## Determination of the Area Under a Curve

In a recent article, Tai (1) describes a method to determine total area under metabolic curves. However, what is exaggerately called "Tai's mathematical model" is nothing but a simple geometrical formula, well known for many years as the trapezoidal rule. This classical method, as well as a series of other approaches, was reviewed and investigated by Wagner and Ayres (2) 17 years ago. Moreover, the derivation of the trapezoidal rule is presented in a circumstantial way, the final equation called "Tai's formula" contains incorrect notations (e.g., 'x̲' must be 'X' with the author's definitions), and the division into different conditions of intercept and passing the origin is absolutely unnecessary.

The validation of the formula by means of comparison with a "true value" is useless and contains several fallacies. First, because of the geometrical interpretation of the trapezoidal rule, it is clear that the expression tends toward the true area under the curve (AUC) if the number of considered curve points increases. Hence, the adequacy of the trapezoidal rule is dependent on the number of curve points and cannot be investigated by a few examples. Second, the AUC value measured graphically by counting the numbers of small units under the curve is not the true AUC value. Like the trapezoidal rule, it is an approximation, which tends toward the true value if the units decrease. Thus, for comparison, not the true but another approximation was used. Third, Student's $t$ tests were misused. Significance tests are generally inadequate tools for comparison of two methods of measurement (3). In addition, the sample size was only $n = 5$, resulting in very low power, and multiple comparisons were made without adjustment. However, even if the sample size had been larger and adjustment for multiple comparisons had been made, in principle, approaches adequate for method comparison should have been used (3).

Finally, the term *total area under a curve* is used in another sense than it is in the pharmacokinetic literature. The word *total* refers to AUC(0-∞), whereas AUC(0-*T*), where *T* is the investigator's last time point, is a partial area. Only the latter can be estimated by means of the trapezoidal rule; computation of the total AUC(0-∞) requires a mathematical or pharmacokinetic model (2). However, "Tai's mathematical model" is no model, it is an application of a simple geometrical rule.

In conclusion, Tai proposed a simple, well-known formula exaggerately as her own mathematical model and presented it in a circumstantial and faulty way.

RALF BENDER, PHD

From the Department of Nutrition and Metabolic Diseases, Heinrich-Heine University, Düsseldorf, Germany.

Address correspondence to Ralf Bender, PhD, Department of Nutrition and Metabolic Diseases, Heinrich-Heine University, P.O. Box 10 10 07, D-40001 Düsseldorf, Germany.

References

1. Tai MM: A mathematical model for the determination of total area under glucose tolerance and other metabolic curves. *Diabetes Care* 17:152–154, 1994
2. Wagner JG, Ayres JW: Bioavailability assessment: methods to estimate total area (AUC 0-∞) and total amount excreted ($A_e^{\infty}$) and importance of blood and urine sampling scheme with application to digoxin. *J Pharmacokinet Biopharm* 5:533–557, 1977
3. Altman DG, Bland JM: Measurement in medicine: the analysis of method comparison studies. *Statistician* 32:307–317, 1983

## Comments on Tai's Mathematic Model

I commend Tai (1) for producing a correct method for calculating the total area under the curve. It uses the trapezoid rule, a basic geometrical concept, which is that the area of a trapezoid is the mean of the length of the two parallel sides times the width. This method has been used by those of us in the field for many years and, in my opinion, does not need a new name. I also have a number of other problems with her paper. Tai considers that the "true value" for the area under the curve is obtained by plotting the curve on graph paper and counting the squares under the curve. This method is subject to a number of errors arising from inaccuracies in plotting the points and lines and in estimating the area of the portions of squares that are bisected by lines whose width is large in relation to the size of the squares. The trapezoid rule is, in fact, the gold standard for calculating areas if the points are joined by straight lines.

The typographical error in the example calculation (which should read: area = 1/2[30(95 + 147) + 30(147 + 124) + 30(124 + 111) + 30(111 + 101)] = 14400) is a problem I cannot criticize. In one of my papers, there are a number of confusing errors in the section describing the effects of different ways of calculating the area under the curve that I was careless enough not to pick up in proof (2).

