Title:

A Tool to Explore Discrete-Time Data: The Time Series Response Analyser

Name(s) of Author(s):

Benjamin J. Narang[1], Greg Atkinson[2], Javier T. Gonzalez[1], James A. Betts[1]

Institutional Affiliation:

[1]Department for Health, University of Bath, Bath, UK.

[2]School of Health and Life Sciences, Teesside University, Middlesbrough, UK.

Running Head:

Time Series Response Analyser

Corresponding Author:

James A. Betts, Department for Health, University of Bath, BA2 7AY, United Kingdom.

E-mail: j.betts@bath.ac.uk

Abstract

The analysis of time series data is common in nutrition and metabolism research for quantifying the physiological responses to various stimuli. The reduction of many data from a time series into a summary statistic(s) can help quantify and communicate the overall response in a more straightforward way and in line with a specific hypothesis. Nevertheless, many summary statistics have been selected by various researchers, and some approaches are still complex. The time-intensive nature of such calculations can be a burden for especially large datasets and may, therefore, introduce computational errors, which are difficult to recognize and correct. In this short commentary, we introduce a newly-developed tool that automates many of the processes commonly used by researchers for discrete-time series analysis, with particular emphasis on how the tool may be implemented within nutrition and exercise science research.

Keywords

Incremental area under the curve; time series data; temporal response; post-prandial.

Introduction

It is common practice within the field of nutrition and metabolism research to analyses serial measurements made over time to determine the temporal pattern of a given response. Typical examples include metabolic control following nutritional challenges (i.e. oral glucose or fat tolerance tests; Berthiaume & Zinker, 2002), monitoring of stable isotope enrichment in various body pools and associated substrate kinetics (Garlick et al., 1989), and markers of physiological response to exercise such as heart rate and oxygen consumption (Gore & Withers, 1990).

Such analyses have become increasingly complex and necessary in recent years both due to technical advancements in measurement tools and due to our growing understanding of the interactions between various nutritional stimuli. Regarding the former, it is undoubtedly a mark of progress that modern technologies have enabled many measurements to be made with higher sampling frequency and thus with greater sensitivity to rapidly fluctuating responses over time. However, such high-resolution temporal data also bring certain analytical challenges (such as the control of type I and II error rates due to the number of multiple comparisons), which can complicate the elucidation and communication of clear conclusions.

While early studies in many areas of nutrition science may have examined simple comparisons of treatments (e.g. 20 g carbohydrate *versus* water/placebo at a single time-point), the state of current understanding in many areas is now such that further progress requires more sophisticated factorial designs with multiple levels within each factor, to examine longer term effects and/or interactions between ingredients that work in concert (e.g. pre-post response to carbohydrate *versus* carbohydrate-protein *versus* water/placebo, *etc.*). This further evolution is necessary to detect more subtle and/or context specific effects but, again, introduces additional

complicating factors, such as the reduced statistical power associated with quantifying interactive effects between all the additional independent variables (e.g. a 3-way ANOVA: 3 conditions*pre-post*multiple time-points), along with the complications arising when the data violate the assumption of sphericity (Huck & Cormier, 1995).

In all the above cases, condensing the time series data down to a summary statistic can simplify the analysis by removing the temporal element. In the above example, the 3-way ANOVA with multiple comparisons at many time-points becomes a 2-way condition*time (pre, post) analysis. Beyond these advantages in relation to statistical analyses, this approach of using summary statistics facilitates the clear communication of the main findings both in simple terms for the general public and with complete reporting of individual responses for the scientific community. For example, graphical presentation of time series data on a line graph does not readily allow for individual or paired responses to be plotted, whereas this consistency of observed responses is easily presented as a histogram showing individual summary statistics (**Figure 1**). Measures of central tendency certainly have a place to illustrate group effects on graphs and figures but individual responses to each experimental condition should still be presented, particularly when sample sizes are relatively small, to facilitate critical evaluation of data (Weissgerber et al., 2015).

