An Introduction to Stereology with Applications to the Glomerulus

Kevin V. Lemley

Department of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Abstract

Background: Stereology is the science of inferring quantitative features of 3-dimensional structures from lower dimensional samples of those structures (probes). It is a statistical discipline and therefore may seem intimidating to many potential users. Without a proper understanding of its principles, though, errors may be made in the quantitative reporting of structural research results. ***Summary:*** This review article attempts to explain and justify the basic principles of stereology as applied to the glomerulus in a simple and accessible way. A few common errors in application are described. The strengths and weaknesses of "biased" (model-based) stereology are described as well as the basics of design-based ("unbiased") stereology. ***Key Messages:*** Stereology is a useful body of theory and practices when quantitation of structural histological features of the glomerulus is desired.

© 2021 The Author(s).
Published by S. Karger AG, Basel

Introduction

Although they are often used interchangeably, the terms "stereology" and "morphometry" in fact have different histories and meanings. Another synonymous phrase which is often used, and a bit more transparent but less specific, is "quantitative morphology." The older term "morphometry" (1860 per the Oxford English Dictionary) refers to the measurement (μέτρια) of form (μορφή), any form. The term "stereology" is of a much later vintage (1963) having been synthesized by Hennig and Elias for the first International Congress on that subject and referring to the study (-λογία) of a solid (i.e., a 3-dimensional object, στερεός). More precisely, the Oxford English Dictionary defines stereology as "the science of reconstruction of 3-dimensional structures from 2-dimensional sections of them." Stereology is a statistical discipline, so "reconstruction" does not mean something like performing exhaustive serial sections, imaging them, and then bringing the section images back together in a 3-dimensional whole. Rather, it means estimating 3-dimensional structural parameters of an object of interest from lower dimensional probes (and more generally 1- or 2-dimensional probes) of that object.

Although standard tissue stereology has not been supplanted yet, modern techniques of tissue clearing and 3-dimensional imaging [1] and non-light-based 3-dimensional imaging (MRI) [2] provide a glimpse into a future with true 3-dimensional imaging. These methods will not be considered in this review.

Because stereology is a statistical discipline, a system of random sampling is of paramount importance. Given the nature of tissue sectioning, a hierarchical system of sampling is needed. To keep the presentation as simple as possible, considerations of sampling will largely be ig-

Correspondence to:
Kevin V. Lemley, klemley@chla.usc.edu

nored. We shall concentrate instead on the basics of what is measured in the lower dimensional probes and how the 3-dimensional quantitative features of the target object are deduced from those measurements.

Densities versus Total Values

The quantitative descriptions of the structural features of objects of interest can be divided into 2 major camps. The first, and the type of feature more often measured directly, is a *density*, viz., the ratio of a measured feature to its containing reference space. An example would be the *volume density* of a definable feature with respect to its reference space, such as the fractional density of the mesangium with respect to the glomerular tuft volume, for example, 20% of the tuft volume is mesangium. How this is estimated will be discussed below. An associated *total feature* of the glomerulus would be the absolute mesangial volume per glomerulus, which may be calculated as the product of the mesangial volume density by the glomerular tuft volume, for example, $0.2 \times 2 \times 10^6 \ \mu m^3 = 0.4 \times 10^6 \ \mu m^3$.

Fractional densities are typically expressed with respect to a reference volume, but the numerator may be a number, a length, an area (surface), or a volume. These densities are typically represented symbolically as N_V, L_V, S_V, and V_V. For example, N_V is the number of particles per unit reference volume (say, podocyte nuclei per glomerular tuft volume), or L_V the length of a (lineal) feature per unit reference volume (say, the filtration slit length per glomerular tuft volume). As we will see below, the reference space may be <3-dimensional, for example, an areal density like A_A (say, the fraction of the glomerular tuft cross section which is mesangium) or L_A (say, the length of the GBM per glomerular tuft cross-sectional area). Areal densities are the usual form in which structural features are measured in histological sections, which are idealized as 2-dimensional. Note here the use of both A and S to express (surface) area in the densities; A is more commonly used as the symbol for the area reference space.

