


2017

# Parts of the Whole: Error Estimation for Science Students

Dorothy Wallace  
*Dartmouth College*, [dorothy.wallace@dartmouth.edu](mailto:dorothy.wallace@dartmouth.edu)

Follow this and additional works at: <http://scholarcommons.usf.edu/numeracy>

 Part of the [Medical Education Commons](#), [Medical Sciences Commons](#), [Physical Sciences and Mathematics Commons](#), and the [Science and Mathematics Education Commons](#)

## Recommended Citation

Wallace, Dorothy (2017) "Parts of the Whole: Error Estimation for Science Students," *Numeracy*: Vol. 10 : Iss. 1 , Article 11.  
DOI: <http://dx.doi.org/10.5038/1936-4660.10.1.11>  
Available at: <http://scholarcommons.usf.edu/numeracy/vol10/iss1/art11>

---

## Parts of the Whole: Error Estimation for Science Students

### Abstract

It is important for science students to understand not only how to estimate error sizes in measurement data, but also to see how these errors contribute to errors in conclusions they may make about the data. Relatively small errors in measurement, errors in assumptions, and roundoff errors in computation may result in large error bounds on computed quantities of interest. In this column, we look closely at a standard method for measuring the volume of cancer tumor xenografts to see how small errors in each of these three factors may contribute to relatively large observed errors in recorded tumor volumes.

### Keywords

measurement error, error analysis, ellipsoid volume, tumor, xenograft

### Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

### Cover Page Footnote

Dorothy Wallace is a professor of mathematics at Dartmouth. She was 2000 New Hampshire CASE Professor of the Year, and the lead PI of the seminal NSF project, Mathematics Across the Curriculum. She recently finished a text in mathematical biology for first-year students, "Situated Complexity." She was a charter board member of the National Numeracy Network and is a founding co-editor of this journal.

## Parts Of The Whole

A Column by D. Wallace

The problem of how best to improve the numeracy of a society is a thorny one, embracing the learning process of a single student but rising in scale to include the management and alteration of an entire system of education. With the issue of quantitative literacy always in mind, this column considers various aspects of the systemic workings of education, the forces acting on classrooms, teachers and students, and mechanisms of both stasis and change. With the issues of volume 9, the column has pivoted to thoughts from developing and teaching “Math 4: Applications of Calculus to Mathematics and Biology,” which Dartmouth biology students can take as an alternative to second-semester calculus (see Rheinlander and Wallace 2011, <http://dx.doi.org/10.5038/1936-4660.4.1.3> in this journal).

## Error Estimation for Science Students

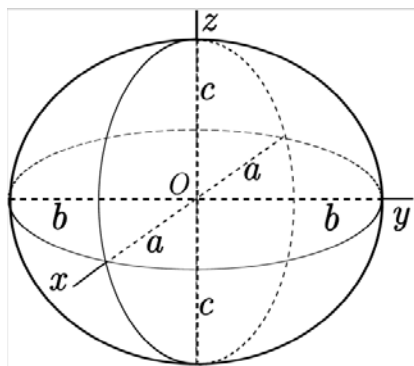
Estimation skills have been considered an essential part of numeracy, and with good reason. A good estimate may require a combination of visually estimating sizes, computing a desired quantity using algebraic expressions, and doing some kind of independent estimate to check that the resulting computation hasn't suffered from inappropriate use of units, transcription errors, or arithmetic blunders. These are all important considerations in scientific pursuits, but there is a second way in which estimation is critical to the ability to make productive use of scientific measurements. In applications, a researcher might want to fit a model to data. No fit is ever perfect, and the researcher (or student) might well wonder how much of this imperfection should be attributed to errors in measurement (or computation) of the data, and how much of it points directly to inadequacies of the model. In scientific applications, an estimate of how much error could reasonably be expected in the data is a useful quantity. As a model is supposed to describe reality, it is unreasonable to expect a model to fit a data set better than the data set itself represents reality.

A typical experiment in cancer research involves injecting a mass of human-derived tumor cells into a mouse. These special mice genetically lack their own immune system, which allows a human tumor to grow in the mouse. An experiment would usually divide a mouse population into control and treatment groups, and the size of the visible subcutaneous tumor, called a “xenograft,” would be measured over time. There are various ways to do this, but the least expensive involves taking length and width measurements with calipers. Researchers assume that the tumor mass is roughly an ellipsoid.