However, I will criticize her totally inappropriate use of "my" formula to calculate total area under the curve (3). As was clearly stated, my formula is for calculating the incremental area under the curve above the baseline and does not give the correct value for the total area. Therefore, her comparison of the accuracy of "her" method with "mine" is a completely meaningless exercise. In addition, to obtain the area she ascribes to my method (i.e., 13,517), she must have used the incorrect final term $tD^2/[2(D+|E|)]$, which, as explained, is only substituted for $t(D+E)/2$ if increment $D$ is positive and $E$ negative (i.e., below the baseline) so as to ignore the area below the baseline. If $D<0$ and $E>0$, then the term becomes $tE^2/[2(E+|D|)]$. In her example, none of the postprandial points is less than fasting.

Finally, she says that my method only permits a single time interval. She is wrong, in that my method is based on the trapezoid rule and can be adapted for different time intervals, a point we made in an earlier and more complete description of the method (4). However, she is correct that the sample simplified formula in our 1991 paper (3) is only appropriate for equal time intervals. We made an error in assuming that readers would be able to modify it for different time intervals. Thus, the formula for the incremental area under the curve ignoring area below the baseline, where $x_1 .. x_n$ are the increments (i.e., the postprandial values minus the fasting value), and if all the increments are positive, is

$$\sum_{i=1}^{n} t_i(x_i + x_{i+1})/2$$

where $t_i$ is the time interval between the ith and $i+1$th points. However, if either $x_i$ or $x_{i+1}$ is negative (i.e., below the baseline), then one of the terms described in the preceding paragraph is substituted (if both are negative, then the area between them is 0).

The lesson here is that calculating areas under the curve is deceptively difficult. I fear I may be responsible for contributing to the confusion.

Thomas M. S. Wolever, DM, PhD

From the Department of Nutritional Sciences, University of Toronto, Toronto, Ontario, Canada.

Address correspondence to Thomas M. S. Wolever, DM, PhD, University of Toronto, Department of Nutritional Sciences, Fitzgerald Building, 150 College Street, Toronto, Ontario, Canada M5S 1A8.

**References**

1. Tai MM: A mathematical model for the determination of total area under glucose tolerance and other metabolic curves. *Diabetes Care* 17:152–154, 1994
2. Wolever TMS: The glycemic index. In *World Review of Nutrition and Dietetics: Aspects of Some Vitamins, Minerals and Enzymes in Health and Disease*. Vol. 62. Bourne GH, Ed. Basel, Karger, 1990, p. 120–185
3. Wolever TMS, Jenkins DJA, Jenkins AL, Josse RG: The glycemic index: methodology and clinical implications. *Am J Clin Nutr* 54:846–854, 1991
4. Wolever TMS, Jenkins DJA: The use of the glycemic index in predicting the blood glucose response to mixed meals. *Am J Clin Nutr* 43:167–172, 1986

## Tai's Formula Is the Trapezoidal Rule

We were disturbed to read the article by M. M. Tai titled "A Mathematical Model for the Determination of Total Area Under Glucose Tolerance and Other Metabolic Curves" (1). The author seems to claim "Tai's formula" as a new method of computing area under a curve. The formula given is simply the trapezoidal rule, published in many beginning calculus texts (for example, see Swokowski [2] or Faires and Faires [3]). Although we do not have a first reference, it is our understanding that the trapezoidal rule was known to Isaac Newton in the 17th century. Further, her article omitted any reference to the magnitude of error of the area approximation when the true curve is unknown, as is the case for measuring glucose tolerance.

The trapezoidal rule is used in undergraduate calculus courses to illustrate and develop the calculus of definite integrals. Calculus students begin estimating area under a known curve by dividing the x-axis into small intervals and totaling the area of the resulting trapezoids. The exercise demonstrates that the error in the area calculation decreases as the length of the x-axis intervals is decreased. Definite integrals are then defined by taking the limit of the trapezoid's summation as the x-axis intervals go to zero.

