ELSEVIER

Journal of Neuroscience Methods 91 (2001) 91–99

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

Computer experiments to determine whether over- or under-counting necessarily affects the determination of difference in cell number between experimental groups

Barbara Heller [a], Frank Schweingruber [b], Dilek Guvenc [c], Alfred Heller [d,*]

[a] *Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL, USA*
[b] *Evanston, IL, USA*
[c] *Department of Mathematics, Bilkent University, Ankara, Turkey*
[d] *Department of Neurobiology, Pharmacology and Physiology, University of Chicago, 947 East 58th Street, Abbott Hall, Chicago, IL 60637, USA*

Received 21 July 2000; received in revised form 11 January 2001; accepted 11 January 2001

Abstract

Computer cell counting experiments were performed in order to examine the consequences of over- or under-counting. The three-dimensional reaggregate culture laboratory environment for cell counting was used as a model for computer simulation. The laboratory environment for aggregate and cell sizes, numbers and spatial placement in gelatin blocks was mimicked in the computer setup. However, in the computer, cell counting was set to be either ideally unbiased, or deliberately biased in regard to over- or under-counting so as to compare eventual results when using the various cell counting methods. It was found that there was no effect of the cell counting methods used in determining whether there was a significant difference in cell number between two experimental groups. In addition, it was found that under the conditions of these simulations, the optical disector method behaved similarly, on the average, as the ideal method of counting cell centers and in both of those cases, the average ratio between actual cell number in a flask and estimated number was close to 1.00.

Keywords: Cell counting; Computer simulation; Cell culture; Dopaminergic neurons; Reaggregates; Dopamine; Optical disector

1. Introduction

While computer simulations have been used for assessing bias and accuracy for various cell counting methods, it remains an open question as to whether over- or under-counting affects the expected frequency of observed differences between experimental groups at a given significance level. As is well known, estimates of cell number from sampled sections can be biased in the sense that, even on the average, they consistently overestimate or possibly underestimate the actual number of cells in the structure depending upon the counting method being used. It has been a matter of controversy as to whether methods that do not compensate for over- or under-counting must necessarily lead to the observation of spurious differences between experimental groups (Coggeshall and Lekan, 1996; Saper, 1997; Guillery and Herrup, 1997). In this paper, computer experiments are performed which mimic a certain experimental environment for counting dopaminergic neurons grown in cell culture in the laboratory. In the computer experiments, the laboratory methods for sampling and slicing of sections are copied. However, in the actual computer cell counting, over- and/or under-counting is artificially produced and, in fact exaggerated, so that even more severe conditions of miscounting are created than those expected in the laboratory. Under those severe counting conditions, repeated computer experiments reveal the frequency with which an experimental group shows a significant difference from a control group when in fact, there is no difference in actual cell number and also in those cases where there is a true difference between them. In other words, how often one could expect to obtain the

* Corresponding author. Tel.: +1-302-7023513; fax: +1-312-7023774.
E-mail address: effe@midway.uchicago.edu (A. Heller).

0165-0270/01/$ - see front matter
PII: S0165-0270(01)00333-8

correct comparison between the two groups even though bias is introduced into the cell counting.

Computer simulation has been used for investigating variability of the fractionator method and for cell number estimation in vivo under a variety of stochastic models of spatial placement of cells (Glaser and Wilson, 1998; Schmitz, 1998; Schmitz et al., 1999). Computer simulation has also been used to investigate spatial arrangement of cells for purposes of devising cell counting techniques for tissue culture wells (McShane and Palmatier, 1994; McShane et al., 1997).

In this paper, computer experiments are performed in order to examine the consequences of over- and under-counting of cells by comparing with the ideal method of counting cell centers and with the optical disector method. Since cell counting must take place in a specific laboratory environment, the computer experiments here are modeled on the counting of dopaminergic neurons in three-dimensional reaggregate cultures (Won et al., 1989; Heller et al., 1992, 1993, 1997). For modeling purposes, this culture system has the advantage of being more easily mimicked than in vivo systems in which it is difficult to model the spatial arrangement of cells realistically by use of spatial stochastic processes (Schmitz, 1998; Schmitz et al., 1999). But, on the other hand this culture system environment is, in fact, three-dimensional as opposed to the essentially two-dimensional situation in monolayer cultures.

