Histol Histopathol (2002) 17: 639-648

http://www.hh.um.es

Histology and Histopathology

Cellular and Molecular Biology

Review

Counting particles in tissue sections: Choices of methods and importance of calibration to minimize biases

C.S. von Bartheld

Department of Physiology and Cell Biology, University of Nevada School of Medicine, Reno, USA

Summary. Investigators must choose between counting methods to quantify microscopic particles in tissues. The conventional profile-based ("model-based" or "2D-") counting methods have been criticized for their potential biases due to assumptions about shapes, sizes, and orientation of particles when converting profile counts into cell numbers. New stereological methods ("design-based" or "3D-") methods such as the optical disector or physical disector were initially introduced as being inherently unbiased. Recent calibration analyses and comparisons of results from different investigators have revealed the potential for significant biases in the most efficient and most frequently used design-based method, the optical disector. This review aims to objectively assess the strengths and limitations of current profile- and disector-based cell counting methods by examination of studies in which these methods have been calibrated against the "gold-standard", counts obtained by 3-dimensional reconstruction of serial sections. Advantages and disadvantages of each counting method and the associated embedding and sectioning techniques are compared and frequent mistakes and pitfalls of each technique are discussed. The importance of a calibration step for each technique is emphasized, and a protocol is provided for a quick and simple calibration by a "sampling" 3-D reconstruction of limited serial sections. Trends in the usage of counting methods are analyzed in four major journals. It is hoped that this review will be helpful, for both investigators and manuscript reviewers, in clarifying some of the contentious issues in the choice and implementation of appropriate methods for particle counting in tissue sections.

Key words: Stereology, Optical disector, Calibration, Bias, Sampling

Offprint requests to: Christopher von Bartheld, Department of Physiology and Cell Biology, Mailstop 352, University of Nevada School of Medicine, Reno, NV 89557 USA. Fax: 775-784-6903. e-mail: chrisvb@physio.unr.edu

Brief historic overview

Counting of cells in organs, nuclei or ganglia is crucial to assess effects of agents, treatments or genes, and for quantitation of processes in development, aging and disease. Cell counts have been essential in the discovery of important concepts in cell biology, for example apoptosis. In general, microscopic particles such as cells can not be counted in tissues directly, but rather have to be counted in histological sections. This has been done traditionally by counting particle profiles in sections. Since many particles appear as profiles in more than one section, profile counts have to be converted into estimates of particle number. How often a particle will appear in sections depends on the particle size, shape and orientation. Therefore, one has to make assumptions about particles when applying a correction factor that converts the number of particle profiles into estimates of particle number (Abercrombie, 1946; West, 1993; Coggeshall and Lekan, 1996; Hedreen, 1998a), and this method will yield reasonably correct estimates only when these assumptions are largely met. Such profile counting methods are also known as "model-based" or "2D-" counting methods.

About 15 years ago, new, conceptually very elegant, methods were developed which promised to revolutionize particle counting by providing unbiased estimates of particle numbers (Gundersen et al., 1988a,b; Williams and Rakic, 1988). These methods identify particles uniquely, so they are counted only once, by either comparing sections (physical disector), or by randomly sampling particle numbers in 3D space in relatively thick sections (optical disector). The optical disector is spelled with one "s", because the "di-sector" refers to the use of "two" optical sections. The average density of particles (N_v) is applied to the entire reference space which is estimated by point-counting with the Cavalieri method (Howard and Reed, 1998). The number of particles (N) in the reference volume (V_{ref}) is calculated as $N = V_{ref} \times N_v$. Hallmarks of the technique are that particles are uniquely identified (by examination

of additional physical or optical sections) and that they have to be sampled randomly, so that each particle has the same chance of being sampled. The new methods, especially the optical disector, appeared to be highly efficient and promised to evolve as the method of choice for virtually all applications in quantitative morphology (Bolender et al., 1991).

In the past five years, problems with the design-based methods began to surface. The new methods proved to be reluctantly embraced by investigators, and the large majority of studies continued to use the older techniques (Coggeshall and Lekan, 1996). Questions arose whether the new methods were indeed unbiased, or at least less biased than the conventional techniques (Clarke, 1992; Popken and Farrel, 1996, 1997; Guillery and Herrup, 1997; Hatton and von Bartheld, 1999; Schmitz et al., 1999; von Bartheld, 1999, 2001; Benes and Lange, 2001). It was realized that, for one entire decade after its introduction, the optical disector had never been calibrated, and it came as an embarrassing surprise when recent calibration analyses showed that the optical disector, even when used "correctly" (as recommended), can have a substantial systematic bias (Hatton and von Bartheld, 1999; von Bartheld, 1999). Many advocates of the optical disector had believed that, on theoretical grounds, the disector method was "inherently unbiased" and therefore should not and could not be calibrated (Cruz-Orive, 1994; Mayhew and Gundersen, 1996; Howard and Reed, 1998).

An intense dispute developed between those using the conventional profile-based methods (Abercrombie, 1946; Konigsmark, 1970; Clarke and Oppenheim, 1995; Guillery and Herrup, 1997; Benes and Lange, 2001), and those advocating the newer, often termed "unbiased" stereological methods such as the optical and physical disector (Gundersen et al., 1988a,b; Williams and Rakic, 1988; Mayhew, 1992; West, 1993, 1999; Coggeshall and Lekan, 1996; Howard and Reed, 1998), although less dogmatic positions advocating a compromise have recently appeared (Saper, 1999; Geuna, 2000; von Bartheld, 2001). The current review attempts to provide a balanced evaluation of profile- and design-based counting methods and to objectively answer questions about the strengths, limitations, biases, and efficiencies of each of the two most commonly recommended counting methods, profile-counting and the optical disector. It is hoped that a review with emphasis on practical issues such as choices of fixation, embedding and sectioning techniques, sampling, calibration and avoidance of common mistakes will be a helpful guide for those within or entering the field of quantitative morphology.