Shortcomings exist with the trapezoidal rule, even if the true curve is known, that are not mentioned in the article. Modeling a curve by a series of connected line segments will either over- or underestimate the actual area, depending on the direction of curvature in the true curve. In the case of the glucose tolerance response, the true curve is unknown; even so, the trapezoidal rule is the best possible approximation of the area based on linear segments given minimal assumptions about the true curve. Most statements of the trapezoidal rule include the upper bound of the possible error stated as $M(b - a)^3/12n^2$, where $M$ is the maximum rate of curvature over the $x$ segment, $[a, b]$, and $n$ is the number of x-axis intervals.

Tai stated that her "standard (true value). . . is obtained by plotting the curve on graph paper and counting the number of small units under the curve" (p. 153). By definition, the sum of the area of the small units, which she erroneously refers to as the "true value," should be exactly the area found by the trapezoidal rule. The formula should have been 100% accurate because she defined truth to be exactly the sum of the area of the graph paper trapezoids.

We hope that our comments as mathematics and statistics practitioners help to clarify the origin of the trapezoidal rule and its properties as an approximation of the area under the true curve.

Jane H. Monaco, MS
Randy L. Anderson, PhD

From the Department of Public Health Sciences, Bowman Gray School of Medicine, Wake Forest University, Winston-Salem, North Carolina.

Address correspondence to Randy L. Anderson, MD, Department of Public Health Sciences, Bowman Gray School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157.

Acknowledgments— The authors are funded in part by National Institutes of Health NHLBI Grant NIH 2-194-811-9443; NINDS Grant NIH NS 20618-08; and NINDS Grant NIH 1PO1-NS-27500-02.

References

1. Tai MM: A mathematical model for the determination of total area under glucose tolerance and other metabolic curves. *Diabetes Care* 17:152–154, 1994
2. Swokowski EW: *Calculus with Analytic Geometry*. 3rd ed. Boston, MA, Prindle, Weber and Schmidt, 1983, p. 260–261
3. Faires JD, Faires BT: *Calculus*. 2nd ed. New York, NY, Random House, 1988, p. 497–498

# Modeling Metabolic Curves

While not denigrating the Tai model (1) for the total area under metabolic curves, nor denying its validity, we believe the Γ variate function also deserves to be better known. It cannot only be used to approximate the total area under the curve (AUC) but also to describe these curves functionally with estimates of their characteristic parameters and to separate their secretion and clearance phases. The formula for this function is

$$(1) \qquad y = y_0 + At^a e^{-bt}$$

where $y \varepsilon$ {insulin, C-peptide, glucose concentrations}, $y_0$ is a basal value, $t$ is time, and $A$, $a$, and $b$ are found by simply fitting the data to the straight line form of (1):

$$(2) \qquad \ln(y - y0) = \ln A + a\ln t - bt$$

in which ln($x$) represents the natural logarithm. By differentiating $y$ with respect to $t$, one can find the secretion and clearance rates as the first and second terms respectively of

$$(3) \qquad dy/dt = Aat^{a-1}e^{-bt} - Abt^a e^{-bt}$$

The total above basal area under the relevant response curve, AUC, is then found by integrating (1) with respect to $t$ between the limits of 0 and ∞ to obtain

$$(4) \qquad AUC = A\Gamma(a + 1)/b^{a+1}$$

in which Γ($x$) represents the Γ function, values for which are in standard tables (2) (hence the name of the function).

An example of the use of the Γ variate function in this context may be found in Shannon et al. (3). The Γ variate model does not require seeding with initial values, nor does it require many sampling times to increase its accuracy.

A. G. Shannon, PhD
D. R. Owens, MD

From the School of Mathematical Science, University of Technology, Sydney, Australia.
Address correspondence to A. G. Shannon, PhD, University of Technology, Sydney, P.O. Box 123, Broadway, New South Wales 2007, Australia.