Modeling a specific laboratory environment is essential in creating computer experiments if they are to have any meaning. Unavoidably then, the conclusions do not apply in all generality. It is not being claimed here that, in all cases, bias in cell counting does or does not affect the eventual conclusion as to whether two groups have significantly different number of cells. Here is an example in which over- and/or under-counting is artificially introduced in order to determine whether or not it makes a difference in the eventual determination of differences between control and experimental groups. Alternatively one could only assume in advance of investigation of the specifics of an experimental situation that bias in cell counting necessarily affects experimental results.

Three modeling situations are investigated. Those cases in which no difference in mean cell number between the two groups is present in the computer model, those cases in which the experimental group is 20% less than the control groups and those in which the experimental group is 30% less than the control group. In each case, the computer experiment is repeated 50 times and the frequency with which the experimental group is significantly different than the control group is recorded. In some of the computer experiments there is artificially produced over-counting, in some under-counting. In each case, the results are compared with the ideal situation in which cell centers are counted with no bias or where the optical disector method of counting is used.

2. Materials and methods

2.1. The laboratory environment

When dissociated fetal cells are suspended in rotatory culture, the cells collide within the vortex of the culture medium formed by the rotatory motion of the fluid and form clusters (aggregates). Within hours after the formation of the aggregates, the cells within them form clusters of 'like' cells by migration. The cells subsequently put out axons which form selective specific synaptic connections with appropriate target cells resulting in the construction of projection pathways analogous to those observed in the intact brain (Won et al., 1989; Heller et al., 1992, 1993, 1997). Experiments involving this 'three-dimensional reaggregate' system typically yield estimates of cell number or cell density for the various experimental groups. Statistical tests are performed to compare the groups. We consider the case of aggregates prepared from cells of the mesencephalon and striatum, which results in reconstruction of the dopaminergic nigrostriatal projection. Aggregates from a given flask are embedded in gelatin, cut into 50 μm sections, and the dopaminergic neurons visualized by immunocytochemistry utilizing antibodies against tyrosine hydroxylase (TH). The total number of dopamine neurons in a flask is estimated by counting TH-positive cells in a random sample of 30 aggregate sections from each flask. Typically, each flask produces about 1500–2500 aggregate sections. TH cells are counted if they meet two or more of the following criteria, (1) the cell has a diameter of 10–20 μm; (2) the cell exhibits neuronal processes; or (3) the cell nucleus can be visualized.

By using computer-assisted image analysis, the volumes of the sampled aggregate sections are estimated and the total aggregate volume for the flask is measured (Vidal et al., 1995; Heller et al., 1997). The sum of the cell counts for the 30 sampled aggregate sections from a flask divided by the sum of the corresponding section volumes provides an estimate of cell density for the flask. Total aggregate volume for the flask multiplied by cell density gives an estimate of the total number of TH-positive cells in the flask. Typically, in an experiment, five to seven replicate flasks are prepared for each experimental group.

To ensure that a representative sample of aggregates is obtained for cell counting, after 15 days in culture, aggregates from all flasks are pooled and subsequently redistributed into an appropriate number of experimental flasks. There are two modes of pooling and redistribution corresponding to the experimental situation.

One is that treatment is applied to the experimental groups after the pooling and redistribution operation has been performed (Mode 1, 'combined' pooling). The other situation is when the experimental groups are separate from the start, as for example when the different groups originate from different types of animals. Then the pooling and redistribution takes place within groups, not across groups as above (Mode 2, 'separate' pooling). For Mode 1 pooling, cell densities between groups are compared. For Mode 2 pooling, total numbers of cells per flask are compared between groups.

2.2. The computer program

To provide a convenient user interface, the program is constructed as a plug-in running in Adobe Acrobat under Windows 95/98. The program is written in 'C' with all run parameters provided by six text 'setup' files. These six files control, (a) building a single flask; (b) multiple slice runs of a previously built flask; (c) running repeated experiments with various test and control flasks; (d) flask slicing parameters for control flasks and test flasks of repeated experiments; (e) flask slicing parameters for repeated experiments. Random numbers for the various internal functions are generated from the 'rand' function provided in the Microsoft development environment. The seed for the random number generator is specified in one of the setup files. The number of calls to the random number generator is small enough that no problems are expected from periodicity. 'Result' text files are generated by the program, which provide some of the statistical summaries for the various modes of operation. These result files are also formatted to allow their input to further statistical analysis programs.

