Chapter 4

Assessment of accuracy, precision and efficiency of manual planimetry and point counting method: A comparison with a gold standard

Xingnan Lin, Arjen J. Van Wijk, Dongyun Wang, Wil G.M. Geraets, Yuelian Liu

Submitted for publication
ABSTRACT

To know the size of area or volume of structures or organs is important in biological or medical field. Two Cavalieri’s principle based methods, manual planimetry and point counting method have been widely used for that purpose. However, due to the lack of reliable gold standard reference, there was considerable controversy in the comparison of accuracy for the two methods. In current study, we created 30 series of artificial sections with different known area of interest by pasting red paper sheet on a white graph paper, mimicking different sections from one sample. Then we used it as a gold standard reference for the comparison of accuracy, precision and efficiency of the manual planimetry method and point counting method. Qwin image analysis program was used to represent manual planimetry method. Two observers applied both methods twice with an interval of two weeks. The results show that both method has comparable high precision. Point counting method exhibits higher accuracy and better efficiency than manual planimetry method. The accuracy of manual planimetry method depends on the workload of manual tracing.

KEY WORDS: Cavalieri’s principle, Gold standard, Planimetry, Point counting, Stereology
INTRODUCTION

To know the volume of organs or other specific structures is of vital importance for disease diagnosis and treatment planning. To this end, Cavalieri’s principle of stereology has been widely used, in which, the volume of samples were estimated by multiplying the areas of interest on a set of parallel two-dimensional medical images multiplied by a known distance between two adjacent images\(^1\). Considering the latter aspect is usually fixed, the validity of the estimated volume is highly dependent on the accuracy and precision of estimated area of interest\(^2\). Planimetry and point counting are two commonly used methods based on the Cavalieri principle for estimating volume. Planimetry usually involves manually tracing the boundaries of the area of interest and the area is estimated by the amount of corresponding pixels\(^3\). Point counting estimates the area of interest by multiplying the number of points hitting the object by the area one point represent in the grid\(^4\). Apart from precision and efficiency, accuracy is an important aspect for the application of the two methods. However, lots of previous researches compared the precision and efficiency, instead of the accuracy of the two method due to the lack of the real volume of the testing samples\(^3, 5-9\). Consequently, the validity of these studies was inevitably impaired\(^1, 10, 11\).

In order to solve this problem, the fluid displacement technique was introduced as a gold standard for comparing the accuracy of manual planimetry and point counting methods\(^1, 12\). Using the fluid displacement technique, the object of interest is immersed in a vessel filled with water. The volume of overflowing water was measured as the actual volume of the object of interest\(^1, 13, 14\). However, the use of the fluid displacement technique has some drawbacks, which include the manpower and material cost for collecting samples\(^3, 9, 12\) and the impairment of the accuracy of the gold standard due to the shrink of biological tissue during the fixation of the sample\(^15\). Furthermore, if the object of interest, such as a lung, has internal cavities that are open to the outside, then the cavities fill with water when the organ is submersed, leading to an underestimation of the volume\(^16\).

In order to overcome the difficulties mentioned above, we developed a new simple method to estimate the accuracy of manual planimetry and the point counting
method. A series of artificial sections with different known area of interest were created by pasting red paper sheet on a white graph paper, mimicking different sections from one sample. These areas of interest can act as gold standard reference because their sizes were pre-established according to the design of the study. Next, Qwin image analysis program, a versatile software widely used for volume estimation in medical field, was chosen to represent manual planimetry method. The purpose of the current study was to compare the accuracy, precision and efficiency of the manual planimetry and the point counting method with the presence of a new simple gold standard.

MATERIALS AND METHODS

Establishment of Gold standard reference

Thirty artificial sections were created with white graph paper and red sheets in the current study. In each section, a white graph paper with small grids (5×5mm) was used as background. Red sheet, with known area, were cut into random shapes and then pasted on the white graph paper to mimic the area of interest on radiographic or histomorphometric photos. A piece of square red sheet, which available at stationery shops with fixed area of 89x89 mm², was used as a benchmark. One benchmark can be easily and precisely divided into halves, quarters and eighth parts by paper knife. Using these properties, we took 1+1/8 benchmark sheet, cut it into random shapes and pasted the fragments on a white graph paper to obtain the first section (Figure 1). The second section was obtained by cutting 1+2/8 benchmark sheet into fragments which were pasted on another white graph paper. In this way, a series of 30 sections were created with known red area of 1+1/8, 1+2/8, 1+3/8, ..., 1+30/8 times the benchmark area. The area of interest can be written as A(i) = (1+i/8) 89x89 mm² (i=1,2,....30). Manual planimetry and point counting method were used to estimate the area of interest on the 30 sections.
Figure 1. Artificial sections used in current study. (A): A benchmark red sheet with area of 89×89 mm². (B): Sample 1 with size of (1+ 1/8) benchmark sheets. (C): Sample 15 with size of (1 + 15/8) benchmark sheets. (D): Sample 30 with size of (1 + 30/8) benchmark sheets.