## Estimating the Volume of an Ellipsoid from Measurements

Figure 1 shows a diagram of an ellipsoid. One can see how three measurements would determine this object completely. Cut in half through the origin in the  $xy$ -plane, the cross section is an ellipse with major and minor semi-axes of lengths  $a$  and  $b$ . Cut through the origin in the  $yz$ -plane gives a cross-sectional ellipse with major and minor semi-axes of lengths  $b$  and  $c$ . So the three numbers, the semi-axes  $a$ ,  $b$ , and  $c$  determine the ellipsoid, whose volume is given by

$$V = \frac{4}{3}\pi abc \tag{1}$$



**Figure 1.** An ellipsoid aligned with  $x$ -,  $y$ -,  $z$ -Cartesian coordinates is determined by three semi-axes, with lengths denoted here as  $a$ ,  $b$ , and  $c$ .<sup>1</sup>

In the remainder of this discussion we will assume that  $2a$  is the largest measurement and  $2c$  is the smallest.

As a preliminary analysis we can ask how small errors in estimating the lengths  $a$ ,  $b$ , and  $c$  translate into errors in estimated volume. Starting with  $a = 2$ , and  $b = c = 1$ , we can compute the result of overestimating each of these by 10%. Table 1 shows the percent error in volume that comes from overestimating one, two, or all three of these parameters.

**Table 1.**  
**Errors in Ellipsoid Volume Resulting from Measurement Error**

Quantities	Longest	Second longest	Shortest	$V$	% error in $V$
	$a$	$b$	$c$		
True measurement	2	1	1	8.3776	0
10% error in $b$	2	1.1	1	9.2153	10%
10% error in $a$	2.2	1	1	9.2153	10%
10% error in $a$ and $b$	2.2	1.1	1	10.1369	21%
10% error in all three	2.2	1.1	1.1	11.1506	33.1%

Perhaps the most interesting thing about Table 1 is that errors propagate more than a student might expect. A 10% error in one measurement alone gives a 10% error in volume, but multiple errors have an effect that is more than additive. A

<sup>1</sup>Peter Mercator (2012), tri-axial ellipsoid with distinct semi-axes  $a$ ,  $b$  and  $c$ , *Wikipedia Commons* [https://commons.wikimedia.org/wiki/File:Ellipsoid\\_tri-axial\\_abc.svg](https://commons.wikimedia.org/wiki/File:Ellipsoid_tri-axial_abc.svg) (no changes made) accessed Nov. 16, 2016.

10% overestimate in all three measurements gives a 33.1% error in volume, more than the 30% that one might expect. Of course, the reason is given by a simple algebra calculation:

$$\begin{aligned} V_{\text{overestimate}} &= \left(\frac{4\pi}{3}\right)(a + .1a)(b + .1b)(c + .1c) \\ &= \frac{4\pi}{3}abc + \frac{4\pi}{3}(0.1a)bc + \frac{4\pi}{3}a(0.1b)c + \frac{4\pi}{3}ab(0.1c) + \text{extra stuff} \\ &= V_{\text{true}} + 3(0.1V_{\text{true}}) + \text{extra stuff} \end{aligned}$$

Here the extra stuff comes from the distributive property applied to the expression for the estimated volume. Even with this simple example, students could be invited to complete the algebra problem and ponder what would happen to the error if other sorts of geometric measures were taken (e.g., area versus circumference of a rectangle). They could also investigate how errors might cancel; for example, an overestimate of  $a$  and an underestimate of  $b$  would give a different table from the one above.

In addition, error analysis should lead to a discussion of why percent error is a better measurement of error in some cases than absolute error. For this example it could be argued that someone using a ruler or calipers to measure  $2a$ ,  $2b$ , or  $2c$  is likely to have comparable-sized errors no matter what length of  $2a$ ,  $2b$ , or  $2c$  is being measured. In that case the numbers recorded are likely to be off by the same amount, rather than the same percent. This would change the analysis just done.