2.3. Building groups of 'flasks' for the computer simulation

In the computer simulation, an experiment consists of two groups, a 'control' group and a 'test' group. For each of these groups, the operator specifies the number of replicate flasks, the number of aggregates and the total aggregate volumes for each flask. Since the data from laboratory experiments includes those numbers, computer experiments are set up so that number of replicate flasks per group, number of aggregates in each flask and total aggregate volume for each flask are exactly those observed in an actual laboratory experiment. Each flask is represented in the computer by a rectangular block ('gelatin block') with dimensions specified in a setup file.

Individual aggregates are modeled as ellipsoids with randomly determined diameter sizes in which the mean and S.D. of each of the three diameters are specified in a setup file. The mean diameter sizes are set so that the volumes of the subsequent aggregate sections and the total aggregate volumes per flask match the laboratory experiment as closely as possible. The actual aggregates produced in the laboratory look almost spherical under the microscope (potato shaped) and are typically modeled with mean diameters ranging from 300 to 450 μm depending upon the particular laboratory experiment which is being copied. Due to the random element in the three diameter sizes, computer aggregates vary in shape in a 'potato-like' fashion similarly as do the laboratory aggregates. The aggregate ellipsoids are placed in random location and with random three-dimensional orientations within their respective 'gelatin' blocks. Aggregates are placed so that none of them overlap with each other in three-dimensional space.

Cells within the aggregates are modeled as spheres of random diameter for which the mean and S.D. are specified. Typically, the cells of both groups are given a mean diameter of 15 μm similarly as in the laboratory. For this study, the mean diameter of cells is given the same value in the setup file for each flask and consequently does not differ between flasks in a group and between groups. Cells are placed within aggregates in a spatially clustered manner. In laboratory aggregates, 'like' cells sort out and the position of dopamine cells within aggregates in three-dimensions is not random, the cells being located in clusters (Hemmendinger et al., 1981). Random clusters are formed according to a 'Poisson spatial random process' (Diggle, 1983). In this type of clustering, a certain specified number of 'parents' are assigned points within the aggregate at random spatial positions. Then each parent is assigned a random number or 'offspring' according to a Poisson distribution. The positions of the offspring relative to their parents are assigned according to a trivariate normal distribution with specified S.D. The amount of clustering is adjusted by the operator by specifying more or fewer parents (fewer parents, more clustering) and higher or lower S.D. for the trivariate normal distribution (lower S.D., more clustering). Cells that overlap other cells are rejected so that no cells overlap each other in three-dimensional space.

The number of cells in each aggregate is random according to a Poisson process, depending upon the volume (known to the computer) of the particular aggregate for that cell. The mean for the Poisson process is specified by the operator in a setup file so that it can be set to be the same for the control group as the test group or different between the two groups. Note that the actual number of cells per flask in each group is *random*. It is the *mean* of the probability distribution, which is set by the operator.

The laboratory process of pooling the aggregates between the flasks of a group (Mode 2) and then redistributing into a second set of flasks, or pooling and redistributing aggregates from all of the flasks and then

applying a treatment to the test group (Mode 1) is copied in the computer. In the case of Mode 1, the mean number of cells per (pre-pooled) flask is set to be the same for both groups. Then a survival factor is designated in a setup file for each group. If the survival factor is 1.0, then no cells are 'killed'. If the survival factor is less than 1.0, the corresponding numbers of cells are eliminated from the flasks of the given group. The actual number of cells eliminated is random according to a Poisson distribution with mean density calculated from the specified survival factor. After pooling and redistribution, the average mean density of cells is similar between flasks and comparison is made between one flask and another by looking at cell density or estimates thereof. When each group is kept separate in the pooling and redistribution (Mode 2), the operator controls the mean number of cells per aggregate in each pre-pooled flask for each group separately. Note that before pooling, the mean number of cells in an aggregate is as designated in the setup file. In Mode 2, after pooling and redistribution, the *mean density itself* is random for each group and could be quite different between groups depending upon the total aggregate volume for the flask. Therefore, one concentrates on the mean of the *total number* of cells in a flask as to whether it is the same or different between groups.