Manual planimetry method

The 30 artificial sections were scanned by Ricoh Aficio MP C4502 scanner (Ricoh Electronics, Inc., Toyota, Japan) and saved in “tiff” format with 600 dpi. These digital images were imported to the Qwin image analysis program (Leica Microsystems Image Solutions, Rijswijk, the Netherlands) for area estimation. A polygon selection tool was used to delineate the outermost boundaries of area of interest, namely every fragment of the red sheet on each artificial section (Figure 2A). The total area
of interest on each artificial section was automatically calculated by the Qwin image analysis program by summing all selecting areas up. Considering the possible distortion of images after scanning in both length and width, a magnification correction was carried out as following: magnification of length = measured length / real length; magnification of width = measured width / real width. Therefore, the final magnification of area = magnification of length × magnification of width.

**Figure 2.** Two methods used for estimation of area of interest in current study. (A): Estimation of the area of interest by Qwin image analysis program. (B): Estimation of the area of interest by the point-counting method.

**Point counting method**

The point counting method consists of overlying each section with a systematic array of test points, which are randomly positioned on the area of interest to avoid bias\(^6\). At each superimposition, the number of test points hitting the area of interest on the section is counted (Figure 2B). A point is counted either if it is completely inside the area of interest or the right-hand corner of a crossing point overlaps with
the area of interest. The area of interest is then calculated as $A = n \times d^2$, where $A$ represents area of interest; $n$ represents the number of points hitting the area of interest; $d$ represents the distance between two neighboring test points. A reasonable precision of the volume estimate is obtained by counting 100-200 points on 7-12 sections in total per organ. That means for the estimation of each area of section, the number of hitting points should be from 8 to 29 to guarantee precision. To this end, we chose a transparent sheet with grid size of 1.25cm for the estimations for the areas of interest. The smallest number of hitting points within the 30 sections was 50 (in the first sample), which was somewhat conservative and higher than the above-mentioned threshold to guarantee the precision of point-counting method in the current experiment.

**Intra- and interobserver variability**

To evaluate intraobserver and interobserver variability, two observers familiar with both Qwin image analysis program and point counting method were asked to independently estimate the area of interest on the 30 sections. Each observer carried out estimation independently by both methods twice with an interval of two weeks. Reproducibility of estimations was expressed in terms of coefficient of variability (CV) values.

**Statistics**

To determine the agreement between the real area (golden standard) and the estimated areas (accuracy or validity), as obtained using the point counting method or manual planimetry method, for observer 1 and 2 and both estimations, the Intraclass Correlation Coefficient (ICC: 2-way mixed model, absolute agreement) was used (single ICC reported). The ICC was also calculated to determine the agreement between repeated estimations within each method (precision or reliability). To complement these analyses, the difference between the real area and the estimated area was calculated, and expressed as a percentage relative to the real area, i.e. $(\text{estimated-real})/\text{real} \times 100$, where a negative difference represents an
underestimation of the real area, and a positive score an overestimation of the real area.

RESULTS

In the present study, 30 artificial sections with different areas were constructed. The mean area was 23268 mm$^2$ (SD=8716.5), the smallest area was 8911 mm$^2$ and the largest was 37625 mm$^2$.

Agreement with real area (accuracy)

For both observers, the ICC between real and estimated area, for both methods, and both estimations, and the average of both estimations was calculated. Results are displayed in Table 1 and suggest a strong extent of agreement between the estimated area and the real area. In addition, the ICCs for the point counting method is systematically (slightly) higher than for the computer method.

Table 1. ICC’s between the estimated and the real area for both observers and their measurements, and within each method for both observers separately (E1 represents first round of estimation; E2 represents second round of estimation).

<table>
<thead>
<tr>
<th>Observer</th>
<th>E1</th>
<th>E2</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point counting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within point counting</td>
<td>0.995</td>
<td>0.991</td>
<td>0.994</td>
</tr>
<tr>
<td>Planimetry</td>
<td>0.978</td>
<td>0.974</td>
<td>0.976</td>
</tr>
<tr>
<td>Within planimetry</td>
<td>0.998</td>
<td>0.9996</td>
<td>0.9996</td>
</tr>
</tbody>
</table>

Observer 1
Observer 2
Since the ICC is a relative measure (the ratio of within and between subjects variability) it is also informative to know the extent of disagreement between estimated area and the real area. Therefore, the difference between the estimated area and the real area was expressed as a percentage relative to the real area. Descriptive statistics regarding this analysis are presented in Table 2.