Conventional, profile-based (2D) methods

The conventional profile-based counting methods are being used by the overwhelming majority of quantitative studies (Coggeshall and Lekan, 1996; this review, see below). These methods have been described in detail by Abercrombie (1946), Konigsmark (1970) and Clarke and Oppenheim (1995). Briefly, one sections through the tissue of interest, collects about every 10th section, counts all particle profiles in the section, extrapolates for the entire set of sections, and then applies a correction factor to account for the overcount due to the presence of split particles in multiple sections (Abercrombie, 1946; Clarke, 1993). This method is typically used in relatively thin paraffin sections or cryosections. Advantages of this technique are the better morphology and resolution of particles in the thin sections, the conceptually easy implementation, and the widespread availability of necessary equipment in virtually every routine histology lab. Disadvantages include the uncertainty about the extent of bias due to

Table 1. Advantages and disadvantages of cell counting methods.

COUNTING METHOD	ADVANTAGES	DISADVANTAGES
Profile-counting (Abercrombie-corrected) (Abercrombie, 1946)	Minimal, ubiquitous equipment Easy concept Compatible with thin sections Relatively efficient and fast Large data base from previous work	Biased due to assumptions about particle size, shape Large bias when section thickness is small relative to particle height Larger bias with heterogeneous particles
Optical Disector (Howard and Reed, 1998)	Relatively efficient Unaffected by heterogeneous particles Requires relatively little extra equipment	Requires thick sections Antibodies may not penetrate entire section thickness Differential z-axis distortion may introduce bias
Physical Disector (Howard and Reed, 1998)	Largely unbiased Applicable with thin sections	Requires tedious alignment of two sections Requires special equipment Inefficient, time consuming
Empirical Method (Coggeshall et al., 1990)	Unbiased	Inefficient, time consuming, tedious
3D-Reconstruction from Serial Sections (Coggeshall et al., 1990) (Hatton and von Bartheld, 1999)	Unbiased	Extremely inefficient Time consuming, tedious Impossible with large numbers or densely packed particles

assumptions about the shape, size and orientation of particles which need to be made when applying a correction factor (Table 1). When the particles are relatively small compared to the section thickness and the particle shape is close to a sphere, such assumptions used for the correction factor may be acceptable and appear to provide reasonably accurate estimates (Clarke, 1992, 1993; Clarke and Oppenheim, 1995). When the particles are heterogeneous, and are relatively large compared with the section thickness, the bias may be significantly larger.

Optical disector

The optical disector is currently used in about 5% of quantitative studies (Coggeshall and Lekan, 1996; this review, see below). The optical disector has been described in detail by Gundersen et al. (1988a,b), Williams and Rakic (1988), West (1993), and Howard and Reed (1998). Advantages of the method include that it is theoretically unbiased, no assumptions have to be made about the size, shape, or orientation of particles, and the optical disector is highly efficient and relatively easy to use. Disadvantages include that the method requires thick sections, that its accuracy can be affected by differential shrinkage and compression of sections, and that some special equipment such as a sensitive z-axis encoder is needed. The limitation of thick sections presents a particular problem in studies employing immunolabel (which are a major fraction of quantitative morphological studies), because the antibodies may not penetrate the entire section thickness, and such sections can not be used for the optical disector.

Other counting methods listed in Table 1, such as the physical disector, 3D-reconstruction of serial sections, and the "empirical method" are very rarely used, because they are too time-consuming and tedious and require specialized equipment, such as for alignment of section pairs for the physical disector (for some solutions, see Guntinas-Lichius et al., 1993; Korkmaz et al., 2000).

Suitability of histological techniques for cell counting

Not all embedding, sectioning and staining techniques are equally suited for different cell counting methods. Investigators should carefully evaluate the needs of the particular tissues in terms of tissue processing for optimal recognition of particles, and then decide which counting method will be most compatible with the requirements for reasonable tissue processing. For example, if particles can not be identified in thick sections, the optical disector may not be suitable. If the antibody needed for labeling cells does not bind the antigen in paraffin-embedded tissue, paraffin sections may not be an option.

In general, there are advantages and disadvantages associated with each histological technique with regard to counting methods (Table 2). As listed in Table 2, paraffin embedding and sectioning have several important advantages, including excellent morphology (especially after Carnoy's or Methacarn fixation which renders significantly better morphology than fixation with aldehydes) and easy preparation of serial sections for calibration analyses. There is differential distortion of sections in the z-axis, so one has to be careful with the use of guard spaces (Hatton and von Bartheld, 1999). Guard spaces are zones of at least 4 µm thickness outside the counting frame at the lower and upper margins of the sections. They are supposed to aid in the identification of particles of interest and to avoid problems associated with cutting artifacts such as uneven surfaces and "lost

Table 2. Suitability of histological techniques of embedding and sectioning for cell counting methods.

HISTOLOGICAL TECHNIQUE	ADVANTAGES	DISADVANTAGES
Paraffin	Easy, safe, inexpensive Excellent staining and morphology Ubiquitous equipment Ribboning makes reconstruction of serial sections easy Even surfaces	Differential distortion of z-axis May be incompatible with preservation of antigenicity for immunolabeling
Cryo-embedding/sectioning	Ubiquitous equipment No distortion of z-axis Compatible with most immuno labeling Easy, safe, and inexpensive	Mediocre staining and morphology Serial sectioning may be difficult
Vibratome	Equipment widely available Compatible with immunolabeling	Uneven surfaces Serial sections difficult to reconstruct due to uneven surfaces Differential distortion of z-axis
Methacrylate resin		Poor staining and morphology Cumbersome embedding and sectioning Wetting required for thick sectioning Differential distortion of z-axis
Celloidin resin	Thick sections obtained easily Good-excellent morphology	Tricky staining and processing of tissue Serial sections tedious