2.4. Slicing and sampling

Each 'flask' represents a 'gelatin block' to be sliced. Typically, each block is sliced into 50 μm sections. Section thickness is specified in a setup file. Also specified is how far to go into the block before sectioning starts and when in the block to stop it. After sectioning, a specified number of aggregate sections are selected at random. Based on the system used in the laboratory, typically a sample of 30 sections from about 1500–2500 aggregate sections are selected. The data consists of the sum of the counted number of cells over the 30 sections, the sum of the sectional volumes, and the total aggregate volume for each flask as in the laboratory. In contrast to the laboratory situation, we also know the *actual number* of cells in each flask.

2.5. Cell counting

The operator can choose from a set of various possible modes for counting the cells in an aggregate section. (1) Count a cell present only if at least x% of the diameter is in the section, where x is specified in a setup file. On the average, this would result in over-counting if x is less than 50% and under-counting is x is greater than 50%. (2) Count the cell present if the center of the cell is in the section. Since single points are being counted here, it is expected that this would result in neither over- nor under-counting. (3) Use an optical disector method. Do not count a cell section if any portion of it touches the top of the aggregate section but count all the other cell sections that are in the aggregate section. This mimics the optical disector method of counting in a thick section (Gundersen et al., 1988; West, 1999). Note that it is not necessary to include the idea of a guard region since we are not studying the effect of laboratory slicing in these particular computer experiments. These computer experiments are focused on the effect of bias in counting.

This part of the computer experiment is not intended to be a simulation of a laboratory experiment. This is the computer tool, which we use to study the effect of bias in counting. We use it to artificially create a bias in counting (method 1) in order to study the effect of the bias in the final result as compared with the case which again can only be done in the computer (methods 2 and 3), where we are assured there is no bias. Since we are deliberately introducing bias in method 1, we are also not using correction factors such as is given in the Abercrombie (1946) method.

2.6. Statistical tests

Cell density and total number of cells in a flask are estimated similarly as in the laboratory as described above. The two groups, control and test group are compared using the nonparametric Wilcoxon rank test (Brownlee, 1960). In the case of Mode 1 pooling, the test is performed upon estimated cell density, and in the case of Mode 2 pooling, on estimated total number as explained above. The Wilcoxon test is performed as follows. The values (densities or totals) for the entire set, both groups, are arranged in increasing order and given the corresponding rank. Then the ranks of each of the flasks in the control group are summed. The value of this sum is examined to see if it is above (or below) the cutoff number for a given significance value. The cutoff values are determined by the corresponding probability distribution of the ranks or by a resampling method (Simon, 1999). Computer experiments are repeated and each time it is determined whether the Wilcoxon test shows a difference between the control and test groups.

Since the test group is always set up to have mean cell number which is the same or less than the control group, a one-sided significance level of 0.05 was used. That means that in the case where there is no difference in mean cell number imposed between control and test group, the significance level becomes 0.1 for a two-sided test and one would expect to see false significance 10% of the time.

3. Results

Two types of computer experiments were performed. The first was of the Mode 2 type in which the two groups were kept separate and pooling was performed within groups. As described above, the number of flasks per group, aggregate volumes for each flask, numbers of aggregates per flask (post-pooling), aggregate size, cell size, and cell numbers per flask were copied from a laboratory experiment in which the test group cell number was significantly different than that of the control group (Won et al., 1997). In the second set of computer experiments, a laboratory experiment was copied in which the pooling and redistribution took place over the whole set of flasks. Subsequently the 'final flasks' were separated into the two control and test groups. 'Treatment' in the case of the computer experiment, consisted of eliminating certain numbers of cells from the test group by using a survival factor less than one. The laboratory experiment being copied in this case showed no significant difference between control and test group (Heller et al., 1997).

3.1. Computer experiment 1

There were seven flasks in each group. Total aggregate volumes ranged from 4.5 to 8.2 mm^3. Numbers of aggregates per flask ranged from 128 to 300. Each computer 'flask' was given the particular aggregate number and volume as in the data from the laboratory experiment. Aggregates were given diameters with a mean of 400 μm so that subsequent sectional volumes, although random, were close to those measured in the laboratory. Similarly, mean cell diameters were 15 μm. Laboratory estimates of cell numbers per flask were about 6500 per flask on the average in the control group. In the computer experiments, the test group mean was specified to be the same, or 20% less, or 30% less, depending upon the particular set of runs. As in the laboratory, sections were 50 μm thick and, for each flask, 30 aggregate sections were selected at random from the entire set of approximately 2000 sections.