Results in Table 2 show how large the differences were between the real area and the estimated area, in how many percent of cases. To illustrate, for the point counting method, observer 1, first estimation, the mean difference was -2.16\% (SD=3.58), the highest overestimation was 4.73\% while the largest underestimation was 10.57\%. In 80\% of all estimated samples, the differences ranged between 4.73\% to -5.23\%. Formulated differently, 80\% of the estimated samples does not differ more than about 5\% from the real area. Comparing all the mean differences found using both methods using paired-samples t-test leads to the conclusion that, on average, the difference between the estimated and the real area is larger in manual planimetry method (all p < 0.001). However, the standard deviations around the mean are consistently lower for the manual planimetry method, indicating less variation. If the range between the minimum and maximum are compared across methods, it is clear that the difference scores for the manual planimetry method lie within a smaller range than the point counting method. Another observation is that in the use of point counting method, observer 2 seems to systematically overestimate the area’s relative to observer 1, i.e. more positive deviations. In addition, it is clear that the manual planimetry method yields more negative scores than point counting method, indicating a higher level of underestimation. To conclude, the analysis so far give the impression that the planimetry method yields slightly more consistent results (better precision), however, results from the point counting method seem to agree more with the real area (better accuracy).
Table 2. Percentage deviation of the estimated area relative to the real area, for both methods, observers and estimations. (E1 represents first round of estimation; E2 represents second round of estimation)

<table>
<thead>
<tr>
<th></th>
<th>Point counting</th>
<th></th>
<th></th>
<th>Planimetry</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observer 1</td>
<td>Observer 2</td>
<td></td>
<td>Observer 1</td>
<td>Observer 2</td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>-2.16</td>
<td>-4.61</td>
<td>3.06</td>
<td>1.10</td>
<td>-6.05</td>
<td>-7.77</td>
</tr>
<tr>
<td>E2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-3.59</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-3.15</td>
</tr>
<tr>
<td>SD</td>
<td>3.58</td>
<td>3.16</td>
<td>3.86</td>
<td>5.08</td>
<td>1.94</td>
<td>1.31</td>
</tr>
<tr>
<td>Max</td>
<td>4.73</td>
<td>2.08</td>
<td>13.62</td>
<td>18.63</td>
<td>-1.37</td>
<td>-4.65</td>
</tr>
<tr>
<td>90%</td>
<td>2.84</td>
<td>.00</td>
<td>7.50</td>
<td>6.27</td>
<td>-3.74</td>
<td>-5.84</td>
</tr>
<tr>
<td>80%</td>
<td>0.64</td>
<td>-2.70</td>
<td>5.60</td>
<td>5.08</td>
<td>-4.05</td>
<td>-6.58</td>
</tr>
<tr>
<td>70%</td>
<td>-.26</td>
<td>-3.24</td>
<td>4.88</td>
<td>3.70</td>
<td>-4.57</td>
<td>-7.45</td>
</tr>
<tr>
<td>60%</td>
<td>-.98</td>
<td>-3.91</td>
<td>4.01</td>
<td>2.39</td>
<td>-5.29</td>
<td>-7.60</td>
</tr>
<tr>
<td>50%</td>
<td>-2.06</td>
<td>-4.09</td>
<td>3.37</td>
<td>1.00</td>
<td>-6.49</td>
<td>-7.75</td>
</tr>
<tr>
<td>40%</td>
<td>-3.00</td>
<td>-5.13</td>
<td>2.66</td>
<td>-0.22</td>
<td>-6.93</td>
<td>-8.13</td>
</tr>
<tr>
<td>30%</td>
<td>-4.29</td>
<td>-5.80</td>
<td>0.95</td>
<td>-1.97</td>
<td>-7.53</td>
<td>-8.53</td>
</tr>
<tr>
<td>20%</td>
<td>-5.23</td>
<td>-6.84</td>
<td>-1.02</td>
<td>-3.45</td>
<td>-7.85</td>
<td>-8.75</td>
</tr>
<tr>
<td>10%</td>
<td>-7.68</td>
<td>-9.94</td>
<td>-2.46</td>
<td>-5.19</td>
<td>-8.73</td>
<td>-9.41</td>
</tr>
<tr>
<td>Min</td>
<td>-10.57</td>
<td>-11.63</td>
<td>-4.26</td>
<td>-7.12</td>
<td>-8.94</td>
<td>-10.39</td>
</tr>
</tbody>
</table>

Association between real area and over- and underestimation

In order to test if the difference between estimated and real area is associated with the real area, Pearson correlations were calculated between the difference score and the real area (Table 3). As the results clearly show, the size of the difference between real and estimated area is very strongly related to the real area for the
manual planimetry method only (with the exception of estimation 2 by observer 2 using the point counting method). In other words, the larger the area to be estimated, the larger the difference between real and estimated area (for the manual planimetry method).