caps", i.e., small object fragments unnoticed or missing at section surfaces (Hedreen, 1998a). Cryosections are frequently used, most often in studies employing immunolabeling. Cryosectioning renders only mediocre to good morphology, but this may be sufficient for many counting tasks, and there seems to be no distortion of the z-axis in frozen sections (Hatton and von Bartheld, 1999; see also Harding et al., 1994), allowing one to use guard spaces with cryosections as originally proposed for the optical disector. Vibratome sections are also used in immunolabeling studies, but due to the uneven surfaces and the mediocre morphology, such sections are not ideal for the optical disector method. Methacrylate resin embedding and sectioning was originally proposed as the method of choice for the optical disector (Gundersen et al., 1988a; Howard and Reed, 1998), but this resin turned out to suffer from numerous serious drawbacks, including poor morphology and staining properties, poor sectioning qualities, extreme difficulty in obtaining serial sections for calibration analyses, and a differential distortion of the z-axis which can introduce significant bias when guard spaces are used (Hatton and von Bartheld, 1999). Instead of methacrylate resin, celloidin resin is now being increasingly used (Heinsen et al., 2000). Thick celloidin sections are obtained more easily than thick glycolmethacrylate sections, and the morphology is better, but tissue processing for celloidin is tricky, and it can be difficult to achieve even staining in thick celloidin resin sections.

In summary, because of the optimal identification of particles, we recommend the use of paraffin sections for the optical disector as a first choice, followed by cryosections or celloidin resin. For use with profile counting and correction with the Abercrombie formula, either paraffin, cryosections or thin resin sections can be used, because the morphology in thin resin sections is excellent. One has to remember though, that the larger the particles are relative to the section thickness, the more bias can be expected due to 'lost caps' (Clarke, 1992, 1993). Availability of equipment as well as experience and preferences of individual investigators will also determine which counting method will be employed with success.

Is the optical disector unbiased?

When the design-based methods were introduced, they were thought to be inherently unbiased (Gundersen et al., 1988a,b; Williams and Rakic, 1988; Cruz-Orive, 1994; Mayhew and Gundersen, 1996; Howard and Reed, 1998). Mathematically, and in a perfect world, there should be no bias associated with the new methods. Accordingly, there seemed to be not only no need for calibration, but "unbiasedness was a built-in property of design-based methods" and the results of the design-based methods were considered to be the new "calibration standard" (Mayhew and Gundersen, 1996). For many years, such arguments discouraged attempts to calibrate the optical disector. Three main arguments have since been raised against the notion of the unbiasedness of design-based methods.

(1) All methods have some bias (Guillery and Herrup, 1997; Saper, 1999), and the design-based methods are no exception. There are multiple sources of potential error, including observer fatigue, difficulty in identifying objects in sections, in identifying the borders of the reference space accurately, and measuring section thickness precisely (Clarke and Oppenheim, 1995; Guillery and Herrup, 1997; Hedreen, 1998a; Benes and Lange, 2001). (2) The optical disector, as recommended in its original form, is vulnerable to a distorted distribution of particles due to differential shrinkage or differential compression of tissue sections in the z-axis that occurs during or after sectioning (Hatton and von Bartheld, 1999). The resulting bias can be largely eliminated by minimizing the guard space, but minimizing the guard spaces may also interfere with recognition of particles or fragments (Gundersen et al., 1988b; Hedreen, 1998a), and thus it limits the power of the optical disector as it makes the optical disector vulnerable to the "lost caps" problem (Hedreen, 1998a). (3) Use of the same design-based method (optical disector) by different groups of investigators for the same structure in the same species has resulted in widely divergent estimates of particle numbers (2-3 fold difference), indicating biases of as much as 50-100% (Guillery and Herrup, 1997; Calhoun et al., 1998; Insausti et al., 1998; Schmitz et al., 1999; von Bartheld, 2001), and such divergence is in the same range as those for profile-based (2D-) counting methods (Schmalbruch, 1987). Thus, it is now becoming clear that the new design-based methods, even when they are used "correctly" (Geuna, 2000), are not unbiased, but can have substantial biases (Saper, 1999). An important question remains to be answered: how much are the design-based methods biased, and how much are they biased compared with the conventional methods? If they are less biased than conventional methods, then they should generally be preferred over the old methods. But has this been truly established?

How bad are the conventional methods really?

The introduction of the new design-based counting methods has elicited an intense debate regarding questions of the validity and accuracy of the profile-based counting techniques. At one extreme, it was claimed that no new profile-based counting should be accepted (Mayhew and Gundersen, 1996) and all previous profile-based results needed to be re-done or re-evaluated with the new methods (Coggeshall and Lekan, 1996). On the other extreme, it has been argued that, when used appropriately, the profile-based methods render results which are at least comparable with, if not better than, the new methods, and that it can often be justified to use the old methods (Clarke, 1992; Clarke and Oppenheim, 1995; Guillery and Herrup, 1997; Benes and Lange, 2001). Some recent data indicate that

the old methods, when used properly, render biases that are for many purposes acceptable and are within the same range as those that can be obtained with the physical or optical disector (Schmalbruch, 1987; Pover and Coggeshall, 1991; Clarke and Oppenheim, 1995; Guillery and Herrup, 1997; Hatton and von Bartheld, 1999). Since many investigators are interested to find out if there is a relative and significant difference in numbers between treatment and control groups rather than attempting to estimate absolute numbers, it has been argued that in many cases the systematic bias will cancel out of the equation, and thus conventional methods may suffice for this task (Saper, 1996; but this may not always be true, see Guillery and Herrup, 1997, and arguments below). It should be emphasized that the profile-based techniques must be used prudently, and that, when information about absolute numbers is required, major biases should be ruled out by calibration.

Should all counting methods be calibrated?

