Automated systematic random sampling and Cavalieri stereology of histologic sections demonstrating acute tubular necrosis after cardiac arrest and cardiopulmonary resuscitation in the mouse.

**Authors:** Rumie Wakasaki, Mahaba Eiwaz, Nicholas McClellan, Katsuyuki Matsushita, Kirsti Golgotiu and Michael Hutchens

**DOI:** 10.14670/HH-18-012
**Article type:** ORIGINAL ARTICLE
**Accepted:** 2018-06-14
**Epub ahead of print:** 2018-06-14

This article has been peer reviewed and published immediately upon acceptance. Articles in "Histology and Histopathology" are listed in Pubmed. Pre-print author’s version
Automated systematic random sampling and Cavalieri stereology of histologic sections demonstrating acute tubular necrosis after cardiac arrest and cardiopulmonary resuscitation in the mouse.

Short title: Automated Cavalieri stereology

Rumie Wakasaki†

Mahaba Eiwaz†

Nicholas McClellan†

Katsuyuki Matsushita†

Kirsti Golgotiu†

Michael Hutchens†,*

†Anesthesiology & Perioperative Medicine
Oregon Health & Science University
Portland, OR

* Portland Veterans Affairs Medical Center
3710 SW US Veterans Hospital Road
Portland OR

*Correspondence should be addressed to: Michael P. Hutchens, M.D., M.A., Research & Development Division, R&D 5 3710 SW US Veterans Hospital Road, Portland OR 9723
hutchenm@ohsu.edu

Acknowledgements: NIDDK K08DK090754 to MPH. This material is the result of work (by MPH) which was supported with resources and the use of facilities at the Portland Veterans Affairs Medical Center. The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

Keywords: stereology, cavalieri principle, acute kidney injury, tubular necrosis, morphometry.

Word Counts: Abstract: 249 Manuscript: 2846
Abstract

**Introduction:** A technical challenge in translational models of kidney injury is determination of the extent of cell death. Histologic sections are commonly analyzed by area morphometry or unbiased stereology, but stereology requires specialized equipment. Therefore, a challenge to rigorous quantification would be addressed by an unbiased stereology tool with reduced equipment dependence. We hypothesized that it would be feasible to build a novel software component which would facilitate unbiased stereologic quantification on scanned slides, and that unbiased stereology would demonstrate greater precision and decreased bias compared with 2D morphometry.

**Material and Methods:** We developed a macro for the widely used image analysis program, Image J, and performed cardiac arrest with cardiopulmonary resuscitation (CA/CPR, a model of acute cardiorenal syndrome) in mice. Fluorojade-B stained kidney sections were analyzed using three methods to quantify cell death: gold standard stereology using a controlled stage and commercially-available software, unbiased stereology using the novel ImageJ macro, and quantitative 2D morphometry also using the novel macro.

**Results:** There was strong agreement between both methods of unbiased stereology (bias -0.004±0.006 with 95% limits of agreement -0.015 to 0.007). 2D morphometry demonstrated poor agreement and significant bias compared to either method of unbiased stereology.

**Conclusion:** Unbiased stereology is facilitated by a novel macro for ImageJ and results agree with those obtained using gold-standard methods. Automated 2D morphometry overestimated tubular epithelial cell death and correlated modestly with values obtained from unbiased stereology. These results support widespread use of unbiased stereology for analysis of histologic outcomes of injury models.
Introduction