Results for the three sets of computer runs are given in Tables 1–3. For each there are four columns. Each column gives the number of times significance occurred (out of 50) for ten sets of 50 repeats. In the first column, cell counting was done to deliberately induce over-counting. A cell was counted as being in an aggregate section if only 20% of the cell diameter was present in the section. This represents an extreme amount of over-counting as can be seen from the ratio between estimated cells and actual number of cells for the flask, (Table 4). In the second column, there was deliberate under-counting, a cell being counted only when more than 80% of its diameter was present in the section. In the third column, cell centers were counted. The fourth column gives results when the computer version of the optical disector method was used.

Table 1
No difference between groups, number of times out of 50 that significant differences were observed[a]

	Cell decision			Optical disector
	20%	80%	50%	
	6	4	4	5
	5	3	4	5
	4	5	3	4
	7	6	6	8
	1	5	5	6
	5	7	5	7
	3	4	4	3
	4	6	4	5
	7	7	7	8
	4	3	3	2
Total out of 500	46	50	45	53
Percent	9.2	10.0	9.0	10.6

[a] Ten sets of 50 repeats of a Mode 2 type computer experiment when there was no actual difference in cell number between the two experimental groups. Recorded are the number of times out of a possible 50 that significant differences were observed ($P<0.1$) for four different methods of cell counting. Cell decision 20% indicates over-counting, 80% under-counting, 50% is the counting of cell centers.

In Table 1, where there was no difference in reality between the two groups in regard to average cell number per flask, paired comparison t-tests were performed comparing each method of cell counting to the ideal of counting cell centers. There were no significant differences indicating that the four different methods of cell

Table 2
Twenty percent difference between groups, number of times out of 50 that significant differences were observed[a]

	Cell decision			Optical disector
	20%	80%	50%	
	30	25	33	28
	36	31	34	32
	32	34	36	33
	30	26	27	26
	27	31	28	29
	35	30	31	31
	34	32	34	33
	32	29	32	33
	34	33	33	34
	37	36	37	36
Total out of 500	327	307	325	315
Percent	65.4	61.4	65.0	63.0

[a] Ten sets of 50 repeats of a Mode 2 type computer experiment when the 'test' group had 20% less cells, on the average, than the 'control' group. Recorded are the number of times out of a possible 50 that significant differences were observed ($P<0.05$) for four different methods of cell counting. Cell decision 20% indicates over-counting, 80% under-counting, 50% is the counting of cell centers.

Table 3
Thirty percent difference between groups, number of times out of 50 that significant differences were observed[a]

	Cell decision			Optical disector
	20%	80%	50%	
	47	46	46	46
	45	42	45	47
	47	46	47	48
	46	44	47	44
	45	44	44	46
	48	46	47	48
	49	48	49	48
	47	45	45	46
	49	48	48	49
	48	46	47	46
Total out of 500	471	455	465	468
Percent	94.2	91.0	93.0	93.6

[a] Ten sets of 50 repeats of a Mode 2 type computer experiment when the 'test' group had 30% less cells, on the average, than the 'control' group. Recorded are the number of times out of a possible 50 that significant differences were observed ($P<0.05$) for four different methods of cell counting. Cell decision 20% indicates over-counting, 80% under-counting, 50% is the counting of cell centers.

counting that were used here did not affect the outcome of the experiment. It is to be noted that false significance occurred approximately 10% of the time, which is appropriate for a 0.1 level two-sided significance test. In Table 2, there was imposed an actual difference between the two groups in that the test group was given a mean cell number which was 20% less than that of the control group. Paired comparison t-tests showed no significant difference in results between the four methods of cell counting. In this case, the chance of picking up the difference between the two groups ranged from 0.61 to 0.65 irrespective of the method of cell counting. In Table 3, where the test group had mean cell number 30% less than the control group, the chance of observing the difference was approximately 0.91, not significantly different between counting methods as in the cases above.

Table 4
Ratio of estimated cells in a flask to actual number of cells[a]

Flask	Cell decision			Optical disector
	20%	80%	50%	
1	1.3665	1.0061	1.2088	1.1263
2	0.9668	0.6421	0.8192	0.7380
3	1.6270	1.1381	1.3910	1.3235
4	1.1803	0.8038	1.0065	0.9268
5	1.0844	0.6886	0.8628	0.8786
6	0.7643	0.5215	0.6357	0.6715
7	1.2371	0.8729	1.0791	1.0928
Average	1.17520	0.81044	1.00044	0.96536
S.D.	0.27841	0.21458	0.25397	0.23002

[a] Comparison of estimated number of cells to actual number of cells in seven 'control' flasks of a typical run of computer experiment 1. Reported is the ratio, estimated cell number/actual cell number for each of four cell counting methods; 20% cell decision indicating over-counting, 80% under-counting, 50% equivalent to counting cell centers, and optical disector method. The average and S.D. over the seven flasks is given for each of the cell counting methods.