Since there is an emerging consensus that all counting methods are somewhat biased, it can be argued on the one hand that it does not matter which method is used. Another approach would be to try to find out exactly how much the methods are biased in order to evaluate such errors and obtain some reasonable estimate of whether or not such bias will affect the general conclusions of the study in question. It has been argued in the past that all counting methods should be calibrated (Coggeshall et al., 1990; Pover and Coggeshall, 1991; Coggeshall, 1992), and that such a calibration is particularly important for new methods (Geuna, 2000). However, the same authors who demand a calibration analysis for the model-based counting methods, seem to "exempt" the new design-based methods (Coggeshall, 1992; Coggeshall and Lekan, 1996; Geuna, 2000), thus creating a double standard. Calibration analyses of counting methods have been carried out. This was first done by Coggeshall and coworkers for the physical disector (Coggeshall et al., 1990; Pover and Coggeshall, 1991), and later, in a more limited fashion, by Clarke and Oppenheim (1995) for the profile counting technique. The physical disector was also calibrated, albeit indirectly, by Popken and Farel (1996, 1997), and optical disector-counting as well as Abercrombie-corrected profile-counting were calibrated by Hatton and von Bartheld (1999). In this study, both methods were directly compared with the "gold standard", a complete 3D reconstruction of serial sections.

While such an in-depth analysis obviously is beyond the scope of most projects, it became clear that a limited, "sample" serial reconstruction of only a few sections can provide a reasonably useful estimate of the bias involved in a particular set-up and with the specific parameters of a project. Thus, if absolute numbers of particles are required, or if the differences between treatment groups are small, it is recommended to include a simple and quick calibration step (as shown in the protocol below) to assess the magnitude of the bias in the particular setting and to evaluate whether the counting method meets the required precision and the bias is under control. If investigators include such a calibration step ("quality control") and report the results in their publications, it would also reveal the biases and their variabilities and ranges, and possibly would lead to a better understanding of how the biases were generated. On a more theoretical note, the use of a calibration step to adjust estimates of counts is in part an application of the "empirical method" which was initially introduced by Gaule and Lewin (1896), and revived by Coggeshall in the 1990s (Coggeshall et al., 1990; Pover and Coggeshall, 1991; for review, see Hedreen and Vonsattel, 1997; Hedreen, 1998b).

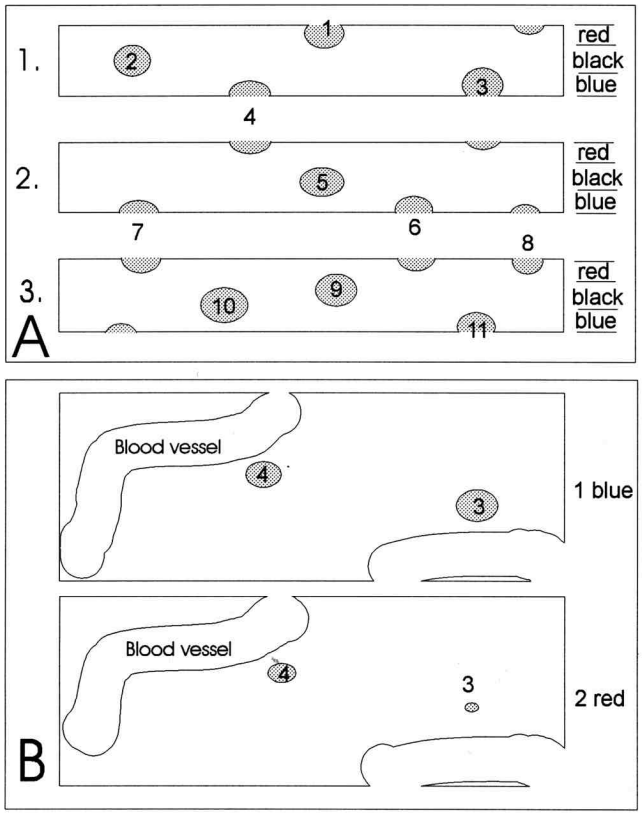

Fig. 1. The top panel **(A)** shows three tissue sections with particles as they would appear when collapsed in the z-axis of adjacent sections (1.-3.). Each particle which has its center within the section is numbered from 1-11 in the sequence as they would appear when focusing from the top of section 1 to the bottom of section 3. As explained in the text, for each section, three transparencies should be prepared so that all particles intersecting the top surface will be drawn in red, particles not intersecting either surface are drawn in black, and particles intersecting the bottom surface are drawn in blue. The bottom of the figure **(B)** shows the microscope view of section number 1, focus on the bottom surface (to be drawn in blue), and section number 2, focus on the top surface (to be drawn in red). Landmarks such as blood vessels or other prominent structures should be used to unambiguously identify fragments of cells or cell nuclei in adjacent surfaces. The data collection for the 3 sections is shown in Table 3.

A quick and simple calibration step

I recommend a simple verification step to make sure that the potential bias of a counting method is "under control". Although this protocol will mostly be used to examine only one region, samples can be obtained from several regions if it is suspected that variations in packing density or other tissue parameters may introduce different, region-specific biases. There is no equipment needed for the calibration step beyond that already needed for an optical disector or a profile counting analysis. About 5-10 serial stained tissue sections (the same ones to be used for disector or profile counting) are needed, as well as a microscope with either a video system or a camera lucida (drawing tube), transparencies, and three different color markers.

1. Take 5-10 serial 20-30 µm tissue sections on glass slides. Place slides on the stage of a microscope with camera lucida, select area of interest, and using a 40x objective, draw landmarks (blood vessels, ventricles, prominent cells or nuclei) within the area of interest (Fig. 1B).

2. Prepare 3 different transparencies to draw on. On the first, using a red pen, mark all cells (or nuclei of cells) which are located at the upper surface of the tissue section, which approach the surface very closely, or which are cut at the upper surface. This can be evaluated by focusing up and down. If in doubt, confirm by going to higher magnification (63x or 100x oil). The second transparency of the same area is prepared in black color by drawing all cells or nuclei which are clearly in the core of the tissue section and do not approach or touch either the upper or the lower surface of the section. The third transparency is used to draw in blue color all cells or nuclei which touch or intersect the lower surface of the tissue section (Fig. 1A).

3. Overlay transparencies of each of the colors and identify each cell or nucleus and decide whether it truly belongs to the "core" (black) transparency, or to the red (intersecting upper surface) or to the blue (intersecting lower tissue surface). When this is unambiguously resolved for each cell or nucleus, procede accordingly with the next tissue section, and so on.