Despite the above benefits of summary statistics and the common use of time series experimental designs within the scientific literature, the general approaches and precise methods of analysis vary considerably between laboratories and experiments (Wolever, 2004; Matthan et al., 2016). In addition, calculations requiring multiple stages and various equations are time consuming and susceptible to human error. This short commentary introduces a downloadable spreadsheet, the Time Series Response Analyser (TSRA), designed specifically

to automate and standardize many common processes, thus minimizing both the time spent analyzing data and the probability of computational errors. The TSRA is freely available under the 'Author Guidelines' section of the *IJSNEM* website (https://journals.humankinetics.com/view/journals/ijsnem/ijsnem-overview.xml/). This commentary will highlight a range of time series analysis procedures that can be computed with the tool, and briefly discuss their utility in the context of exercise and nutrition research.

Area under the curve (AUC)

The methodological approach to an AUC calculation is particularly variable (Wolever, 2004) and manual calculation is highly susceptible to human error. The AUC can be calculated using denominations of the trapezoidal rule, where time series data are integrated to form a single value characterizing the overall response, representative of an area (e.g. blood glucose concentrations measured in $mmol \cdot l^{-1}$ at serial time-points over a standard oral glucose tolerance test are expressed as the product of concentration and time; $mmol \cdot l^{-1} \cdot 120$ min). **Figure 2** illustrates a range of AUC options, each of which is described in this section.

Total AUC is the most straightforward approach, in which an area is calculated relative to the line representing an ordinate of zero (Matthews et al., 1990). This practice can provide a valid estimate of the overall exposure to the parameter of interest (i.e. including the value measured at baseline – e.g. if contrasting 24 h plasma testosterone concentrations between males and females). However, by the same reasoning, total AUC can be limited by the variation commonly observed at baseline, despite the best efforts of researchers and participants to replicate experimental conditions (Altman, 1985). In cases where baseline differences are apparent and/or it is the response to a stimulus that is of primary interest, the incremental AUC

106 relative to another nominal value (generally baseline) may be a more appropriate alternative
107 (Wolever & Jenkins, 1986).
108
109 Naturally, certain exposures can cause the dependent variable to drop below the value
110 to which incremental AUC is being calculated. For example, the postprandial response to a
111 standard oral glucose tolerance test is typically measured across two hours, as the blood glucose
112 concentrations of healthy participants tend to return to baseline within this time period (Babraj
113 et al., 2009). Therefore, the blood glucose concentrations of highly insulin sensitive individuals
114 could feasibly fall below the value measured at baseline, which for an incremental AUC
115 calculation provides multiple options for analysis. In this instance some researchers may
116 choose to terminate the calculation at the time-point at which the measured value falls below
117 the incremental reference value (Ha et al., 1992), while others will include any subsequent
118 positive segments if the value returns above baseline. Within this latter approach, researchers
119 could consider negative areas to equal zero (Hofman et al., 2004), or subtract them from the
120 calculation (Gannon et al., 1989). It should be noted that, while the subtraction of negative
121 areas follows the principle of mathematical integration, this process is rarely justified but may
122 occasionally be applied in error. In theory, unless this subtractive process is rationalized, values
123 representing AUC should always be positive. Moreover, some of the incremental AUC
124 variations can be applied to the nadir rather than the baseline value (Vorster et al., 1990), which
125 may be of interest when variables tend to decrease in response to a stimulus, such as
126 postprandial concentrations of non-esterified fatty acids (Bickerton et al., 2007), or the 'hunger
127 hormone' ghrelin (le Roux et al., 2005). Alternatively, the AUC could be calculated relative to
128 a pre-determined absolute value or clinical reference threshold that is indicative of a certain
129 outcome (Monnier et al., 2003). It is beyond the scope of this commentary to discuss each of
130 these methodologies in any greater detail as they ultimately depend on the context. Suffice to

say, whilst some AUC calculations are relatively simple, others can become mathematically complex, particularly those that consider the intersection of certain thresholds. In these instances, the probability of conceptual and computational errors with manual calculations are increased, and the clarity with which the AUC values have been derived is reduced.