Unless a structure is completely serially sectioned and reconstructed, quantitative structural features are estimated from samples of the total structure of interest. This is analogous to measuring some feature of a population of individuals (e.g., their average height) in a sample of the population and estimating the population mean height from the arithmetic mean of the heights in the sample. Not surprisingly, the latter statistic (the sample mean) is called an estimator of the population feature. As with population sampling, with stereology such sampling can lead to bias in the estimate. Unlike population sampling, with stereological sampling, the bias may in some cases be inherent to the measurement procedure itself.

Let us start with some examples of biased and unbiased estimators. The easiest examples come from descriptive statistics: the sample mean, $\bar{x}$, and the sample variance, s_n^2. The former, calculated as the arithmetic average of all the x values in the sample, is an unbiased estimator of the underlying population mean, μ. The latter, calculated as the mean of the squared differences of those x values from $\bar{x}$, is a biased estimator of the population variance, σ^2. The sample variance tends to slightly underestimate the actual population variance. Fortunately, an easy fix (due to Bessel) corrects the bias. Rather than dividing the sum of squared differences by the number of measurements in the sample (n), this sum is divided by the slightly smaller number, $n - 1$, yielding an unbiased estimator, s_{n-1}^2, of σ^2.

Similarly, some stereological estimators are inherently unbiased, while others are inherently biased. The volumetric density (V_V) of one component with respect to the body as a whole (the reference volume) can be estimated without bias by measuring the areal density (A_A) of that same component with respect to the area of a 2-dimensional section through the body of interest. For example, if you have a specific stain for podocyte cytoplasm (e.g., GLEPP1), measuring the percentage of the total cross-sectional glomerular tuft area that is GLEPP1-positive will give an unbiased estimator of the percentage of the tuft volume that is made up of podocyte cytoplasm. This is the principle of Delesse (1848), originally derived to estimate the mineral composition of rocks from areas of the particular mineral measured in the cut faces of those rocks relative to the entire area of the cut face. Thicker sections tend to increase the apparent area over the idealized 2-dimensional area due to overprojection at the edges (Holmes effect); this is minimized by using constant and thin sectioning.

The length density per reference volume (L_V) of a linear feature can likewise be estimated without bias by measurements performed in a 2-dimensional probe of the reference space. In this case, what is measured is the number of intersections of the linear feature per reference area (N_A) and the formula is only a little more complicated than that for the volume density above: $L_V = 2 \times N_A$. Think of estimating the total length of the podocyte filtration slits per glomerulus. If you have an estimate of the mean filtration slit length density (L_V) and of the reference (tuft) volume, V, then the estimated total filtration slit length per glomerulus (L) is just $L = L_V \times V = 2 \times N_A \times V$.

Fig. 1. The number of particle profiles (transects) in the histological section is a function of the volumetric density of the particles (N_V), but also their sizes and the thickness of the section (increasing with all of these). The thinner section (**a**) contains fewer profiles than the thicker section (**b**). From [3].
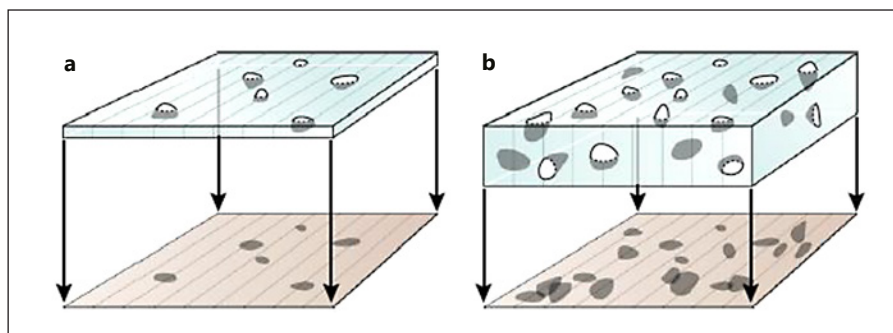
Unfortunately, not all the structural features we would like to estimate are amenable to such unbiased stereological measurement. For example, the average number of discrete particles in a reference volume (e.g., podocytes in glomeruli) or the number density of such particles per reference volume (N_V) cannot be estimated from measurements in single histological sections without making simplifying assumptions about the underlying geometry. This is what is meant by model-based stereology (often referred to as "biased" stereology). An entirely different approach to such estimation is discussed below (design-based stereology).