### Estimating the Volume of a Solid Tumor Xenograft

The research paper by Ayers et al. (2010), on which the following discussion is based, describes a method often used for actually measuring a tumor xenograft by hand and computing its volume:

The two longest perpendicular axes in the  $xy$ -plane of each xenograft tumor were measured to the nearest 0.1 mm by three independent observers (reviewers 5–7) familiar with collecting caliper measurements of xenograft tumors in mice. The depth was assumed to be equivalent to the shortest of the perpendicular axes, defined as  $y$ . Measurements were made using a digital vernier caliper while mice were conscious.... (Ayers et al. 2010)

The authors then give the estimated volume ( $V_{Xen}$  here for the as-measured xenograft volume) as

$$V_{Xen} = \frac{xy^2}{2} \quad (2)$$

It's a basic exercise in quantitative reasoning to translate this description to the terms of the formula for volume of an ellipsoid. The measurement  $x$  is equal to  $2a$ . The measurement  $y$  is equal to  $2c$ . And  $2c$  is taken to be equal to  $2b$ . This

computation is routinely used for tumor xenografts and introduces several new sources of error.

## Pi Is Not Three

Even if the measurement  $x$  is exactly equal to  $2a$ , and  $y$  is exactly equal to both  $2b$  and  $2c$ , the formula (1) above for volume gives

$$V_{true} = \frac{4\pi}{3} abc = \frac{4\pi}{3} \left(\frac{x}{2}\right) \left(\frac{y}{2}\right) \left(\frac{y}{2}\right) = \frac{\pi}{3} \left(\frac{xy^2}{2}\right),$$

whereas, recall, the xenograft volume using the formula described in Ayers et al. gives

$$V_{Xen} = \frac{xy^2}{2}$$

for the as-measured value (as if  $\pi = 3$ ).

As a point of cultural, if not mathematical, literacy, the number  $\pi$  is not equal to 3. For the example in Table 1, where  $a = 2$ ,  $b = c = 1$ , the true volume is 8.3878 and the estimated volume using formula (2), with  $x = 4$  and  $b = 2$  is 8. The xenograft calculation yields a relative error of  $-0.04507$ , or in other words, an underestimate by about 4.51% of the true volume of the tumor.

It may seem silly to set  $\pi = 3$  when we all know otherwise, but students can benefit from a discussion of whether it is really all that unreasonable. After all, we round off numbers all the time and hope the consequences are negligible. A calculation will show that a  $-4.51\%$  error has been made in rounding off  $\pi$  to 3, which has resulted in a  $-4.51\%$  error in the volume calculation.

## An Ellipsoid Need Not Have Circular Cross Section

Visualizing the process of measuring diameters of a tumor trapped under the skin of a mouse, it is easy to see that the depth of the tumor is impossible to measure with any accuracy. Ayers et al. (2010) recognize this and point out that "... caliper measurements are commonly collected along the longest two dimensions of the tumor  $xy$ -plane only, with the  $z$ -axis dimension assumed to be the same as the shortest dimension."

As the skin of the mouse is putting pressure on the tumor, the  $yz$ -cross section could very well be flattened so that the depth is reduced and the  $y$  measurement is expanded slightly, as when a balloon is pressed downward under a flat pan. So even if the tumor were inclined to grow symmetrically under the skin, the resulting pressure is likely to distort the geometry slightly. All of these considerations cause one to expect that the unmeasurable third axis (denoted  $c$ ) is likely to be overestimated by the measurement " $y$ ." Table 2 describes these errors in terms of a series of tumors with round and flattened cross sections.

**Table 2.**  
**Comparison of Errors Due to Lack of Circular Cross Section of Ellipsoid**

Quantities	Longest $a$	Second longest $b$	shortest $c$	$V$ (eq. 1) $\frac{4\pi}{3}abc$	longest $x = 2a$	Second longest $y = 2b$	$V_{xen}$ (eq. 2) $\frac{xy^2}{2}$	error $\frac{V_{xen} - V}{V}$
no compression	2	1	1	8.38 @&	4	2	8	-4.51% #
compression	2	1	0.9	7.54	4	2	8 @	6.10% ##
compression with elongation along y-axis	2	1.1	0.9	8.29 &	4	2.2	9.68	16.71% ###

# Error due to setting  $\pi = 3$ .

## Additional error due to overestimating  $c$ . Compensates a bit for the error due to setting  $\pi = 3$ . Compare the values denoted with @.

### If a xenograft is both somewhat compressed in the  $z$ -axis and also elongated in the  $y$ -axis, a larger overestimate will occur. Note that the true volume ( $V$ ) is close to that of the example with circular cross section. Compare values denoted with &.