**Table 3.** Pearson correlation between estimated area and deviations from that of artificial sections for two methods (point counting and manual planimetry), two observers, and two estimating sessions. (E1 represents first round of estimation; E2 represents second round of estimation)

<table>
<thead>
<tr>
<th></th>
<th>Point counting</th>
<th>Planimetry</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observer 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>0.17</td>
<td>0.96*</td>
</tr>
<tr>
<td>E2</td>
<td>0.30</td>
<td>0.92*</td>
</tr>
<tr>
<td><strong>Observer 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>-0.23</td>
<td>0.95*</td>
</tr>
<tr>
<td>E2</td>
<td>0.39*</td>
<td>0.95*</td>
</tr>
</tbody>
</table>

* p<0.05

**Agreement within methods (precision)**

In order to determine the precision for each method, the ICC between estimation 1 and 2 for each method and observer was determined (Table 1). For observer 1, the ICC for the point counting method was 0.993 and for computer method 0.998. For observer 2, the ICC for the point counting method was 0.988 and for computer method 0.9996.

**Time needed for the estimation (efficiency)**

In order to compare the two methods on the time needed to perform the estimations, a repeated measures ANOVA was performed. Results show a significant main effect for time, F (7, 23) = 194.1, p <0.001, implying that not all estimations took equally long. Mean scores and standard deviations are presented in Figure 3. The means plot in Figure 3 summarize the story well, point counting method was faster.
than manual planimetry method, Observer 2 was slower than Observer 1, which was only true for the manual planimetry method.

![Figure 3. Mean (and standard deviation) time needed to perform the estimations.](image_url)

**DISCUSSION**

In the field of stereology, manual planimetry and point counting are two frequently used techniques based on the Cavalieri principle for estimating volumes or areas. In the current study, a new simple gold standard method was set up as a conference for the comparison of the precision, accuracy and efficiency between the manual planimetry and the point counting method.

In the field of science, accuracy and precision are two terms which are easily to be confused. The accuracy of a measurement system is the degree of closeness of measurements of a quantity to that quantity’s true value. The precision of a measurement system, related to reproducibility and repeatability, is the degree to which repeated measurements under unchanged conditions show the same results. Accuracy and precision are totally different terminologies in science and have no any necessary connections. For example, a measurement method can be...
Point counting and manual planimetry

precise but not accurate, accurate but not precise, neither, or both. A measurement method is considered valid only if it is both accurate and precise. Without the proof of good accuracy, the validity of tested methods was inevitably impaired. This opinion is supported by Gong\textsuperscript{21} and he thought the accuracy of the manual planimetry and point counting method is less straightforward due to the impossibility of ex vivo volume measurement for the accessed organ or tissue.

In some previous studies in which the accuracy of manual planimetry and point counting was compared with the presence of fluid displacement technique for the estimation of real volume of samples, the results varied a lot\textsuperscript{1,9,22}. In one study for the estimation of volume of human temporal lobe, point counting method proved better accuracy than manual planimetry method\textsuperscript{1}. On the contrary, in another study to estimate liver volume, manual planimetry method provided more accurate results than point counting method\textsuperscript{9}. In one research for volume estimation of intracranial volume\textsuperscript{22}, manual planimetry method and point counting method were proved with equal low accuracy because the results from both methods were significantly different from the estimated volume of skulls measured by fluid displacement technique. Several factors associated with fluid displacement technique may play a role in the inconsistent results mentioned above. Firstly, fluid displacement technique usually requires huge manpower and material cost\textsuperscript{3,9,12}. The process involves retrieving samples by surgery, trimming samples by removing neighboring irrelevant tissue and fixation of samples. Therefore operators who are capable of some special professional techniques are needed to guarantee the validity of experiments. In addition, the more steps involved in the process, the higher possibility to induce systematic error. Secondly, the shrinkage of biological tissue during the fixation of sample processing causes irreversible destruction and impairment to the accuracy of gold standard due to the\textsuperscript{15}. Finally, if the estimated organ under test, such as lung, has internal cavities that are open to the outside, when the organ is immersed the cavities fill with water, leading to an underestimation of volume unless the holes in the organ have been plugged in a watertight manner before immersion\textsuperscript{16}. 