4. Find the same area of view in the adjacent tissue section (use landmarks to unambiguously identify the area). Prepare the same kind of camera lucida drawings as for the first section. We recommend reconstruction of at least 100 particles.

5. Overlay transparencies for the adjacent surfaces from two adjacent tissue sections (make sure that you compare the upper of the former and the lower surface of the subsequent adjacent tissue section). Again identify each cell or nucleus. You will experience that sometimes you cannot find a cell fragment in the next section even though the preceding section clearly contained a split cell or nucleus. These missing fragments apparently are "lost caps" (see Hedreen, 1998b).

6. Go through all sections and count all particles within the sample area, collapsing all those profiles that appear in two adjacent tissue sections so that they are counted only once (Table 3). The total number is the true number of objects in this sampling area (unless you missed a particle or mistakenly identified it).

7. Compare with the number derived by optical disector (or profile) counting of the same area in the same sections. Divide the disector or profile-derived numbers by the 3D-reconstruction-derived number. The % above or below 100 gives the % overcount or undercount. This is the approximate bias of the method for the region of interest in this particular set-up.

Does it matter which counting method is used?

Yes, it does matter which counting method is used, because the requirements differ for different studies, and not every counting method can be used for every task. Some of these requirements are obvious. If particles are to be counted which can only be identified in relatively thin sections, then the optical disector is not suitable, because this method requires thick sections. If particles are highly heterogeneous in size and if they may change sizes after treatment, profile-counting may yield unacceptable biases. If particles can not be stained in resin sections, an embedding and sectioning technique has to be used which allows the staining of those particles. If particles are to be counted which are very densely packed, and with no landmarks in the tissue, the physical disector (Gundersen et al., 1988a; West, 1993) or the optical disector (Clarke, 1993) may not be suitable. On the other hand, there are many tasks where one could choose either model-based or design-based counting methods with the sectioning and staining parameters. If this is the case, it is usually preferable to use the design-based methods, because they are not affected by the possibility of biases due to changes in the

Table 3. Data collection of 3D serial reconstruction shown in Fig. 1. Number of profiles and number of cells.

TRANSPARENCIES	RED FRAGMENTS	BLACK WHOLE CELLS	BLUE FRAGMENTS	TOTAL NUMBER		WHOLE CELLS
Tissue section 1	1	1	2	4		=4
Tissue section 2	2	1	3	6	2 identical: -2	=4
Tissue section 3	3	2	1	6	3 identical: -3	=3
		Total Fragments: =16				Total cells: =11

sizes, shapes, or orientations of the particles. Most often, investigators are not so much interested in the absolute numbers of particles, but rather whether any significant changes occur between two groups, for example two stages during development, or a treatment group and a control group. When such changes in numbers are huge (such as a 90% loss of cells), then it does not matter for the validity of the main conclusion if the counting method has a relatively small bias of, e.g., 8%. However, when one attempts to detect discrete differences, for example a 15% change in numbers, with a standard error of 8%, then it is possible that relatively small biases of 8% due to changes in the size of the particles with the treatment could be interpreted as a false-positive result. A change in particle size of 8% may not be visible in the sections and go unnoticed, but may cause a false-positive or false-negative result, thus leading to a false conclusion.

It has been argued that when profile-based counting methods are used "correctly", i.e. the particle diameter is relatively small compared with the section thickness, such biases are relatively small, unless the change in particle size and/or shape and orientation is quite large (Clarke, 1992; Clarke and Oppenheim, 1995; Guillery and Herrup, 1997). Nevertheless, since counting techniques are available which are not affected by this kind of bias, investigators should use the most powerful tool when there is any chance that the result and conclusion of the study could be affected.

Calibration analyses will reveal whether the biases of the counting methods are reasonably small relative to the requirements of the task. For example, a 5% bias in a profile-counting technique will be preferable to a 15% bias in an optical disector counting technique and vice versa. Both design-based and model-based counting techniques can have substantial total biases associated with them, depending on the specific parameters of the studies. It is not known if these are due to observer bias, sampling bias, systematic bias, identification of particles bias, bias due to differential shrinkage or compression of sections, or to lost caps. Therefore, the best we can do is to try to assess the magnitude of total bias in each study, and to use the sectioning-, staining- and counting method which can be expected to provide the overall most truthful (least-biased) estimate. In some cases, it may also be relevant with which sectioning and counting method the investigator's lab is more familiar and experienced with, and what kind of facilities are available. However, if the optical disector method is clearly preferable over the profile-counting method, but a z-axis encoder is not available, then the lab should invest in the relatively minor expenses associated with the necessary equipment and implementation of the proper counting technique.

Frequent mistakes with 2D methods

When the particles to be counted are relatively large compared to the section thickness, there will be larger errors, because this will result in multiple fragments, and each fragment is potentially associated with errors such as those resulting from "lost caps" (Hedreen, 1998a). On the other hand, when the particle is small relative to the section thickness (e.g., a 2 µm nucleolus and a 10 µm section), there will be relatively few split nucleoli, and thus few lost caps, and the potential bias due to lost caps is much smaller and likely negligible.

Users of the profile-based methods often count nucleoli rather than nuclei, for the reasons stated. However, if the cells in question contain more than one nucleolus per cell nucleus, the number of nucleoli will be greater than the number of cells, and the ratio of nucleoli/cells has to be established to correct for the nucleoli overcount. This seems to be a trivial necessity, but it is surprising how many studies disregard this and publish nucleolar counts without either considering, adjusting for, or reporting the nucleoli/cell ratios. This appears to be the main reason for the so-called overcount of neurons in Coggeshall's calibration study (Coggeshall et al., 1990; Clarke and Oppenheim, 1995).

Another common misconception is that an overcount due to multiple profiles can be prevented by simply spacing the counting of sections so that profiles from the same particle are counted only once. This, of course, does not work, because profile-counting is a sampling procedure, and the samples are extrapolated to the entire reference space.