Since the number of cells in any flask in the computer experiment is known, it is possible to compare the estimated number of cells obtained by counting from a sample of 30 aggregate sections with the actual number of cells present in the flask. In Table 4 is shown the ratio of estimated number of cells in a flask to actual number of cells in that flask for a typical run of control flasks (mean total number of cells = 6500) when there is deliberate over-counting (cell decision = 20%), under-counting (cell decision = 80%), neither over- nor under-counting (cell decision = 50%), and counting by the computer version of the optical disector method. Although there is considerable flask-to-flask variation, due to sampling variability, the average and S.D. for the optical disector method is close to that of the 50% cell decision whereas the 20% cell decision (over-counting) yields ratios 17.5% higher on the average, and the 80% cell decision (under-counting) yields ratios 19% lower on the average. The estimated S.D. for all four cell counting methods were similar to each other.

3.2. Computer experiment 2

In this set of computer runs a laboratory experiment was copied in which pooling and redistribution took place for the whole set of flasks. Subsequently to pooling, the 'final flasks' were separated into two groups; a control group and a test group into which a DA antagonist was added which did not produce a loss of cells.

The computer experiments were performed as in Section 3.1 above except that aggregate numbers and total aggregate volumes were set to resemble this laboratory experiment in which there were six flasks per group. The main difference between this set of computer runs and those of Experiment 1 was that the pooling of aggregates in the original flasks and redistribution into final flasks was done according to 'Mode 1' (combined) pooling. In this mode, all of the flasks were originally given the same mean parameter for numbers of cells. Then after pooling and redistribution, each group had a random number of cells eliminated according to a specified mean survival factor. The test group was made to have fewer cells than the control by giving it survival factors less than 1.0. Another difference in the Mode 1 pooling as compared with Mode 2 above, was that *density* of cells for each flask was estimated and compared between groups instead of total number of cells. As in Section 3.1, comparisons were made between the various methods of cell counting.

Table 5
Percent of times significance occurs ($P<0.05$) out of 200 runs for various survival factors and various cell counting methods[a]

Survival factor for test group	Cell decision			Optical disector
	20%	80%	50%	
1.0	9.5	10.5	9.0	12.0
0.85	54.0	44.5	50.5	50.5
0.80	64.5	60.0	62.5	61.5
0.75	83.5	80.0	81.0	83.0
0.70	95.0	90.5	93.0	91.5

[a] The percent of times out of 200 repeats of computer experiment 2 (Mode 1 type) that significant differences occurred at a one-sided significance level of 0.05, when the survival factor is 1.0 for both groups and then, in succession, 0.85, 0.80, 0.75, 0.70 for the test group as compared with full (1.0) survival for the control group. Four different types of cell counting were 20% indicating over-counting, 80% under-counting, 50% equivalent to counting cell centers, and optical disector.

In Table 5 is shown the percent of times out of 200 that significant differences occurred at a one-sided significance level of 0.05, when the survival factor was 1.0 for both control and test groups, and then in succession, 0.85, 0.80, 0.75, 0.70, for the test group with survival factor 1.0 for the control group each time. As can be seen from the table, there was very little difference between the various cell counting methods. It is noteworthy that when there was actually a 30% drop in cell density between control and test group, there was more than a 90% chance of the density estimate picking up the difference (irrespective of cell counting method) similarly as in the case of Experiment 1.

3.3. Optical disector method of counting

In order to compare the optical disector method of counting with the ideal method of counting cell centers, a simulated gelatin block from one flask (similar to Section 3.1) was sliced and a sample of 30 aggregate sections selected at random. An estimate of the total number of cells was made using the cell count and the sectional volumes of each of the 30 aggregate sections. This was compared with the actual total number of cells in the simulated flask. The sampling procedure was then repeated on the same simulated gelatin block 500 times with random selections for sampling each time. The ratio of estimated cells to actual number of cells for each of the two counting methods mentioned above was calculated each time.