A critical technical challenge in translational models of organ injury is determination of the extent of cell injury or death. Acute kidney injury (AKI) studies which quantify renal cell death have typically used one of three methods: blinded scoring of hematoxylin and eosin stained sections by a renal pathologist, quantification of area or total intensity of specific staining (e.g., morphometry), or unbiased stereology of specific staining. Subjective scoring of 2-dimensional sections may suffer from bias, therefore quantitative measures are desirable. However, total fluorescence- or area-based (2D) morphometric quantification may suffer from bias due to transformation from 3 to 2 dimensions. (Nyengaard, 1999; Marcos et al., 2015; Brown, 2017) Unbiased stereology was devised in the 19th century to eliminate this bias, however performing unbiased stereologic quantification on microscopic sections is frequently conducted with expensive and cumbersome specialized equipment (a computer-controlled, 3-axis moving microscope stage (Gundersen, 1986; Nyengaard, 1999). Therefore a critical problem in evaluating translational organ injury models would be solved by development of a low-cost, technology-independent, unbiased stereology quantification tool. We hypothesized that it would be possible to develop a platform-independent novel software component which would allow both area morphometry and unbiased stereologic quantification of cell death in an AKI model. We further hypothesized that unbiased stereology results would demonstrate greater precision and reduced bias compared with quantitative 2D morphometry. To test these hypotheses, we developed the novel tool as a component (“macro”) for the freely-available and widely-used image analysis program, ImageJ. Using histologic sections from an existing model of acute cardiorenal syndrome, which produces significant tubular epithelial cell death, we quantified cell death by three methods: gold-standard stereology using a controlled stage, unbiased stereology using the novel software, and quantitative area (2D) morphometry, also using the novel software.
Methods

Animals and Experimental Groups

Animal experiments were approved by the Institutional Animal Care and Use Committee of Oregon Health & Science University, and were conducted in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals. The LRP2\textsuperscript{fl/fl} ApoE\textsuperscript{cre} strain was a kind gift of Thomas Willnow, Ph.D. (Max Delbruck Institut, Berlin, Germany) and was bred in our laboratory (n=16, 11 male and 5 female, 9±1.4 weeks old, and 24±3.5 gm body weight) prior to use in these experiments. This strain has been extensively described elsewhere; mosaic deletion of the primary protein endocytosis receptor in the renal proximal tubule causes low molecular weight proteinuria, but this strain is otherwise phenotypically normal. (Leheste \textit{et al.}, 2003; Bachmann \textit{et al.}, 2004; Theilig \textit{et al.}, 2007) We have previously demonstrated estrogen-mediated sex difference in renal cell death after CA/CPR which is abrogated by ovariectomy. (Hutchens \textit{et al.}, 2010; Hutchens \textit{et al.}, 2014; Ikeda \textit{et al.}, 2015) therefore females in this study underwent ovariectomy 7 days prior to CA/CPR.

Cardiac Arrest and Cardiopulmonary Resuscitation (CA/CPR)

We used the same procedure for CA/CPR as in previous experiments (Hutchens \textit{et al.}, 2010; Hutchens \textit{et al.}, 2011). Mice were anesthetized using 2% isoflurane in 45% oxygen. After induction of anesthesia, a 22 gauge venous catheter was placed orotracheally and a 0.61 mm outer diameter polyethylene tube was placed in the jugular vein. The electrocardiogram was monitored via transcutaneous needle probes. Rectal temperature was maintained between 36.5-37.5 °C using a heating lamp. Cardiac electrical activity was stopped with intravenous potassium chloride. Cardiac arrest was confirmed by electrocardiographic silence and absence
of visible chest wall cardiac contractions. Isoflurane and oxygen were then discontinued. 7 minutes and 30 seconds later, mechanical ventilation with 100% oxygen was started. 8 minutes after the induction of cardiac arrest, 300 chest compressions/min were started; 8-16μg of epinephrine was administered intravenously over 30 seconds. If return of spontaneous circulation, defined as visible cardiac contractions on the chest wall with spontaneous electrocardiographic R waves, was not observed after 180 seconds, resuscitation efforts were stopped. When spontaneous respiratory rate reached 60 breaths per min the tracheal tube and jugular vein catheter were removed and mice were placed in a recovery cage placed on a 37°C warming pad for 2h before return to standard housing.

*Preparation of histologic sections for stereology and morphometry, and measurement of serum creatinine and urea nitrogen*

24 hours later, mice were anesthetized with 4% isoflurane in oxygen and air mixture. After blood was aspirated from the left ventricular apex via thoracotomy, mice were perfused with saline followed by 4% paraformaldehyde and the right kidney was collected, fixed in 4% paraformaldehyde overnight and subsequently paraffin-embedded. Whole blood samples were immediately centrifuged at 5000 rpm for 5 minutes in serum separator tubes (BD, Franklin Lakes NJ). Serum creatinine and serum urea nitrogen were measured by creatine aminohydrolase assay on an autoanalyzer (Abaxis, Union City CA). 6μm-thick sagittal sections were cut perpendicular to an imaginary line drawn from the middle of the caudal pole to the middle of the cranial renal pole. (i.e., the entire kidney was sectioned) at 160 μm intervals. The first section was obtained at random distance from the pole. Five sections were mounted on each slide. Sections were then stained with Fluoro-Jade B (Histo-Chem, Jefferson, AK) which stains dead cells bright green.
Morphometry and unbiased stereology.

