figures, such studies do demonstrate the

   (a) biocomplexity and

   (b) the scientific wonders of the human body and

   (c) how we have evolved and function

   .

4. Perhaps you can understand the dilemma of the Editor of this journal when he receives manuscripts with poor sampling or 2misuse of statistics on limited samples or pretty micrographs chosen as typical or representative.

There- fore, considering all the above, when authors intend to measure a few features in sections, a minimum of requirements seem mandatory to assure meaningful analyses. [2]

References

[1] S. H. Batterywala and H. Narayanan. Efficient dc analysis of rvj circuits for moment and derivative commutations of interconnect networks. *12th International Conference on VLSI Design*, pages 169–174, 1999.

[2] Kevin Dale, Kalyan Sunkavalli, Micah K. Johnson, Daniel Vlasic, Wojciech Matusik, and Hanspeter Pfister. Video face replacement. In *Proceedings of the 2011 SIGGRAPH Asia Conference*, volume 30, pages 130:1–130:10, 2011.

[3] Norbert Frohlich, Bernhard M. Riess, Utz A. wever, and Qinghau Zheng. A new approach for parallel simulation of vlsi circuits on a transistor level. *IEEE Transaction on Circuits and system -1:Fundamental Theory and Applications*, 45(6):601–613, 1998.

[4] Al Gist, Adam Beguelin, Jack Dongara, Weicheng Jiang, Robert Manchek, and Vaidy Sunderam. *PVM 3 User Guide and Reference Manual*. Oak Ridge National Laboratory, Engineering Physics and Mathematics Division, Mathematical Science Section, Oak Ridge, Tennessee, 378331, USA, September 1991. operated by Martin Marietta Energy Systems.

[5] Roger Grosse, Micah K. Johnson, Edward H. Adelson, and William T. Freeman. Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In *International Conference on Computer Vision*, pages 2335–2342, 2009.

[6] Micah K. Johnson and Edward H. Adelson. Shape estimation in natural illumination. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2553–2560, 2011.

[7] data sampling in multidimentional NMR. http://www.ncbi.nlm.nih.gov/pubmed/21773916.

[8] Pvm http site. http://www.epm.ornl.gov/pvm/.

[9] Sampling techniques. http://www.epm.ornl.gov/pvm/.