References

1. Tai MM: A mathematical model for the determination of total area under glucose tolerance and other metabolic curves. *Diabetes Care* 17:152–154, 1994
2. Abramowitz M, Stegun IA: *Handbook of Mathematical Functions*. New York, Dover, 1965
3. Shannon AG, Ollerton RL, Owens DR, Luzio S, Wong CK, Colagiuri S: Mathematical model for estimating pre-hepatic insulin secretion from plasma C-peptide (Abstract). *Diabetes* 40 (Suppl. 1):13A, 1991

# Reply From Mary Tai

I would like to thank all of the readers who have reviewed and responded to the publication of Tai's model. I am particularly grateful to those who have confidence in my intention of publishing and have considered Tai's model an effective tool in calculating total area under a metabolic curve. However, three of the readers expressed their concerns on the following issues. My replies are presented as follows.

## To Dr. Bender

**The originality of Tai's model.** While a doctoral candidate working on my dissertation at Columbia University in 1981, I needed to calculate total area under a curve. During a session with my statistical advisor, and after examining several alternative methods, I worked out the model in front of him. The concept behind it is obviously common sense, and one does not have to consult the trapezoid rule to figure it out. The trapezoid rule is really not Nobel Prize material, such as the double helix or jumping genes. I also used the formulas to calculate the areas of a square or a triangle without knowing whose rules were being followed. Fortunately, I do not have to answer that for you.

**Why I call it Tai's model.** I never thought of publishing the model as a great discovery or accomplishment; it was not published until 14 years later, in 1994. Because of its accuracy and easy application, many colleagues at the Obesity Research Center of St Luke's-Roosevelt Hospital Center and Columbia University began using it and addressed it as "Tai's formula" to distinguish it from others. Later, because the investigators were unable to cite an unpublished work, I submitted it for publication at their requests. Therefore, my name was rubber-stamped on the model before its publication.

According to *Merriam Webster's Dictionary*, a model can be defined as "a type of design of product;" "a description used to visualize something that cannot be directly observed;" or "a system of postulates, data, and inferences presented as a mathematical description of an entity. . . ." Even if Tai's model were based on the trapezoid rule concept, according to the definition of a model, I have worked out a "design" (mathematical expression) for the "structure units" (individual areas) on my own. In other words, I have presented the original concept into a functioning mathematical description that can be easily observed and applied. Following the above definition, I therefore carefully named the mathematical description as Tai's "model" rather than "formula" to indicate that I have used existing formulas for small area calculations.

My intention in publishing the model is therefore to share, rather than to gain honor or glory with its publication, because there is none. Many other investigators probably thought about the same thing, but maybe they did not bother to follow up or produce a model (or the same model). You indicated that I probably did work this out on my own and I am grateful for your "probability," because I did indeed do so with a witness present. Maybe I can address the model as my creation based on fact rather than your doubtful "probability." Besides, if I do not address the model as "Tai's," other investigators who wish to cite it will.

**The precision of Tai's model.** Because Tai's model is based on the calculations of individual squares and triangles, its precision is obviously absolute. You are correct in saying that I have verified the validity of the formula by comparison with its approximation, meaning counting squares.

**The size of *n*.** Following the statistical principle that you consider elementary, it is correct that *n* does represent numbers of data sets. However, in this case, elemental principle simply does not apply. The hypothesis here is the validity of the formula. The acceptance or rejection of the hypothesis is not based on the findings of each individual data set, as is a general rule in an experimental study. It should not be difficult to see that each data set here represents the findings from its respective research protocol and answers its individual research questions rather than answering the validity of Tai's model. Furthermore, because the same formula was used for each data set, the degree of accuracy on the resultant total area obtained will be exactly the same for each set. Therefore, increasing *n* of the data set does not increase statistical power as you suggested.

I introduced other formulas simply for the purpose of comparison. Because the formula cannot be compared with its approximation and there are limited formulas available, I decided to count, because every published curve has been based on counting squares. I also believe, if one increases the N I am talking about, meaning the number of methods, one can better verify the validity of Tai's model.