For example, the number of podocyte nuclear transects (one typically counts podocytes from their nuclei, given their extended cytoplasm) per tuft cross-sectional area (N_A) is a function of N_V, but also of the section thickness and the podocyte nuclear size (increasing with each of those), as is shown in Figure 1. A number of "correction factors" have been developed to account for these dependencies. The simplest (and oldest) is due to Wicksell (1925): $N_V = N_A/D$, where D is the mean (caliper) diameter of the particle (its apparent profile diameter when viewed from all possible perspective angles – shown in Fig. 2). A more commonly used and more complex formula is due to Weibel and Gomez [4] wherein $N_V = (\kappa/\beta) \times (N_A)^{3/2}/(A_A)^{1/2}$ and κ and β are, respectively, a size distribution coefficient and a shape factor (e.g., $\beta = 1.382$ for a spherical particle), N_A is the number of nuclear transects per cross-sectional area, and A_A is the areal density of the podocyte nuclear transects relative to tuft cross-sectional area. The thick and thin method [5, 6] empirically establishes the effects of particle size and section thickness, by measuring N_A at 2 (or more) different section thicknesses and deriving N_V from the difference in apparent N_A measured at the 2 thicknesses. An adaptation to single sections has also been proposed [7].
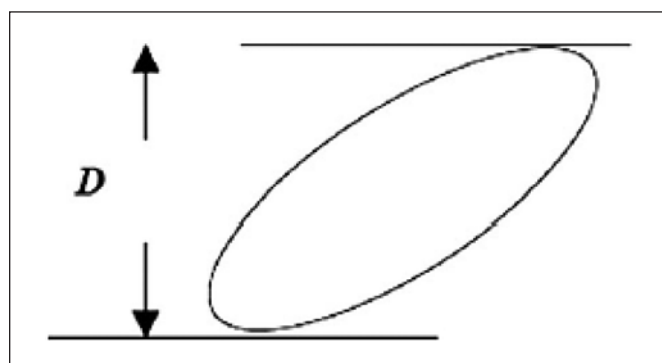

Fig. 2. The so-called caliper diameter of a particle (other than of a sphere) depends on the orientation of the perspective angle. The average caliper diameter is the average over all orientations and represents the statistical 2-dimensional profile of the particle. From [3].

Why do people even use a biased method for stereological estimates? First, the fundamental measurements with these methods are typically made in a single histological section, which is much easier than the unbiased methods (described below). As in statistics, parametric (model-based) methods are often more powerful than nonparametric ("design-based") methods. This is the payback for bringing external information (about the population parameters) into the mix. Finally, the bias may be small compared to the differences in glomerular parameters we are interested in testing [8].

Design-Based versus Biased (Model-Based) Stereology

In the 1980s, Gundersen and others developed a body of estimation methods that did not lead to the problems outlined above. I shall illustrate a single technique, the

disector, for unbiased estimation of the density N_V of particles (e.g., podocytes) with respect to a reference space (e.g., the glomerular tuft). This was first described in 1984 by the pseudonymous stereologist, DC Sterio [9], as a solution to the above-mentioned problem.

The disector principle is based on the fact that, if one cuts sections serially through an object, any embedded particle *uniquely* first appears in one section, regardless of its size or shape. To establish the first appearance of the particle, a pair of consecutive (although not necessarily adjacent) sections, a *reference* section and a *look-up* section, separated by a given height, h, is compared. This height must be relatively small compared to the particle size (caliper diameter) so that no particles are lost between the disector pair. Any particle absent from the first section and present in the second has clearly "appeared" between the former and latter sections. That particle can therefore be assigned to the disector volume corresponding to the section area A times h. Thus, $N_V = Q/(A \times h)$, where Q is the number of particles that appeared in the disector volume. (To make this system more efficient, both appearances and disappearances may be counted in the section-pair, since each is a unique occurrence, effectively sampling twice the volume at once.) The disector pairs do not have to represent physical histological sections. An "optical" version of the (physical) disector can be created by scanning through a Z-stack with the thin focal plane of confocal microscopy and counting particles when they first come into focus [10]. This approach has largely supplanted the more tedious physical disector.