Geometric visualization is considered part of quantitative literacy. This discussion offers an entry into a discussion of how the dimensions  $a$ ,  $b$ , and  $c$  will change if the ellipsoid gets “flatter” or “elongated.” Simplifying assumptions, such as the one that yields  $V_{xen}$ , introduce errors. Students could be asked for other examples in science where such assumptions are routinely made.

## The Result of Measurement Errors Using Calipers

For this discussion, we assume that the true dimensions of the xenograft are those of a slightly flattened ellipsoid with  $a = 2$ ,  $b = 1$ ,  $c = 0.9$ . These dimensions give a true volume of 7.54 using equation (1). In addition to the errors described above, the calipers themselves do not give a true measure of  $x$  or  $y$ . Ayers et al. (2010) describe the problem: “Caliper measurements of subcutaneous xenografts are affected by contributions to the measure from epidermis and adipose tissue, as well as fur if present, each of which introduces error.”

Epidermis and other tissue tend to make the measurements larger rather than smaller. Table 3 shows the result of errors of this sort with respect to the true volume of the hypothetical ellipsoidal tumor. Just a 10% error in measurement of the two axes of the ellipsoid, coupled with the approximations made by equation (2) yield a spectacular 41% error in tumor volume calculation.

From Table 3 it is also easy to see that a 10% error in measuring the short axis results in a larger volume error than a 10% error in measuring the long axis, even though the absolute size of these errors might lead one to expect the reverse. That is, an error of 0.2 in measuring the longest axis results in a smaller error in estimated volume than an error of 0.1 in measuring the shorter axis. On the one hand, students could use algebra to understand how this happens. On the other

hand, beginning a lesson with this phenomenon might be an excellent hook for the topic of error estimation.

**Table 3.**  
**The Result of Measurement Error on Volume Calculations for Xenografts**

True volume (eq. 1) $V = 7.54$	Longest $a$	Second longest $b$	Shortest $c$	$V_{xem}$ (eq. 2) $\frac{xy^2}{2}$	error $\frac{V_{xem} - V}{V}$
True dimensions	2	1	0.9	8	6.10%
10% error in $b$	2	1.1	0.9	9.68	28.38%
10% error in $a$	2.2	1	0.9	8.8	16.71%
10% error in both	2.2	1.1	0.9	10.65	41.22% #

# Note that a 41% error in estimating tumor volume results from only two measurement errors of size 10%. However, these errors interact not only with each other, but also with an error due to the assumption that the shortest axis is as long as the intermediate one, and also the (somewhat compensating) error stemming from taking  $\pi = 3$ .

## The Xenograft Might Not Be an Ellipsoid At All

It is worth including a cautionary note about careful reading of research papers here. It is certainly possible that, faced with a jiggling blob under the skin of a mouse and armed with a pair of calipers, the researcher cannot help but conclude that the tumor mass cannot possibly be described as an ellipsoid of any sort. The researcher may still take caliper measurements and use equation (2) for the xenograft volume. An honest researcher will note somewhere that the answer is questionable. For example, in the paper of Ayers et al. (2010), a table of volumes is reported. Those tumors that do not appear to be ellipsoidal are clearly denoted with an asterisk. However, the “caliper volume” given by the xenograft volume formula is still reported. It is an interesting statistical question as to whether the formula does significantly worse in estimating volumes of non-ellipsoidal tumors.

## Actual Volume Error

When looking at examples of this sort, students should always be invited to exercise their imaginations. In this case the obvious question is, “How would you come up with a better way to estimate tumor volume, so you could see exactly how bad your estimates are?” This question could come after soliciting general opinion on the question, “Do you think a 41% error ever really happens using the caliper technique?” Of course, Ayers et al. also wanted to know how bad caliper estimates were, as they were comparing them with another technique that uses ultrasound. So they needed an independent and trustworthy assessment of the actual tumor volumes in each mouse. Here is what they did (Ayers et al. 2010):

Animals were sacrificed immediately following ultrasound imaging and xenograft tumors excised and stripped of non-tumor tissue if present. Tumor mass has been shown to directly correlate with volume measured by water displacement . . . Mass was determined to the nearest 0.1 mg using a calibrated analytical balance. Xenograft tumor volume was calculated from tissue mass assuming a



density of  $1 \text{ mg/mm}^3$ . This value was used as the true tumor volume (TTV) for comparison purposes.

A summary of their results for only those tumors that were considered ellipsoidal is in Table 4.