Another misconception is that when using profile-based methods and comparing relative numbers from two different groups, any systematic bias always cancels out of the equation (Saper, 1996). This argument could be taken to support the notion that when using relative numbers (% difference between two groups), one can use ANY method, no matter how much biased, because the bias cancels out in the end. This is not necessarily true. When the treatment changes the size, shape or orientation of the particles to be counted, the bias will not cancel out of the equation in assumption-based methods. The potential bias must be measured to assure that it does not interfere significantly with the conclusions of the study. In contrast to profile-based methods, the design-based methods are immune to this source of bias.

Finally, conclusions from density analyses of profile-counted sections need to be treated with great caution, because they may lead into the "reference trap" (Braendgaard and Gundersen, 1986), i.e., hydration or dehydration of the specimen or section can result in an erroneously low or high density. When the total number of particles is estimated based on densities alone rather than estimates of total numbers, conclusions may be false. An example for this kind of bias is found in the apparently erroneous concept that the normal aging brain necessarily loses significant numbers of neurons, which has since been re-evaluated (West, 1993; Gallagher et al., 1996; Wickelgren, 1996; Morrison and Hof, 1997; Long et al., 1999).

Frequent mistakes with 3D methods

Mistakes can also be made with the physical or optical disector. One rather simple but consequential mistake is that investigators follow the dubious advice of using methacrylate embedding and sectioning for the optical disector, and then try to recognize particles in poorly stained material which may result in significant biases.

A frequent question is how many samples and particles have to be scored for the optical disector to obtain a reliable estimate. While Gundersen and colleagues (Gundersen et al., 1988a,b; West et al., 1991) recommended sampling of only about 100-200 particles, subsequent calibration and computer simulation studies showed that the number of sampled particles should be significantly higher, at several hundreds (Popken and Farel, 1996; Schmitz and Hof, 2000). Larger sampling numbers also appear to be necessary when tissues contain hugely different particle densities (Benes and Lange, 2001).

Another example is that the section pairs in the physical disector are spaced too close together, resulting in a substantial underestimate because the "lost caps" problem (invisible fragments) can skew the particle count (an effect which becomes more "diluted" and apparently insignificant with increasing distances between sections (Pover and Coggeshall, 1991; Hedreen, 1998a).

With the optical disector, substantial bias can occur when the section is embedded in material which renders differential distribution of particles due to shrinkage or compression of the section in the z-axis, and when a guard space is used which does not take into account such differential distribution of particles (Hatton and von Bartheld, 1999). The resulting bias can be substantial, as discussed above, even though the investigator may have followed earlier recommended protocols "correctly". Guard spaces have to be used with caution, because they may violate the principal rule of design-based counting that "all particles must have the same chance of being counted". This rule was initially not observed (and guard spaces indifferentially recommended), because it was thought that the particle distribution would never change with tissue processing. Since it was shown that differential shrinkage or compression due to tissue processing is widespread in both paraffin sections and methacrylate resin sections (Hatton and von Bartheld, 1999), such shrinkage/compression must be monitored closely when guard spaces are to be used (as discussed above). Finally, Popken and Farel (1996) showed that the orientation of the tissue can affect the amount of bias rendered with the optical disector, for reasons which are not yet understood.

Trends in the use of counting methods

In 1988, Williams and Rakic examined the citation rates for the main profile-based counting method, the Abercrombie correction factor (Abercrombie, 1946). At the time, the paper was cited at a rate of about 100/year (Fig. 2) and had been cited 1,200 times. In the meantime, the citation total has increased to 2,600, and together with the second-most cited profile-based technique, Konigsmark (1970), the total citation number for the two papers combined approaches 3,300. This compares with a citation number for the design-based methods (Gundersen et al., 1988a,b), first introduced to a wider audience in 1988, for a total of about 1,600. It is interesting that the general trend in the 1980s, an increase of the citation of the Abercrombie paper, appears to "flatten" just at the time when the citation rate of the optical disector method begins to rise in the early 1990s, indicating that the new design-based methods may have caused this "dent" or flattening in the general trend of the Abercrombie citation rate (Fig. 2). Since the newer methods have been cited for a much shorter time period, does this mean that the design-based methods are rapidly catching up?

Since only a very small fraction of papers using 2D-counting actually cite the original literature in which the technique was developed, a more reliable measure of usage of the two main counting techniques is obtained by surveys which sample research articles in journals. In 1994, several neuroscience journals were sampled to determine the usage of different counting methods

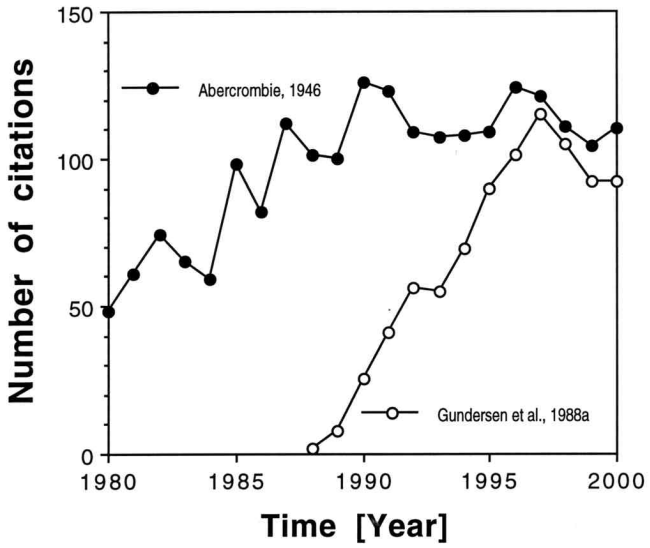

Fig. 2. The number of citations/year according to the Institute for Scientific Information (ISI), Web of KnowledgeSM, is plotted for a 20-year span from 1980-2000 for the two main original sources of counting methods, Abercrombie, 1946 (Anat. Rec. 94:239-247), black dots, and Gundersen et al., 1988a (APMIS 96:857-881), white dots. Note that the gradual increase of citations for the Abercrombie paper in the 1980s comes to a halt in the beginning of the 1990s when citations for the Gundersen et al. paper make their debut.