The above description of the disector method makes clear that "unbiased" stereological methods may be quite labor-intensive. Other design-based methods for directly estimating total numbers of particles within a reference space (the fractionator) have been described [10]. Image analysis using machine learning may someday automate some of the more time-consuming tasks involved in stereological studies.

The Measurements Themselves

There are basically 3 classes of measurements made on the 2-dimensional histological images: counting numbers of objects, measuring lengths of "1-dimensional" objects (e.g., the transect of the GBM in the section plane), and measuring areas of "2-dimensional" objects (e.g., glomerular tuft cross sections). The latter 2 may be measured directly or estimated with a sampling strategy. A histological section is of course not really a 2-dimensional ob-
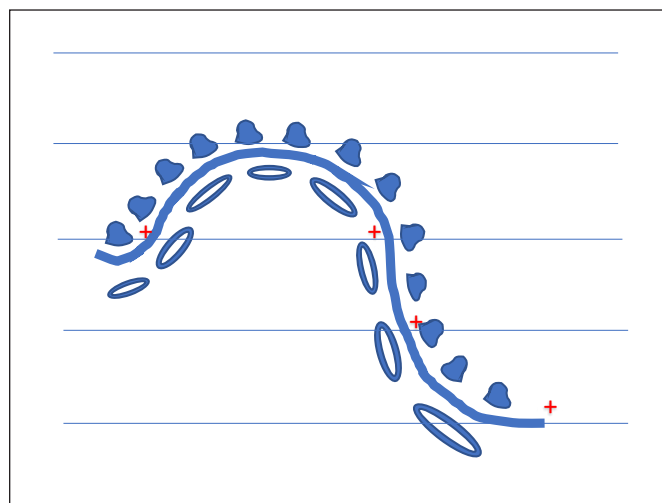

Fig. 3. The length density (L_A) of the GBM (thick blue line) is estimated by the number of intersections with a raster of test line probes (thin blue lines). The red +s show the locations of the 4 intersects in the image. L_A is estimated without bias by $2 \times I \times d$, where I is the number of intersects in the test area and d is the raster spacing. The surface density of GBM per tuft area (S_V) may be estimated from $S_V = (4/\pi) \times L_A$.

ject – it has a thickness. The focal plane of the microscope has a depth of a few microns, although it appears as a 2-dimensional body in the histological image. Similarly, the GBM has an obvious thickness (width), which we ignore when we are measuring its length in a histological section.

The direct length and area measurements are usually done by computerized planimetry, a great advance from earlier times when areas were estimated by cutting up printed pictures of the images and weighing them! A cursor is used to trace the GBM or outline the tuft area and a more or less precise value for the length or area is obtained. The open-source ImageJ is a commonly used planimetry program available from the NIH (https://imagej.nih.gov/ij/download.html).

The sampling approach differs in each application but basically involves that fact that the chance that a given probe intersects the object of interest is proportional to the length or area of that object. Measuring GBM length might be done by laying a raster of horizontal lines over an image (shown in Fig. 3) and counting the number of discrete intersections of the raster lines with the GBM. The denser the raster, the more precise the estimate (and the greater the effort). If the raster spacing is d (e.g., 3 μm), and the number of intersections of the raster with the
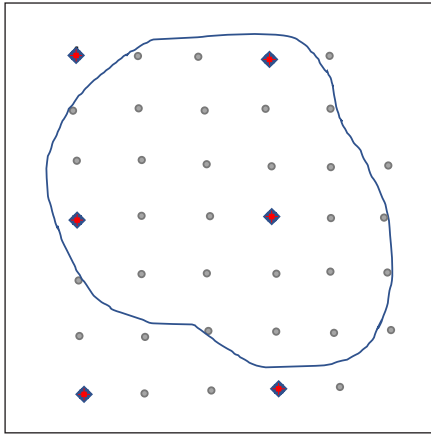
Fig. 4. Point-counting grid showing denser sampling points (filled circles) and less dense red diamonds. These are in the relation of 9 to 1 (diamonds count as points also) so that 1 diamond represents the area of 9 points. The outlined area contains 27 points/diamonds.