For the optical disector method of counting, the average ratio of estimated cells to actual number of cells was 1.026 with a S.D. of 0.197 when the cells were clustered spatially within the aggregates. The average ratio was 1.030 with a S.D. of 0.098 when the cells were spatially random. When the cell centers were counted, the average ratio and S.D. were very similar to that of the optical disector method. For clustered cells, the average was 1.028, S.D. 0.196 and for unclustered cells, 1.034, S.D. 0.100.

By contrast, when the cells were counted with a 20% cell decision (a cell was counted if at least 20% of the cell section was in the aggregate section), the average ratio for clustered cells was 1.224, S.D. of 0.224 and for unclustered cells, an average of 1.217, S.D. of 0.110. Here, we see the effect of over-counting. In the under-counting situation where a cell was not counted unless at least 80% of the cell section was in the aggregate section, for clustered cells the average ratio was 0.831, S.D. = 0.166 and for the unclustered cells, the average ratio was 0.851 with S.D. = 0.087.

These results indicate that the optical disector method gives similar estimation results to the ideal method of counting cell centers. In this simulation, they are similar both in regard to average ratio and S.D. Both were close to the ideal of 1.00.

In the case of over- or under-counting, it appears that S.D. varies as does the mean. That is, the S.D. was less in the case of under-counting where the mean was less and more in the case of over-counting where the mean was greater.

4. Discussion

The purpose of running this set of computer simulation experiments was to discover the effect of deliberately induced over- or under-counting in the eventual results of a cell counting experiment. Actual laboratory experiments involving cell counting were mimicked in the sense that the computer experiments contained similar numbers of flasks per group, similar volumes, numbers, and sizes of aggregates and similar numbers of cells to be counted. Also, the laboratory procedure of pooling aggregates and redistributing them into fresh flasks was copied. In addition, the laboratory procedure of sectioning and subsequent random sampling of 30 aggregates from about 2000 sections was copied. What was not copied was the method of cell counting since, in the computer experiments, either ideal counting of cell centers or deliberate over- or under-counting could be performed.

Before performing these computer experiments, it was not known how over- or under-counting would effect the observation of significant differences between experimental groups with regard to cell number. In the actual laboratory experiments, according to the criteria for cell counting, it would seem evident that, if anything, under-counting might occur. However, if there were under-counting it would not seem to be extreme. In the computer experiments, severe cases of over- and

under-counting were induced deliberately to see the frequency with which they affected the final result.

In the computer, it was possible to avoid bias by counting cell centers. Results from that method were used as the standard with which to compare the other cell counting methods. In the case where there was in reality no difference between the two groups in regard to mean cell number, false significance occurred from 9 to 10.6% as would be expected from a one-sided 0.05 significance level which corresponds to a two-sided 10% level. Remarkably, this result occurred also in the cases where there was over- or under-counting. Consider the cases where there was in reality a difference between the two groups. When the test group had 20% less mean number of cells than the control group, the ideal method of counting picked up the difference 65% of the time, over-counting picked it up 65.4% of the time and under-counting 61.4% of the time on the average and according to statistical tests performed on these results, again remarkably the over- and under-counting methods showed no significant difference from the ideal method of counting cell centers. The results are similar in the case where in actuality the test groups had mean number of cells 30% less than the control group. The results were also similar in the case of Experiment 2 with 'Mode 1' pooling. As to be expected, the computer version of the optical disector method gave results very similar to those of the ideal method.

The disector methodology is strongly recommended when absolute numbers of cells are being estimated (Gundersen et al., 1988; West, 1999). But in the case of experiments dedicated to comparing experimental groups, it is an open question as to how critical the cell counting method is. Possibly it depends upon the particular circumstances. Here, we present an example where deliberately induced bias in cell counting evidently did not effect the end result of distinguishing difference in cell number between experimental groups. Of course, it would not be claimed that such a result would hold in general. But similarly, one could not claim that bias necessarily leads to spurious end results irrespective of the environment and conditions for cell counting.