We used three methods to quantify renal cell death on all sections from each mouse. The gold standard in this study was estimation of volume fraction of Fluoro-Jade B stained cells with reference to the total volume of renal tissue (i.e., according to the Cavalieri principle), quantified using an upright fluorescence microscope (Leica DM4000B) with computer-controlled motorized stage (Proscan II, Prior Scientific) and dedicated software (NEWCast, Visiopharm, Hørsholm, Denmark). This system moves the microscope stage to random locations within a user-defined area, imposes a grid for counting and reports the number of user-selected “positive” grid points and user-selected “area” grid points. A ratio of positive grid points to area is used to calculate the volume fraction of positive tissue with reference to the volume of tissue. In this study we used a point grid consisting of two marks (cell outcome: reference space) in a fixed relation of 4:64. The reference space is defined as the entire kidney, and quantified by counting any point in the large grid which falls on kidney tissue. The estimated volume fraction of necrotic tubular epithelium was calculated according to the formula:

\[ Vv = \frac{\sum P(\text{necrotic tubular epithelium})}{\sum P(\text{kidney})} \times 16, \]

where \( \sum P(\text{necrotic tubular epithelium}) \) is the sum of grid points intersecting with a dead tubular epithelial cell and \( \sum P(\text{kidney}) \) is the sum of points intersecting with kidney tissue (the reference space). The constant 16 is required to correct for the difference in point number used to determine the reference space (kidney tissue points, 64) at low magnification and the smaller number of points (4) used to determine cell outcome at high magnification.

Results were expressed as the volume fraction of necrotic tubular epithelium (%). This point ratio and sampling frequency performed adequately in prior studies, demonstrating a coefficient of error below 0.1 using Gundersen’s formula. (Gundersen et al., 1999)
Following gold-standard stereology, each slide was scanned using a slide scanner (Zeiss Axioscan) for analysis with the novel tool. The novel tool determines the total area of each kidney slice and after a threshold is set, determines “stained” or “nonstained” for each pixel within the kidney area. The user may exclude voids within each slice to reduce this potential source of artifact. For each scanned image, the tool reports the ratio of stained to total pixels within the kidney area (morphometric area quantification). Simultaneously, the novel tool facilitates unbiased stereology by imposing a random grid, and using that grid to export a stochastic series of high-power fields (i.e. truly random) acquired from the kidney slice image, superimposed with a random point grid which may then be manually counted and used to calculate as above. The tool reports the fluorescence value under each crosshair; with a binary threshold this value may be used to determine “stained” vs. “nonstained” status, and completely automated stereologic quantification is enabled.

The comparison methods were, respectively, morphometry (specifically the ratio of stained area to total kidney area, averaged over 4 slices per mouse), and automated systematic random sampling conducted using the novel tool. For the systematic random sampling, the definition of “dead” and the definition of the reference space (i.e., all points intersecting kidney) was the same as for the gold standard method. Details of the use and function of the novel tool are presented in figure 1, extensive instructions for use are provided in supplemental file 1 and the commented source code is presented in supplemental file 2.

Statistical analysis

We used Prism version 7.0 for statistical analysis. Random numbers in this study were obtained from a random number generator. Two group comparisons were carried out using Student’s t-test for continuous data, Fishers exact test for categorical data (sex), and McNemar’s test for nominal data (survival). Statistical significance was inferred when p<0.05.
Correlation analysis was performed by Pearson’s method. We performed Bland-Altman analysis to quantify agreement between two stereologic methods. Bland-Altman analysis was also employed to quantify disagreement and bias between morphometric measurement and stereology.