Fig. 5. Changes in the volume of the reference space (glomerular tuft) cause changes in apparent podocyte number (as reflected in nuclear transects per tuft cross section) without a change in total glomerular podocyte number. Each reference space contains the same number of particles. Each section (1 and 2) intersects less particles in the larger space (**b**) than in the smaller space (**a**). From [3].

GBM is I, then the estimated total GBM length will be: $(\pi/2) \times I \times d$.

Area estimation is often accomplished by point counting. A 2-dimensional grid of points is projected onto the image and the numbers of points falling on the object of interest and on the reference space are counted. As above, the number of intersections is proportional to the area of the object, so the ratio of points on the object of interest to points on the reference space is an estimator for A_A. Each point has an area value equal to the size of the square of which it makes up one corner. To keep the point counting to a minimum when measuring the areal density of a "rare" object in a larger reference space, denser and less dense grids may be superimposed together, say a denser system of points and the upper left "point" of each 3×3 box of points is replaced by a diamond. The diamonds may be used to estimate the area of the larger space and the points + diamonds of the smaller space (each diamond is worth 9 points of area), as illustrated in Figure 4.

"Do more less well!" [11] or go where the variance is. The former statement by Ewald Weibel reflects the fact that often too much effort is put in to overly precise measurements at the level of the microscopic field of view, rather than spending time/effort sampling more of the kidney or sampling more individuals. The total measured variance is the sum of variances at all sampling levels: from biological variance at the level of the individual subject to variance due to various levels of tissue sampling and the final measurements. We know that in the statisti-
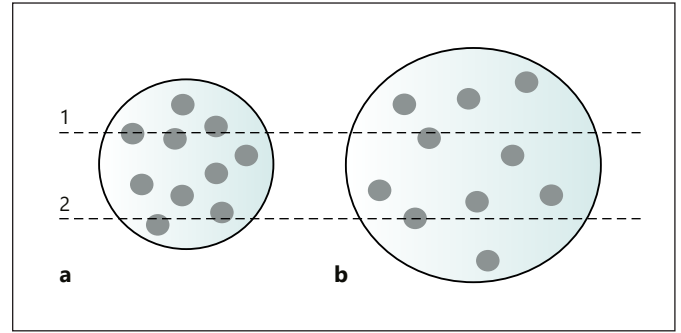
cal estimation theory, the standard error of the mean (which sets the precision of our estimate of a population parameter) decreases inversely with the square root of sample size. "Do more less well!" means greater effort should be placed at the sampling level (individual, tissue, field of view) where the variance is greatest, as added effort there will increase precision the most. As a rule of thumb, tissue level point counting using 100–200 total points often gives optimal precision/effort.

The Reference Trap [12]