Table 4. Frequencies in the Use of Counting Techniques in Neuroscience Journals 1994-2001, and one Histology Journal in 2001. Sampled Journals include Brain Research 890-892:262, 2001 (BR), The Journal of Comparative Neurology 429-431:361, 2001 (JCN), The Journal of Neuroscience 21:1-950, 2001 (JN), and Histology & Histopathology, 16:1-881 (HH). The first 100 research articles (71 for HH) published in the year 2001 were sampled for each Journal. Vague estimates include qualitative descriptions of particles such as "most", "few", "some", or "many".

JOURNALS	BR		JCN		JN		HH
	1994* (%)	2001 (%)	1994* (%)	2001 (%)	1994* (%)	2001 (%)	2001 (%)
Tissue sections	33.7	47.0	97.5	89.0	38.3	50.0	87.3
Particle counts	10.1	18.0	32.5	32.0	9.5	21.0	31.0
Vague estimates		10.0		46.0		10.0	33.8

	NEUROSCIENCE JOURNALS							HISTOLOGY	
	(1994*)		(2001)					(2001)	
Journals	Total	%	BR	JCN	JN	Total	%	HH	%
Profile counts	112	75.7	16	25	13	54	84.4	20	90.9
Corrected profile counts	17	11.5	1	4	0	5	7.8	0	0.0
Serial reconstructions	11	7.4	1	1	0	2	3.1	1	4.5
Design-based counts	8	5.4	1	2	0	3	4.7	1	4.5
Total Numbers or %	148	100.0	19	32	13	64	100.0	22	99.9

* Data for 1994 are from Coggeshall and Lekan (1996).

(Coggeshall and Lekan, 1996). In the year when the samples were taken, 1994, design-based methods were used in only about 5% of all papers employing particle counts (Table 4). To determine if usage increased in the past 7 years, I sampled the same neuroscience journals as well as one histology journal for the beginning of 2001 and calculated the percentage of counting methods comparable to the analysis of Coggeshall and Lekan (1996). As can be seen (Table 4), the use of design-based methods is stagnant, even though the use of histological sections and the use of particle counts has increased over the same time period. These data indicate that profile counting will persist and will not be replaced any time soon by design-based methodology.

It is important to keep in mind that counting methods continue to evolve. We should try to optimize and improve our methods with the aim of minimizing biases in counting. We should evaluate carefully the advantages and disadvantages associated with each method, and use the method which is the most appropriate for the task and the situation.

Acknowledgments. I thank Sven Ebbesson for introducing me to profile-based counting methods and Bob Bolender for introducing me to design-based counting methods. I also thank many colleagues for discussions about counting methods, including Peter Clarke, Rick Coggeshall, Paul Farel, Will Hatton, Hans Jorg Gundersen, Jens Nyengaard, Ron Oppenheim, Howard Rind, Glenn Rosen, Mark West, and Rob Williams. I thank Will Hatton, Jens Nyengaard, Glenn Rosen, and Rob Williams for sharing unpublished results. This work was in part supported by NIH grants NS 35931, EY 12841 and RR 15581 (Center of Biomedical Research Excellence).

References

Abercrombie M. (1946). Estimation of nuclear populations from microtome sections. Anat. Rec. 94, 239-247.

Benes F.M. and Lange N. (2001). Two-dimensional versus three-dimensional cell counting: a practical perspective. Trends Neurosci. 24, 11-17.

Bolender R.P., Charleston J., Mottet K., and McCabe J.T. (1991). Quantitative morphology of the nervous system: expanding horizons. Anat. Rec. 231, 407-415.

Braendgaard H. and Gundersen H.J.G. (1986). The impact of recent stereological advances on quantitative studies of the nervous system. J. Neurosci. Meth. 18, 39-78.

Calhoun M.E., Wiederhold K.H., Abramowski D., Phinney A.L., Probst A., Sturchler-Pierrat C., Staufenbiel M., Sommer B. and Jucker M. (1998). Neuron loss in APP transgenic mice. Nature 395, 755-756.

Clarke P.G.H. (1992). How inaccurate is the Abercrombie correction factor for cell counts? Trends Neurosci. 15, 211-212.

Clarke P.G.H. (1993). An unbiased correction factor for cell counts in histological sections. J. Neurosci. Methods 49, 133-140.

Clarke P.G.H. and Oppenheim R.W. (1995). Neuron death in vertebrate development: in vitro methods. Methods Cell Biol. 46, 277-321.

Coggeshall R.E. (1992). A consideration of neural counting methods. Trends Neurosci. 15, 9-13.

Coggeshall R.E. and Lekan H.A. (1996). Methods for determining numbers of cells and synapses: a case for more uniform standards of review. J. Comp. Neurol. 364, 6-15.

Coggeshall R.E., La Forte R. and Klein C.M. (1990). Calibration of methods for determining numbers of dorsal root ganglion cells. J. Neurosci. Methods 35, 187-194.

Cruz-Orive L.M. (1994). Towards a more objective biology. Neurobiol. Aging 15, 377-378.

Gallagher M., Landfield P.W., McEwen B., Meaney M.J., Rapp P.R.,

Sapolsky R. and West M.J. (1996). Hippocampal neurodegeneration in aging. Science 274, 484-485.

Gaule J. and Lewin T. (1896). Ueber die Zahlen der Nervenfasern und Ganglienzellen in den Spinalganglien des Kaninchens. Centralbl. f. Physiol. 10, 437-440, 465-471.

Geuna S. (2000). Appreciating the difference between design-based and model-based sampling strategies in quantitative morphology of the nervous system. J. Comp. Neurol. 427, 333-339.

Guillery R.W. and Herrup K. (1997) Quantification without pontification: choosing a method for counting objects in sectioned tissues. J. Comp. Neurol. 386, 2-7.