**Table 4.**  
**Xenograft Volume and True Tumor Volume from Ayers et al. (2010)**

Data using caliper, $V_{Xem}$ , $\text{mm}^3$	“true” tumor volume, $\text{mm}^3$	error, $\frac{V_{Xem}-TTV}{TTV}$ , %
602	471	27.81
658	550	19.64
1231	906	35.87
924	996	-7.23
1061	1130	-6.12
1576	1755	-10.20
2335	2061	13.29
2024	2322	-12.83

The percent error averages 7.53% with a standard deviation of 19.00%. Researchers who are attempting to fit a model to data often average the data over many “experiments” or, in this case, mice. This data set suggests that doing so would reduce errors in the data considerably. The large standard deviation added to the average suggests that an error of 27% would not be unusual in these estimates. The average percent error of a positive 7.53% indicates the presence of a bias toward overestimating tumor volume using the caliper method, which is consistent with the comments made by Ayers et al. (2010) about the presence of extra tissue such as skin or fur between the calipers.

Another interesting point about this study is that the percent error seems to be smaller for larger tumors. This tendency also highlights the difference between absolute error and percent error. As the tumor gets larger, the small amount of skin and fat tissue between the tumor and the calipers may represent a smaller percentage of the total tumor length measurement being taken. Students could be invited to investigate the results of a single measurement error of  $5 \text{ mm}^3$ , for example, at the various tumor sizes reported in this study.

## Summary

Sometimes our job as instructors is to take complicated questions and pull them apart into smaller, more manageable pieces. But it is also important to see how those pieces fit together to create a coherent picture of some bit of scientific inquiry. In the example presented here we see how a simple question, “How big is the error?” may be answered through an elementary statistics calculation. However, the question “Why is the error so big?” is much deeper and draws on a variety of estimation skills and the algebra that relates them to each other.

Caliper estimates of tumor size have errors that come from measurement error, which is to be expected. They also have errors coming from assumptions

about shape, and even nontrivial errors due to round-off (because, friends,  $\pi$  will never be three). Because the formula for volume is multiplicative, errors don't add up in a straightforward way. Some errors make a bigger difference than others in the final estimate of tumor volume, and algebra can tell you why that is so. Some errors tend to be biased in one direction, as is likely the case with calipers. Absolute error is not the same as percent error, and this fact shows up in the data.

At the most basic level, discussions of estimation skills often focus on estimates of computation. About how big will  $3.559 \times 4.2$  be? Surely we want students to see easily that it will be between 12 and 20. But in the sciences, estimations are made all of the time. The results of computing with those estimates are important, by which I mean that other people use them in their own research with the intention of furthering knowledge. So, errors matter, and estimation of error size requires more thoughtfulness than is suggested by the usual examples of estimates in mental arithmetic. Although more sophisticated, the tumor volume example is still within the bounds of quantitative reasoning and numeracy, defined as “the application of basic mathematics skills, such as algebra, to the analysis and interpretation of real-world quantitative information in the context of a discipline or an interdisciplinary problem to draw conclusions that are relevant to students in their daily lives” (Elrod 2014).

For science students we need pithy problems with interesting questions and relevant answers, even for something as deceptively straightforward as estimation to be “practiced extensively, across the curriculum, in the context of progressively more challenging problems, projects, and standards for performance,” (Association of American Colleges and Universities 2007). Fortunately, as we have shown here, there is no shortage.

## References

- Association of American Colleges and Universities. 2007. *College Learning for the New Global Century*. Washington, DC: Association of American Colleges and Universities. <https://aacu.org/publications-research/publications/college-learning-new-global-century>
- Ayers, Gregory D., Eliot T. McKinley, Ping Zhao, Jordan M. Fritz, Rebecca E. Metry, Brenton C. Deal, Katrina M. Adlerz, Robert J. Coffey, and H. Charles Manning. 2010. “Volume of Preclinical Xenograft Tumors Is More Accurately Assessed By Ultrasound Imaging Than Manual Caliper Measurements.” *Journal of Ultrasound in Medicine* 29 (6): 891–901. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2925269/>.
- Elrod, Susan. 2014. “Quantitative Reasoning: The Next ‘Across the Curriculum’ Movement.” *Peer Review* 16 (3): Article 2. <https://www.aacu.org/peerreview/2014/summer/elrod>.