Another source of error in using morphometry to quantify structural features of the glomerulus arises from confusion regarding the density versus absolute number distinction mentioned above. Generally speaking, the objects of interest are total values, such as the total number of podocytes per glomerulus. This may be deduced from the product of the podocyte volume density, N_V, and the glomerular volume, V. What is often reported in publications as glomerular "podocyte number," though, is not even the density N_V, but the number of podocyte nuclear transects per glomerular cross section. It should be intuitively clear that a mouse glomerulus does not have only 10 podocytes! The number of transects per cross section is a function of N_V, nuclear size and section thickness (Fig. 1). It is important not to forget what happens to the *reference space* (hence the trap). For example, if whatever experimental treatment we are using causes glomerular hypertrophy, but no change in podocyte number, N_V will
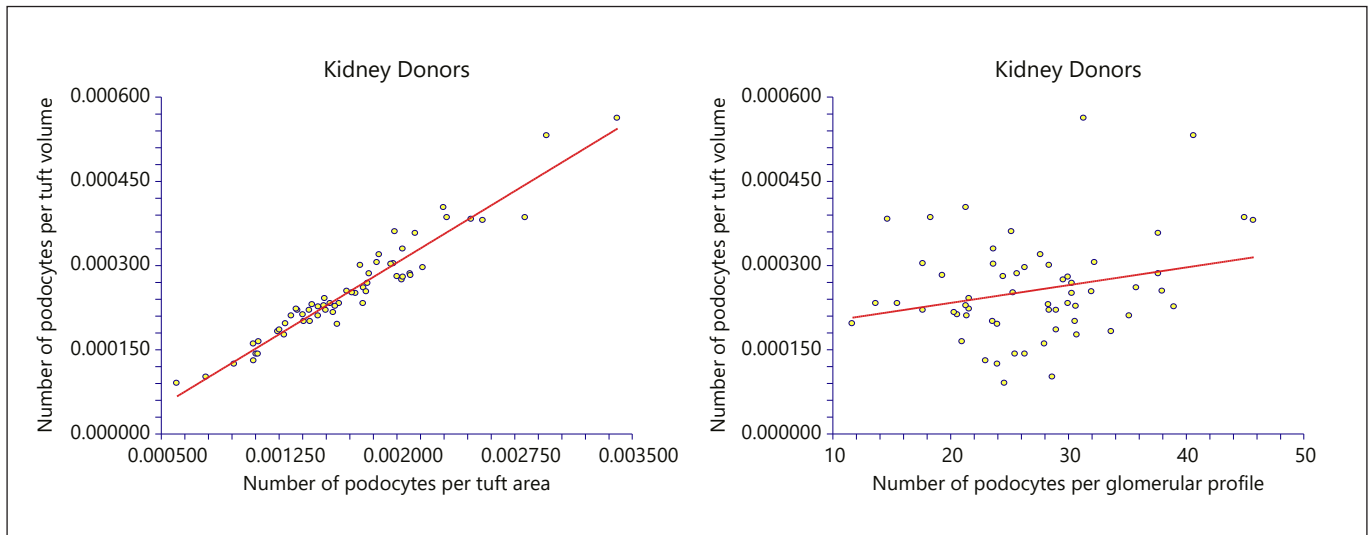
Fig. 6. Relationship between the number of podocyte transects per glomerular tuft area (N_A) and the number of podocytes per glomerular tuft volume (N_V) and between N_V and the number of podocyte transects per glomerular cross section (N_P) in kidney transplant donors. The relationship of the former pair is clearly much stronger ($r = 0.96$) than between the latter pair ($r = 0.25$). N_V was calculated using the Weibel-Gomez method on electron micrographs. From [8].

decrease, which will decrease the number of podocyte nuclei seen in a glomerular cross section (shown in Fig. 5), often leading to the interpretation of a decreased total glomerular podocyte number. Even considering podocyte number per glomerular cross-sectional area N_A (despite its inadequacies) gives a better representation of N_V than the number of podocyte nuclei per glomerular cross section (shown in Fig. 6 [8]).

Weibel versus Cavalieri

We have mentioned methods for estimating densities from histological images, but not how we measure the volume of the reference space (typically the glomerular tuft) in order to turn the densities into total numbers per glomerulus. Here again we have the option of model-based or design-based methods. The most frequently used model-based method for estimating average glomerular tuft volume, V, is due to Weibel [13]: $V = (\beta/\kappa) \times (A)^{3/2}$, where κ and β are the same size and shape factors from above and A is the mean tuft cross-sectional area. The "design-based" method dates from 1635 (Cavalieri) and is similar to the disector principle above: the volume of a single section is approximated by the product of the area of the object of interest in that section (A_i) times the distance between that and the next section (T). The sum

of all the volumes intersecting the object of interest is the Cavalieri estimate of the object's volume, $V = \Sigma (A_i \times T) = (\Sigma A_i) \times T$.