### To Dr. Wolever

After receiving your recent graphic representation of your formula, I began to realize that I have indeed misunderstood your formula as some other readers did. Your incremental area is the area above the baseline rather than the total area under the curve including the baseline area. I apologize for the misapplication of your unique formula, which I do fully support. I also acknowledge that you, too, have indeed used the concept of adding triangles and rectangles in your mathematical model for the total increment. I also appreciate your idea of weighing, because I did weigh the total area under an arc and, as you know, that might be the only way.

### To Dr. Anderson and Ms. Monaco

Tai's model is designed to calculate total area under a metabolic curve that is plotted by connecting experimental points $x_i$, $y_i$ with straight lines as shown in Fig. 1 in my article. Because the metabolic curve is not an arc, the exact area can be calculated without assumption and approximation. If a smooth arc represents the true curve, it is obtainable only when $x_i \rightarrow \infty$ and $\Delta x \rightarrow 0$, as presented in the trapezoid rule or calculus, and it is virtually impossible in an experimental condition.

Finally, I would like to correct some typographical errors in my article:

On p. 153, the correct formula should be

$$\text{Area} = \frac{1}{2}\sum_{i=1}^{n} X_{i-1}\,(y_{i-1} + y_i)$$

and in example 1:

$$\text{Area} = \frac{1}{2}\,30\,[(95 + 147) + (147 + 124) + \ldots]$$

MARY M. TAI, MS, EDD

From the Obesity Research Center, St Luke's-Roosevelt Hospital Center, New York; and the Department of Nutrition, School of Education, New York University, New York, New York.

Address correspondence to Mary M. Tai, MS, EdD, Obesity Research Center, Women's Hospital, 10th Floor, St. Luke's-Roosevelt Hospital Center, 114 Street at Amsterdam Avenue, New York, NY 10025.

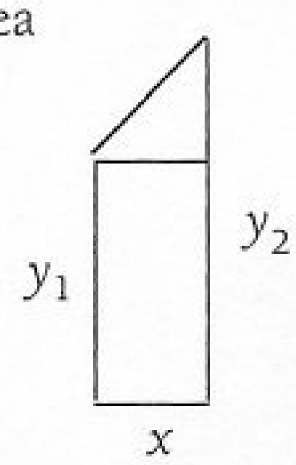
## Addendum to Monaco's and Anderson's Letter

Tai responds that her formula is based on the sum of the areas of small triangles and rectangles and is not based on the sum of the areas of trapezoids (the trapezoidal rule). As is evident in the following figure and algebra, the small triangle and the contiguous rectangle form a trapezoid. The sum of the area of the triangle and the rectangle is the area of the trapezoid. Using her notation,

triangle area + rectangle area

$$= \frac{1}{2}\underline{x}(y_2 - y_1) + \underline{x}y_1$$
$$= \frac{1}{2}\underline{x}(y_2 - y_1 + 2y_1)$$
$$= \frac{1}{2}\underline{x}(y_2 + y_1)$$
$$= \underline{x}\bar{y}$$

= trapezoid area.




The trapezoid area is the mean of the length of the parallel sides, $\bar{y}$, times the width. Summing over all trapezoids under the curve yields the trapezoidal rule, the expression listed as Tai's formula in Tai's article.

Tai invites readers to "do a small problem using any existing geometric concept(s) you prefer and without using the geometric concept behind Tai's formula." We prefer no other method to the trapezoidal rule; rather, our goal is to point out that Tai's formula *is* the trapezoidal rule, as shown above.

JANE H. MONACO, MS
RANDY L. ANDERSON, PHD

From the Department of Public Health Sciences, Bowman Gray School of Medicine, Wake Forest University, Winston-Salem, North Carolina.

Address correspondence to Randy L. Anderson, PhD, Department of Public Health Sciences, Bowman Gray School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157.

T

**Acknowledgments**— The authors are funded in part by National Institutes of Health, NHLBI Grant NIH 2-194-811-9443; NINDS Grant NIH NS 20618–08; and NINDS Grant NIH 1PO1 NS 27500–02.