In the experimental situation there is generally no practical way of counting an identified point such as 'center' in a cell. In addition, aside from serial reconstruction, there is no way of determining the actual number of cells in a sample. As a result, arguments regarding the appropriateness of counting methods must rely on indirect reasoning without experimental verification. From the theoretical point of view, stereological methods (Gundersen et al., 1988) have introduced a great advance in the available techniques for cell counting. However, it still remains an open question as to the final effect of a given cell counting method, in a particular experimental environment, on the determination of differences between experimental groups even in the presence of bias in cell counting. The computer simulation approach provides a means for comparison of cell counting methods since identified points in the cells can be counted and the actual number of cells present can be determined.

Acknowledgements

This research was supported by MH-28942 and DA-09764 to A.H. We wish to express our appreciation to the reviewers for their incisive remarks and suggestions.

References

Abercrombie M. Estimation of nuclear population from microtome sections. Anat Rec 1946;94:239–47.

Brownlee KA. Statistical Theory and Methodology in Science and Engineering. New York: Wiley, 1960:189.

Coggeshall RE, Lekan HA. Methods for determining numbers of cells and synapses: a case for more uniform standards of review. J Comp Neurol 1996;364:6–15.

Diggle PJ. Statistical Analysis of Spatial Point Patterns. New York: Academic Press, 1983:55.

Glaser EM, Wilson PD. The coefficient of error of optical fractionator population size estimates: a computer simulation comparing three estimators. J Microsc 1998;192(Pt 2):163–71.

Guillery RW, Herrup K. Quantification without pontification: choosing a method for counting objects in sectioned tissues. J Comp Neurol 1997;386:2–7.

Gundersen HJG, Bendtsen TF, Korbo L, Marcussen N, Moller A, Nielsen K, Nyengaard JR, Pakkenberg B, Sorensen FB, Vesterby A, West MJ. Some new, simple and efficient stereological methods and their use in pathological research and diagnosis. APMIS 1988;96:379–94.

Heller A, Won L, Choi H, Wainer B, Hoffmann PC. Study of dopaminergic neuronal differentiation and survival by use of three-dimensional reaggregate tissue culture and monoclonal hybrid cell lines. In: Hefti F, Weiner WJ, editors. Progress in Parkinson's Disease Research. New York: Futura Publishing, 1992:403–26.

Heller A, Won L, Choi H, Heller B, Hoffmann PC. Reaggregate cultures. In: Tyson CA, Frazier JM, editors. In Vitro Biological Systems: Preparation and Maintenance, Methods in Toxicology. New York: Academic Press, 1993:27–45.

Heller A, Choi H, Won L. Regulation of developing dopaminergic axonal arbor size in three-dimensional reaggregate tissue culture. J Comp Neurol 1997;384:349–58.

Hemmendinger LM, Garber BB, Hoffmann PC, Heller A. Selective association of embryonic murine mesencephalic dopamine neurons in vitro. Brain Res 1981;222:417–22.

McShane LM, Palmatier MA. Spatial distribution of neurons in tissue culture wells: implications for sampling methods to estimate population size. Stat Med 1994;13:523–40.

McShane LM, Albert PS, Palmatier MA. A latent process regression model for spatially correlated count data. Biometrics 1997;53(2):698–706.

Saper CB. Counting on our reviewers to set the standards. J Comp Neurol 1997;386:1.

Schmitz C. Variation of fractionator estimates and its prediction. Anat Embryol 1998;198:371–97.

Schmitz C, Rub U, Korr H, Heinsen H. Nerve cell loss in the thalamic mediodorsal nucleus in Huntington's disease. II. Optimization of a stereological estimation procedure. Acta Neuropathol 1999;97:623–8.

Simon JL. Resampling: The New Statistics. Arlington, VA: Resampling Stats, 1999:273.

Vidal L, Heller B, Won L, Heller A. Quantitation of dopaminergic neurons in three-dimensional reaggregate tissue culture by computer-assisted image analysis. J Neurosci Methods 1995;56:233–44.

West MJ. Stereological methods for estimating the total number of neurons and synapses: issues of precision and bias. Trends Neurosci 1999;22:51–61.

Won L, Price S, Wainer BH, Hoffmann PC, Bolam JP, Greengard P, Heller A. A correlated light and electron microscopic study of dopaminergic neurons and their synaptic junctions with DARPP-32-containing cells in three-dimensional reaggregate tissue culture. J Comp Neurol 1989;289:165–77.

Won L, Ghetti B, Heller B, Heller A. In vitro evidence that the reduction in mesencephalic dopaminergic neurons in the weaver heterozygote is not due to a failure in target cell interaction. Exp Brain Res 1997;115:174–9.