Results

Cardiac arrest survival and Acute kidney injury

Table 1 summarizes baseline, resuscitation, survival, and serum renal functional test data for the two groups of mice subjected to CA/CPR. Potassium chloride induced cardiac arrest in all mice. There was no difference in pre-arrest age, weight or sex between groups. Nor was there difference in the total time of cardiac arrest, the time taken to resuscitate (CPR time), the epinephrine dose, or survival between groups.

24h after CA/CPR, all mice exhibited elevated serum urea nitrogen and serum creatinine, consistent with acute kidney injury. Neither serum urea nitrogen nor creatinine was different between groups (serum urea nitrogen 69.1 ± 22.6 in Cre+ versus 76.5 ± 29.5 mg/dL in Cre-, n=13, p=0.7 and serum creatinine 0.31 ± 0.06 versus 0.33 ± 0.13, mg/dL, n=13, p=0.11). Therefore acute kidney injury 24h after CA/CPR was not mediated by mosaic deletion of megalin, and the entire cohort was treated as a single group for further analysis.

Evaluation of novel tool-based stereology

To compare the gold standard method to the novel tool, we determined volume fraction of necrotic cells in all mice. The mean volume fraction of necrotic tubular cells, with reference to the kidney volume, was 0.010 ± 0.003 by the gold standard method and 0.0137 ± 0.003 by the novel method. There was strong correlation between estimates by either method (Pearson r
=0.91, \( R^2 = 0.82, p<0.0001 \). Because correlation does not necessarily require agreement between two tests, we performed Bland-Altman analysis, which demonstrated a high degree of agreement between the two measurements (bias \(-0.004\pm0.006\) with 95% limits of agreement \(-0.015\) to \(0.007\)). Therefore the novel tool reliably estimates volume fraction of necrotic cells.

These results are illustrated in figure 2.

**Comparison of automated morphometry and gold standard stereology**

Because morphometric analysis is expeditious and frequently used to quantify cell death in renal injury models, we compared automated morphometry using the novel tool with gold standard stereology in order to test the hypothesis that automated morphometric quantification of cell death approximates stereologic quantification. The results of morphometry and gold standard stereology were dissimilar: the mean morphometric quantification of cell death was 0.064±0.010 (area %), whereas stereology indicated significantly less injury, 0.014±0.004 (volume %), \(p<.0001\), \(n=14\). While there was moderate correlation between the two methods across the entire sample (Pearson \(r=0.57, R^2=0.32, p<0.05, n=14\) pairs), Bland-Altman analysis demonstrated poor agreement (95% limits of agreement \(-0.114\) to \(0.014\)) with substantial bias (0.050±0.033) indicating that area morphometry overestimates gold-standard stereologic quantification of cell death on the same sections. These results are summarized in figure 3.

**Comparison of automated morphometry and novel tool-based stereology**

To ensure that the dissimilarity in morphometry and stereology was not due to dissimilar platforms and methods, we compared automated morphometry and novel-tool based stereology, conducted on the same platform, from the same scanned image files. Results were again dissimilar with area morphometry generating a larger estimate of the injury (automated morphometry 0.064±0.010, novel tool stereology 0.010±0.003, \(p<0.0001, n=14\)).

Wakasaki et al Automated Cavalieri Stereology in AKI
correlation was again observed (Pearson r=0.61, $R^2=0.38$, p=0.02) but agreement was poor with persistent bias (0.054±0.033). These results are summarized in figure 4.

**Discussion**

In this investigation we developed a novel software tool and workflow to facilitate rapid quantification of cell death in a well-characterized acute kidney injury model. The commented source code, and installation instructions are included as supplemental data (Supplementary files 1, 2). To our knowledge this is the first freely-available, fully automated solution specifically designed to evaluate renal cell death. However, the program is designed to be agnostic as to tissue and image source, therefore there are several precedents. Since the advent of digital image handling, a number of investigative teams have developed analytic packages with similarities to ours. STEPanizer(Tschanz *et al.*, 2011), developed in 2011, is developed in JAVA and facilitates multiple stereologic analytic techniques on stored images. In STEPanizer, several types of stereologic interrogation maps may be superimposed on JPEG-format image files, the program then facilitates counting of map/feature interaction. Extensive use of this tool, and interest in such, is evidenced by 87 citations in the US National Library of Medicine publication database (Pubmed). Sallon *et al.* describe an automated tool, written in MATLAB, for analysis of lung sections which utilizes scanned slides and which provides extensive statistical analysis from automated results. A similar MATLAB approach to slide-scanner images demonstrated correlation between automated and manual point counting approaches.(Nault *et al.*, 2015) Our novel software tool contributes to this emerging analytic toolset for whole-slide digital imaging. Advantages of our approach include dependence only on the open-source and widely-used ImageJ platform, a simple, procedural interface, and straightforward path to modifications in the ImageJ macro language. We believe the tool we describe will find wide application and modification in histologic analysis in injury models.
We then tested the novel tool against a gold standard stereologic workflow. The major finding of this study is that the novel tool provides reliable estimates of tubular epithelial cell death which are nearly identical to gold standard methods.