Avoid Double-Dipping

For statistical reasons, it is important that each potential sample space is sampled with equal probability. For example, in sampling glomeruli for an estimate of the volume density of podocytes, if the sampled histological sections are 20 μm apart (for humans, say), it is likely that some glomeruli will be sampled multiple times, while others would be sampled only once. Similarly, using the Weibel formula to estimate glomerular volume from average cross-sectional area, it is important to examine sections far enough apart from each other that transecting the same glomerulus more than once is impossible or at least unlikely.

Simpson's Paradox

Although it is not strictly a feature of stereology, stereological studies may present a good illustration of Simpson's paradox (namely, an apparent association between variables that exists in a larger group, but that dis-
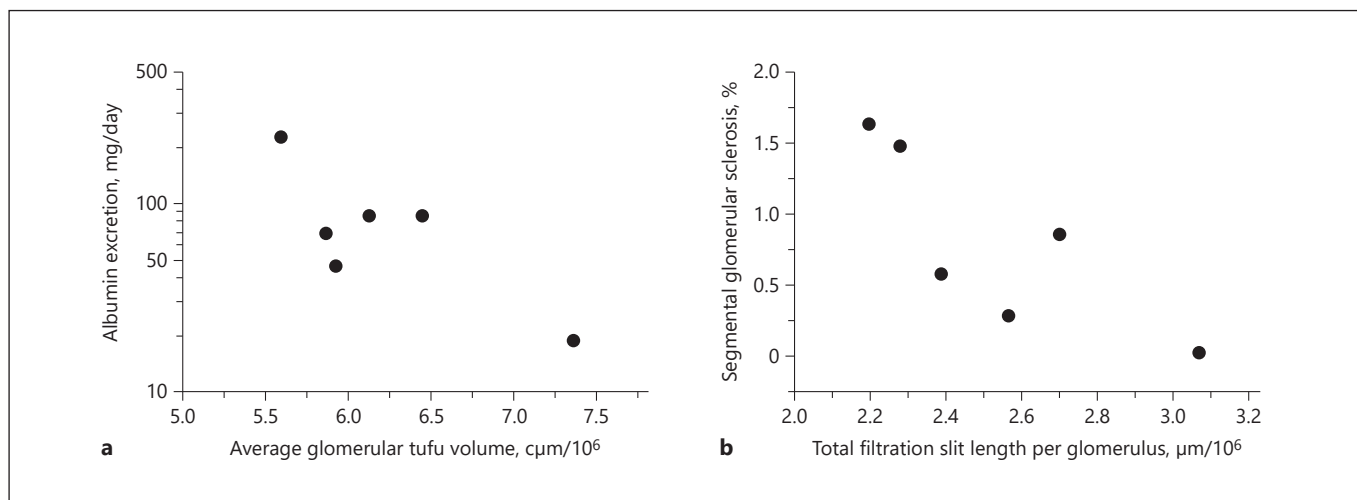
Fig. 7. Despite the fact that the group of nephrectomized rats with glomerular hypertrophy had greater albuminuria than the sham controls, within this group urinary albumin excretion actually decreased significantly with increasing glomerular tuft volume ($r = -0.814$, $p < 0.05$), left panel. Incidence of segmental glomerular sclerosis decreased significantly with total filtration slit length per glomerulus ($r = -0.818$, $p < 0.05$), right panel. From [14].

appears or even reverses when the group is divided into subpopulations). In a stereological study [14] in Sprague-Dawley rats, partial renal ablation (Nx) led to hypertension (SBP 220 ± 21 vs. 128 ± 14 mm Hg), albuminuria (89 ± 72 vs. 11 ± 11 mg/day), and glomerular hypertrophy (tuft volume 3.475 ± 0.517 vs. 1.422 ± 0.116 × 10^6 μm^3) at 8 weeks compared to sham-operated rats. The incidence of segmental glomerular sclerosis was also significantly greater in the Nx rats (0.80 ± 0.65 vs. 0.10 ± 0.17%). The usual interpretation would be that glomerular hypertrophy in the Nx rats was responsible for the albuminuria and sclerosis. However, when the Nx rats were examined as a group, the opposite seemed to be the case (as shown in Fig. 7). Albuminuria actually decreased significantly with increasing glomerular volume and the rate of segmental sclerosis decreased with increasing total filtration slit length. A possible explanation of this reversal is the effect of increasing the total filtration slit length (and thereby filtration slit area) on the intraglomerular hydraulic pressure. The greater the ultrafiltration capacity of the glomerulus (determined largely by the filtration slit area), the lower the intraglomerular pressure needed to achieve the compensatory increased single-nephron glomerular filtration rate.