The TSRA generates AUC results from raw data consistently and instantaneously with a minimal risk of human error. The tool computes AUC for all treatments simultaneously and handles each of the aforementioned methodologies under the input of the user. In addition, the spreadsheet provides transparency by explicitly quantifying the segmental areas that combine to produce the chosen AUC (which can be valuable information in itself to retain some reference to the shape of the response curve despite reducing the individual time points into areas).

Alternative summary statistics in discrete-time series analysis

In addition to the AUC calculations computed by the TSRA, the peak and time-to-peak values for each trial are also included in the output. Errors and inconsistencies in the identification of these summary statistics are considerably less likely to occur when compared to AUC, as their definitions are more precise and their calculations are more straightforward. They can however be particularly informative within certain contexts, and they are therefore briefly discussed in this section. **Table 1** contains definitions, benefits, limitations and examples for each summary statistic included in the TSRA output.

Peak

Of the various alternative summary values that can describe a time series response, the absolute peak is an easily identifiable, interpretable and physiologically meaningful statistic. It

is simply the highest value attained in the dependent variable across the time window through which it was measured. Therefore, rather than representing the totality of a response, as is the case with AUC, this value indicates the maximum *measured* value of the relevant outcome. Critically, this statistic should be determined separately for every distinct trial and individual, accepting that the peak value may occur at different time-points for different response curves. Thus, the contrast of maximum measured values cannot be ascertained from visual inspection of the data when plotted as a time series (i.e. it is possible that no single participant's maximum value occurred at the apex of the group mean line). The utility of a peak value during the response to a physiological challenge has been demonstrated in the diagnoses of various medical conditions such as growth hormone deficiency (Koppeschaar et al., 2004) and constitutional delay of puberty (Grinspon et al., 2010), and is practical in the application of diagnostic research due to the absence of any complex calculations. Despite the simplicity of this summary statistic representing a clear benefit of this approach, contextual limitations do exist. For example, measurement error is likely to be relatively high when a single data-point is used to summarize an overall response, and the accuracy is heavily influenced by the true location of a peak value relative to the frequency with which samples are collected (De Nicolao et al., 2000). The accuracy of this value may therefore be questioned when sampling frequency is insufficient and/or the random within-subjects variability or "noise" in the measurement of the dependent variable is high.

Time-to-peak

Alongside the reporting of the peak value, the time at which this peak occurs is typically reported and interpreted by authors. This "time-to-peak" summary statistic indicates the gradient of the response to the stimulus, demonstrating onset alongside magnitude. For example, both the AUC and peak values may be similar between treatments, yet the time-to-

peak may still reveal important changes in the shape of the response curve (**Figure 3**). This may be useful when assessing the bioavailability of a nutrient or supplement, as it can indicate the net rate of appearance relative to an alternative condition (Matthews et al., 1990). For example, Vinson and Bose (1988) included a comparison of a time-to-peak summary statistic when investigating ascorbic acid bioavailability, in response to the ingestion of equivalent doses of synthetic and naturally-occurring vitamin C. Importantly, unless a substance is not endogenously produced and maintains constant disappearance rates, or in the absence of isotopic tracer methodologies, this method provides fairly limited insight into substrate kinetics. However, the utility of the time-to-peak summary statistic as a diagnostic tool has been demonstrated in the context of insulin sensitivity. Specifically, risk-prediction models for prediabetes were shown to be reliably and independently enhanced by the addition of time-to-peak blood glucose concentration during an oral glucose tolerance test (Chung et al., 2017). Moreover, the use of this statistic in this context theoretically signified the early-phase insulin response, which may have provided additional mechanistic insight beyond alternative summary statistics (Cree-Green et al., 2018).

A further application of time-to-peak has been to inform methodologies that seek to identify certain responses, such as the duration and sampling frequency of an oral fat tolerance test necessary to provide a holistic metabolic profile (Tentolouris et al., 2017). As with all considerations outlined in this paper, the precise calculations and reported outcomes should remain specific to the research question and will therefore depend heavily on the context in which time series data are being analyzed. Moreover, where the magnitude and/or timing of the peak is of interest, additional measurements should be taken throughout the time window within which it is expected to occur.