Because area morphometry is expeditious and has been widely used for this purpose, we built this into the novel tool. This facilitated direct testing of this method against stereology on the same sections. Since gold standard and novel software-based stereology were in close agreement, we interpret our findings to mean that there was overestimation of cell death by area morphometry and poor agreement between area morphometry and stereologic estimation of cell death. We hypothesize three reasons why this may have occurred. First, automated area morphometry counts all pixels above a threshold value as positive, without determining whether those pixels represent tubular epithelial cells, whereas human observers may not count such stained pixels as positive. Second, area morphometry can be expected to count any structures which extend across two sections twice, whereas systematic random sampling is specifically designed to reduce to near-zero this error. Third, area morphometry may systematically underestimate total kidney area because although large voids may be excluded by the user, small voids may not be accounted for, reducing the reference area. Given the increasing availability of whole-slide digital imaging with subsequent unbiased evaluation and persistent bias in area morphometry, our results suggest area morphometry should be viewed as a preliminary method of analysis. Our data support editorial and scientific society endorsements of stereology (Saper, 1996; Madsen, 1999; Hsia et al., 2010).

Our study has limitations. A slide scanner is required to digitize images. No specification is made as to how the analyzed sections are derived from tissue, but a general limitation of stereology is that tissue must be sectioned such that the structure of interest is equally probable to be encountered in all sections. If non-systematically random sections undergo automated analysis, the bias inherent in the sectioning will be reflected in biased analysis. Although
Stereologic methods are defined for estimating the number of objects and area, our novel software is limited to the Cavalieri method, which estimates volume – in the case of tubular epithelial cell death, the volume fraction of necrotic epithelium – with reference to a larger volume, the reference. The novel software is designed to facilitate systematic random sampling of any tissue and stain. Automated analysis is best facilitated with fluorescent stains of binary outcome such as fluoro-jade B. We have recently employed the software to facilitate quantification of tubulointerstitial α-smooth muscle actin staining using FITC-conjugated antibody. Because the novel software produces images which can be human-read, however, stains such as hematoxylin and eosin and periodic acid-Schiff may also be employed. However, although the software may be extended to use for other stains and tissues, our demonstration of efficacy is limited to Fluoro-Jade B stained renal sections, and use for other purposes should be validated. Lastly, although construction of the novel software as an ImageJ macro makes it widely accessible and easy to understand, the ImageJ macro language limits the extent to which sophisticated statistical analysis can be performed, and may limit the speed of analysis.

We conclude that Cavalieri method volume fraction estimation using a novel ImageJ macro reliably quantifies cell death in a model of acute kidney injury, simplifying and extending the use of this important quantification technique.
Figure & Table Captions

Figure 1: Automated systematic random sampling of scanned histologic images using a straightforward workflow.

Figure 2: Novel software-based stereology and gold standard, dedicated-software based quantification closely agree. Quantification of renal tubular cell death (volume fraction of necrotic tubular epithelial cells, VNC) using a gold standard method (dedicated stereology software with a computer-controlled microscope stage) was compared with quantification on the same histologic sections conducted using novel software which facilitates systematic random sampling on scanned slide images and marks random points for manual assessment of positive or negative staining. The gold standard and novel software yielded highly similar results (A). B. Correlation analysis demonstrates high degree of correlation between the two methods (middle, Pearson r=0.91, R²=0.83, p<0.0001, n=14 pairs). C. Bland-Altman analysis was employed to test agreement between methods and demonstrated high degree of agreement between the two measurements.