Conclusion

This has been a review of the basic principles and some applications of stereology to glomerular morphometry, meant to give a practical feeling for the field, but not to give the detailed knowledge needed to apply stereology to structural studies in nephrology. A number of excellent reviews with more details on specific methods are available [10, 15].

Conflict of Interest Statement

The author has no conflicts of interest to declare.

Funding Sources

There are no relevant funding sources for this work.

Author Contributions

K.V.L. conceived this work, wrote it, and revised it.

Lemley

References

1 Puelles VG, Combes AN, Bertram JF. Clearly imaging and quantifying the kidney in 3D. Kidney Int. 2021. Epub ahead of print.

2 Bennett KM, Bertram JF, Beeman SC, Gretz N. The emerging role of MRI in quantitative renal glomerular morphology. Am J Physiol Renal Physiol. 2013;304:F1252–7.

3 Lemley KV, Bertram JF, Nicholas SB, White K. Estimation of glomerular podocyte number: a selection of valid methods. J Am Soc Nephrol. 2013;24(8):1193–202.

4 Weibel ER, Gomez DM. A principle for counting tissue structures on random sections. J Appl Physiol. 1962;17:343–8.

5 Loud AV, Anversa P, Giacomelli F, Wiener J. Absolute morphometric study of myocardial hypertrophy in experimental hypertension. I. Determination of myocyte size. Lab Invest. 1978;38:586–96.

6 Sanden SK, Wiggins JE, Goyal M, Riggs LK, Wiggins RC. Evaluation of a thick and thin section method for estimation of podocyte number, glomerular volume, and glomerular volume per podocyte in rat kidney with Wilms' tumor-1 protein used as a podocyte nuclear marker. J Am Soc Nephrol. 2003;14: 2484–93.

7 Ventakareddy M, Wang S, Patel S, Wickman L, Nishizono R, Chowdhury M, et al. Estimating podocyte number and density using a single histologic section. J Am Soc Nephrol. 2014;25:1118–29.

8 Lemley KV. Simplification of the Weibel-Gomez method for estimating podocyte number. In: Capasso V, Aletti G, Micheletti G, editors. Stereology and image analysis. ECS10 – Proceedings of the 10th European Congress of ISS. Bologna: Esculapio Pub.; 2009. conf-pap.

9 Sterio DC. The unbiased estimation of number and sizes of arbitrary particles using the disector. J Microsc. 1984;134(2):127–36.

10 Puelles VG, Douglas-Denton RN, Cullen-McEwen L, McNamara BJ, Salih F, Li J, et al. Design-based stereological methods for estimating numbers of glomerular podocytes. Ann Anat. 2014;196:48–56.

11 Gundersen HJ, Østerby R. Optimizing sampling efficiency of stereological studies in biology: or "do more less well"! J Microsc. 1981; 121(1):65–73.

12 Brændgaard H, Gundersen HJG. The impact of recent stereological advances on quantitative studies of the nervous system. J Neurosci Meth. 1986;18:39–78.

13 Weibel ER. Stereological methods: practical methods of biological morphometry. London: Academic Press; 1979. Vol. 1; p. 44–5.

14 Tenschert S, Elger M, Lemley KV. Glomerular hypertrophy after subtotal nephrectomy: relationship to early glomerular injury. Virchows Arch. 1995;426:509–17.

15 Howard CV, Reed MG. Unbiased stereology: three-dimensional measurement in microscopy. Oxford: Bios Scientific Publishers Ltd/ Springer; 1998.