Gundersen H.J.G., Bagger P., Bendtsen T.F., Evans S.M., Korbo L., Marcussen N., Møller A., Nielsen K., Nyengaard J.R., Pakkenberg B., Sørensen F.B., Vesterby A. and West M.J. (1988a). The new stereological tools: disector, fractionator, nucleator and point sampled intercepts and their use in pathological research and diagnosis. Acta Pathol. Microbiol. Immunol. Scand. 96, 857-881.

Gundersen H.J.G., Bendtsen T.F., Korbo L., Marcussen N., Møller A., Nielsen K., Nyengaard J.R., Pakkenberg B., Sørensen F.B., Vesterby A. and West M.J. (1988b). Some new, simple and efficient stereological methods and their use in pathological research and diagnosis. Acta Pathol. Microbiol. Immunol. Scand. 96, 379-394.

Guntinas-Lichius O., Mockenhaupt J., Stennert E. and Neiss W.F. (1993). Simplified nerve cell counting in the rat brainstem with the physical disector using a drawing-microscope. J. Microsc. 172, 177-180.

Harding A.J., Halliday G.M. and Cullen K. (1994). Practical considerations for the use of the optical disector in estimating neuronal number. J. Neurosci. Methods 51, 83-89.

Hatton W.J. and von Bartheld C.S. (1999). Analysis of cell death in the trochlear nucleus of chick embryos: Calibration of the optical disector counting technique reveals systematic bias. J. Comp. Neurol. 409, 169-186.

Hedreen J.C. (1998a). Lost caps in histological counting methods. Anat. Rec. 250, 366-372.

Hedreen J.C. (1998b). What was wrong with the Abercrombie and empirical cell counting methods? A review. Anat. Rec. 250, 373-380.

Hedreen J.C. and Vonsattel J.P. (1997). Centenary of Gaule and Lewin, pioneers in cell counting methodology. J. Microsc. 187, 201-203

Heinsen H., Arzberger T. and Schmitz C. (2000). Celloidin mounting (embedding without infiltration) - a new, simple and reliable method for producing serial sections of high thickness through complete human brains and its application to stereological and immuno-histochemical investigations. J. Chem. Neuroanat. 20, 49-59.

Howard C.V. and Reed M.G. (1998). Unbiased stereology. Three-dimensional measurement in microscopy. Springer. New York.

Insausti A.M., Megias M., Crespo D., Cruz-Orive L.M., Dierssen M., Vallina I.F., Insausti R., Florez J., Vallina T.F. and Vallina T.F. (1998). Hippocampal volume and neuronal number in Ts65Dn mice: a murine model of Down syndrome. Neurosci. Lett. 253, 175-178.

Konigsmark B.W. (1970). Methods for the counting of neurons. In: Contemporary research methods in neuroanatomy. Nauta W.J.H., and Ebbesson S.O.E. (eds). Springer. New York. pp 315-340.

Korkmaz A., Ciftci N., Bosnak M. and Agar E. (2000). A simplified application of systematic field sampling and low-cost video recording set-up for viewing disector pairs - exemplified in the rat cochlear nucleus. J. Microsc. 200, 269-276.

Long J.M., Mouton P.R., Jucker M. and Ingram D.K. (1999). What counts in brain aging? Design-based stereological analysis of cell number. J. Gerontol. A Biol. Sci. Med. Sci. 54, B407-417.

Mayhew T.M. (1992). A review of recent advances in stereology for quantifying neural structure. J. Neurocytol. 21, 313-328.

Mayhew T.M. and Gundersen H.J. (1996). 'If you assume, you can make an ass out of u and me': a decade of the disector for stereological counting of particles in 3D space. J. Anat. 188, 1-15.

Morrison J.H. and Hof P.R. (1997). Life and death of neurons in the aging brain. Science 278, 412-419.

Popken G.J. and Farel P.B. (1996). Reliability and validity of the physical disector method for estimating neuron number. J. Neurobiol. 31, 166-174.

Popken G.J. and Farel P.B. (1997). Sensory neuron number in neonatal and adult rats estimated by means of stereologic and profile-based methods. J. Comp. Neurol. 386, 8-15.

Pover C.M. and Coggeshall R.E. (1991). Verification of the disector method for counting neurons, with comments on the empirical method. Anat. Rec. 231, 573-578.

Saper C.B. (1996). Any way you cut it: a new journal policy for the use of unbiased counting methods. J. Comp. Neurol. 364, 5.

Saper C.B. (1999). Unbiased stereology: three-dimensional measurement in microscopy. Book review. Trends Neurosci. 22, 94-95.

Schmalbruch H. (1987). The number of neurons in dorsal root ganglia L4-L6 of the rat. Anat. Rec. 219, 315-322.

Schmitz C. and Hof P.R. (2000). Recommendations for straightforward and rigorous methods of counting neurons based on a computer simulation approach. J. Chem. Neuroanat. 20, 93-114.

Schmitz C., Korr H. and Heinsen H. (1999). Design-based counting techniques: the real problems. Trends Neurosci. 22, 345-346.

von Bartheld C.S. (1999). Systematic bias in an 'unbiased' neuronal counting technique. Anat. Rec. (New Anat.) 257, 119-120.

von Bartheld C.S. (2001). Comparison of 2D- and 3D-counting: the need for calibration and common sense. Trends Neurosci. 24, 504-506.

West M.J. (1993). New stereological methods for counting neurons. Neurobiol. Aging 14, 275-285.

West M.J. (1999). Stereological methods for estimating the total number of neurons and synapses: issues of precision and bias. Trends Neurosci. 22, 51-61.

West M.J., Slomianka L. and Gundersen H.J. (1991). Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. Anat. Rec. 231, 482-497.

Wickelgren I. (1996). Is hippocampal cell death a myth? Science 271, 1229-1230.

Williams R.W. and Rakic P. (1988). Three-dimensional counting: accurate and direct method to estimate numbers of cells in sectioned material. J. Comp. Neurol. 278, 344-352.

Accepted January 2, 2002