Figure 3: Automated area morphometry overestimated tubular necrosis. Depicted are the results of evaluation of renal tubular cell death by automated area morphometry using novel software and stereologic evaluation performed using the gold standard dedicated-software. In the automated morphometry mode the novel software quantified the total flouro-jade stained area relative to the area of kidney. Automated area morphometry and the gold standard yielded dissimilar results, with automated morphometry overestimating stereology-determined cell death in many sections (A). Although there was moderate correlation between the methods, (B, Pearson r=0.57, R²=0.32, p<0.05, n=14 pairs), agreement was poor; land-Altman analysis (C) demonstrated nearly 5% overestimation (bias) and 95% limits of agreement encompassing a wide range (Bias -0.050±0.033 with 95% limits of agreement -0.114 to 0.014). These findings
suggest that use of area morphometry to quantitate tubular cell death in this setting may be expected to produce heterogeneous results which may overstate injury.

Figure 4: Novel software-based stereology and automated area morphometry of identical digitally scanned, fluoro-jade B stained histologic images demonstrates overestimation of cell death by area morphometry. The novel software was used to output systematic random sampled, grid-imposed images, which were manually evaluated. Area morphometry was performed in automated fashion by the novel software. Automated area morphometry (A, mean±SEM displayed), recorded a higher mean estimate for cell death than stereology, with greater variation. Although automated morphometry and stereology methods modestly correlated (B, r=0.61, R^2 = 0.38, p=0.02), agreement between individual measurements was poor. Bland-Altman analysis demonstrated consistent bias (0.054±0.033) and wide 95% limits of agreement (−0.117 to 0.01), encompassing nearly the entire range of injury.
References


Table 1: Baseline and resuscitation, and renal injury data

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Age (weeks)</th>
<th>Body weight (g)</th>
<th>Males /females</th>
<th>Epinephrine dose (µg/g weight)</th>
<th>CPR time (seconds)</th>
<th>Survival rate (%)</th>
<th>24h serum urea nitrogen (mg/dL)</th>
<th>24h serum creatinine (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cre+</td>
<td>8.7±1.5</td>
<td>21.9±3.1</td>
<td>4/3</td>
<td>0.46±0.06</td>
<td>70.8±18.4</td>
<td>88.9</td>
<td>69.1±22.6</td>
<td>0.31 ± 0.06</td>
</tr>
<tr>
<td>Cre-</td>
<td>9.0±1.1</td>
<td>22.0±3.4</td>
<td>5/2</td>
<td>0.52±0.14</td>
<td>90.0±38.7</td>
<td>87.5</td>
<td>76.5±29.5</td>
<td>0.33 ± 0.13</td>
</tr>
<tr>
<td>p value</td>
<td>0.72</td>
<td>0.97</td>
<td>&gt;0.99*</td>
<td>0.39</td>
<td>0.33</td>
<td>0.08**</td>
<td>0.84</td>
<td>0.89</td>
</tr>
</tbody>
</table>

P values shown are derived from Student’s t-test except:

* P value was calculated by Fisher’s exact test.

** P value was calculated by McNemar’s test.
1. Open original kidney image in ImageJ and run macro. Define analysis parameters. (Scale bar 100µm)

2. Automated thresholding allows determination of tissue area and the program selects systematically random regions of interest which intersect tissue.

3. A systematically random grid is overlaid on the regions of interest. Each ROI is exported with and without the grid.

User may manually count point-labeled images

Program reports pixel value under each grid point and deviance from mean pixel value. If a threshold is set, positivity can be set for automated, systematic random quantification.
**A**

Volume fraction of necrotic tubular cells

- **Novel software**
- **Gold standard**

**B**

Volume fraction of necrotic tubules

- **Novel software (VNC)**

**C**

Difference

- **Average**
HISTOLOGY AND HISTOPATHOLOGY

(A) Volume fraction of necrotic tubular cells (VNC)
   - Novel software
   - Area morphometry

(B) Automated area morphometry vs. Volume of necrotic tubular epithelial cells (VNC)

(C) Difference
   - Average