Further considerations

Variability statistics

Another avenue for investigation of time series data is variability. For example, measures of variability in the continuous monitoring of glucose concentrations can be a useful parameter to describe glycemic control (Wijsman et al., 2013). A greater variability in glucose concentration could indicate a reduced ability to appropriately respond to nutritional stimuli, reflecting impaired homeostatic regulation and in the context of glucose metabolism, an increased risk of type-2 diabetes (Ceriello et al., 2008). Within this example, a variety of methods are available to characterize glycemic variability including overall standard deviation, standard deviation across fixed time windows (for variability changes across time), range, interquartile range, percentage coefficient of variation and time spent above/below certain thresholds (Akintola et al., 2015). Rodbard (2009) discussed these methods from a statistical standpoint and provided further context-specific options for alternative perspectives on time series data. Another context in which the variability in a measured marker is of interest within a certain time window is chronobiology. Whilst this is a particularly interesting avenue for time series data analysis in nutrition research, it is beyond the scope of the TSRA primarily because of the circular nature of chronobiological data measured over several biological rhythm periods. The intricacies of biological rhythm descriptions and summaries are discussed from a statistical perspective elsewhere (Landler, Ruxton & Malkemper, 2018). The appropriate application of variability statistics to time series data ultimately depends on the specific research question being addressed, and the information that each option can provide. Further key considerations may be the normality of data distribution, which can influence the appropriateness of certain measures of central tendency and variability, and the associated sensitivity of these approaches to more extreme values. The TSRA computes both the standard deviation and the coefficient

of variation for each individual trial, and provides these simple variability statistics within the standard output. Alternative variability statistics are not calculated by the tool, as the provision of a finite number of complex options may influence the analytical approach taken by the user.

Missing values

Missing values may be the result of missed or inappropriately handled samples, errors in a measurement technique or mistakes during data entry. These can be particularly common in time series data, as the probability of an error is increased when a large number of samples are collected (especially where humans and/or technology are involved!). Missing data pose a problem for the analysis of time series data as the intended temporal resolution within a given trial is transiently reduced. Key considerations include the amount, the pattern and the cause of missing data, each of which may influence the methods by which they are resolved. Regarding the cause, data could be *missing completely at random* (MCAR), where missing values are unrelated to any observed values and are therefore a totally random subset of the data. Alternatively, if missing values are related to observed data, or dependent on the unobserved values themselves, they are considered to be *missing at random* (MAR) or *missing not at random* (MNAR), respectively (Little & Rubin, 1987). Where data are MCAR, techniques typically aim to preserve the observed underlying parameters of the variables for which data are imputed (e.g. means, variances, covariances *etc.*). However, the systematic nature of data MAR and MNAR suggest potential bias may have been introduced in these parameter estimates due to the existence of the missing values. For example, if the accuracy of a measurement technique utilized during time series data collection is confounded outside a certain range, especially high and/or low values are likely to be missing more frequently, eliciting an unrepresentatively skewed distribution (an example of MNAR). Indeed, Bell, King and Fairclough (2014) demonstrated a greater level of bias in time series summary measures with

data MAR or MNAR, compared with MCAR, using a simulated randomized controlled trial. Researchers are therefore recommended to identify the cause of missing time series data and handle this issue accordingly.

Individual time-points for continuous time series data are inherently not mutually exclusive, so it seems appropriate to estimate missing values using known data for a given trial. The precise method by which this process has been conducted may however be ambiguous. As AUC calculations follow the trapezoidal rule, this summary statistic would typically use simplistic linear interpolation to estimate missing values. Briefly, existing points either side of missing values are connected with a straight line, and these are imputed as a function of time using the resulting linear equation (**Figure 4A**). It should be noted that this approach has limitations, particularly if missing values occur where the true response is likely to have reached a peak, as a linear connection would undercut this value (**Figure 4B**). An alternative approach may be to fit a polynomial curve of appropriate order to the known data and impute missing values using the resulting polynomial equation. In the context of time series data, imputing missing values using alternative trials for the same treatment or the same individual are not recommended, as these approaches are likely confounded by inter-individual variability and the effect of treatments, respectively. For a comprehensive review of missing value handling in the context of randomized controlled trials in nutrition, the reader is directed to Li and Stuart (2019).