Comparing with fluid displacement technique, our method of artificial sections with known real area is proved to be a reliable, simple and low-cost method to act as a gold standard reference. Firstly, the area of the commercial available red sheet is fixed that means almost no opportunities of making error for estimating area of artificial sections. This guarantees the validity of accuracy comparison between manual planimetry method and point counting method in our study. Secondly, the process of creating the artificial sections is so simple that even people without special professional techniques and instruments could easily carry it out. Finally, both the manpower and materials cost for the construction of the artificial sections is lowered to the extent that almost everyone can afford it. By virtue of the advantages of this method, we are able to easily set different sizes and shapes of area of interest for each artificial section. Furthermore, we can set as many as 30 sections in the current study to reduce the potential systematic deviation which usually result from limited number of samples derived from human beings or animals. These merits make the method of artificial sections to be an impartial gold standard reference for the area or volume comparison between manual planimetry method and point counting method.

The results from table 1 and 2 show that point counting method is more accurate than manual planimetry method, which is in agreement with one previous research\(^1\). This could be explained by different working mechanisms of the two methods. In point counting method, the counts are made of randomly overlaid lattice points that fall within a closed boundary area and a positive hit is scored by the eye-brain complex only\(^2\). The primary source of error in this technique is the decision errors about whether a point is inside or outside an area’s boundary\(^2\). On the contrary, the errors in the application of manual planimetry method results from not only the decision process which point counting experienced, but also the manual tracing process, which is tedious and demanding for operator’s patience and experience. Tracing a thin line by hand easily introduces deviations due to the mechanical and dynamic limitations of the hand, especially when tracing tortuous structures\(^24\). Therefore, the error probability of the use of point counting is lower than manual planimetry because the borderline tracing is not needed. This opinion is also supported by the results from table 3, in which the accuracy of manual planimetry
method decreased along with the size of area of interest increased. In our experiment, bigger size of area of interest means more manual tracing for planimetry method because of more fragments increase the workload of tracing. The increasing workload has no relationship with the size of area of interest. For example, drawing the boundary for a very big size of circle is much easier than that for a small size of irregular shape. Therefore, to be precise, the accuracy of manual planimetry is mainly influenced by the complex of contour of area of interest, not the size of it. However, the accuracy of point counting is not influenced by the contour of area of interest because the manual tracing is not necessary.

In the current study, the ICC within manual planimetry method and point counting method were very high, which indicated good precision (repeatability) for both methods (Table 1). Although the related results showed higher ICC for manual planimetry method, which may imply that this method yielded somewhat more consistent results, this conclusion needs to be interpreted with some caution because differences are minimal. The good precision for both methods in our study are supported by lots of previous studies with fluid displacement as gold standard.

The working efficiency in area estimation of point counting method is significantly higher than that of manual planimetry method in current study (Figure 3). The result is supported by a lot of previous studies. The extra workload in manual planimetry for manually tracing the boundary will account for the huge time consuming as above mentioned. Furthermore, the manual tracing process need more time for estimating structures with complex contour, such as human body compartments and fetal. On the contrary, the process of determining whether points are inside or outside the area of interest is much more faster than erasing and re-drawing the outline in manual planimetry. However, there were few previous studies showed that more time cost for point counting than manual planimetry method. The possible reason maybe the operator used a purpose designed track ball to outline the boundary of area of interest, making manual planimetry a more convenient process than general condition. It is in agreement with a previous study of Mazonakis et al., who reported that manual planimetry method
spent less time by a factor of 1.5 in comparison with the time of point counting method for estimation the volume of malignant liver lesions. It could be explained by that they spent some time to alter the grid size when applied point counting method. Similar results was found in one study for estimation of mandible cyst\(^3\). Although the kind of tools using to delineate boundaries of cyst were not mentioned, the shape of the cyst was quite similar to an oval or rectangle with regular shape and the workload for manual tracing is very low, increasing the working efficiency for manual planimetry method.

CONCLUSIONS

In conclusion, by virtue of a new simple reliable gold standard reference, we compared the accuracy, precision and efficiency of manual planimetry and point counting method. Although both methods showed comparable high precision, point counting method exhibits higher accuracy and better efficiency than manual planimetry method. The accuracy of manual planimetry method depends on the workload of manual tracing, viz., the more irregularity of the contour of the area for estimation, the less accuracy of the results. On the contrary, the point counting method is not influenced by that.
REFERENCES


20. BiPM, I., IFCC, I., ISO, I., and IUPAP, O. International Vocabulary of


Point counting and manual planimetry