Outliers

Another contentious topic in the initial screening of data is the identification and subsequent handling of outliers. Outlier identification typically uses statistical approaches, such as Tabachnick and Fidell (2007) defining values ≥ 3.29 standard deviations above or

below the mean as outliers (the probability of obtaining a true sample this extreme is 0.1%). However, similar to missing values, continuous time series data are unique in that an outlier may be identifiable by its magnitude in relation to the rest of the response curve. This viewpoint may however lead to the exclusion of certain values simply because they don't follow a relatively smooth pattern which, as measurement error is likely to exist in all samples, may be too subjective an approach. de Souza and colleagues (2015) advocate for data analyses to be conducted with and without suspected outliers, to assess whether the main analysis is robust to these extreme cases. Comprehensive reporting of this sensitivity analysis may then be the most transparent approach to the handling of outliers.

Conclusion

The TSRA has been specifically designed to speed up and standardize the calculation of summary statistics from time series data. Therefore, this tool can be used to validate calculations, and can then be cited in publications to provide transparency and to verify that the reported summary statistics are free from error. In turn, readers can have greater confidence in the reported conclusions.

306

307

308 **References**

309 Akintola, A. A., Noordam, R., Jansen, S. W., de Craen, A. J., Ballieux, B.E., Cobbaert, C. M.,
310 Mooijaart, S. P., Pijl, H., Westendorp, R. G., & van Heemst, D. (2015). Accuracy of continuous
311 glucose monitoring measurements in normo-glycemic individuals. *PLoS One, 10*(10),
312 e0139973.

313 Altman, D. G. (1985). Comparability of randomised groups. *The Statistician, 34*(1), 125-36.

314 Babraj, J. A., Vollaard, N. B. J., Keast, C., Guppy, F. M., Cottrell, G., & Timmons, J. A. (2009).
315 Extremely short duration high intensity interval training substantially improves insulin action
316 in young healthy males. *BMC Endocrine Disorders, 9*(1), 3.

317 Bell, M. L., King, M. T., & Fairclough, D. L. (2014). Bias in area under the curve for
318 longitudinal clinical trials with missing patient reported outcome data: summary measures
319 versus summary statistics. *SAGE Open, 4*(2), 1-12.

320 Berthiaume, N., & Zinker, B. A. (2002). Metabolic responses in a model of insulin resistance:
321 comparison between oral glucose and meal tolerance tests. *Metabolism: Clinical and*
322 *Experimental, 51*(5), 595-8.

323 Bickerton, A. S. T., Roberts, R., Fielding, B. A., Hodson, L., Black, E. E., Wagenmakers, A. J.
324 M., Gilbert, M., Karpe, F., & Frayn, K. N. (2007). Preferential uptake of dietary fatty acids in
325 adipose tissue and muscle in the postprandial period. *Diabetes, 56*(1), 168-76.

326 Ceriello, A., Esposito, K., Piconi, L., Ihnat, M. A., Thorpe, J. E., Testa, R., Boemi, M., &
327 Giugliano, D. (2008). Oscillating glucose is more deleterious to endothelial function and
328 oxidative stress than mean glucose in normal and type 2 diabetic patients. *Diabetes, 57*(5),
329 1349-54.

330 Chung, S. T., Ha, J., Onuzuruike, A. U., Kasturi, K., Galvan-De La Cruz, M., Bingham, B. A.,
331 Baker, R. L., Utumatwishima, J. N., Mabundo, L. S., Ricks, M., Sherman, A. S., & Sumner, A.
332 E. (2017). Time to glucose peak during an oral glucose tolerance test identifies prediabetes risk.
333 *Clinical Endocrinology, 87*(5), 484-91.
334 Cree-Green, M., Xie, D., Rahat, H., Garcia-Reyes, Y., Bergman, B. C., Scherzinger, A., Behn,
335 C. D., Chan, C. L., Kelsey, M. M., Pyle, L., & Nadeau, K. J. (2018). Oral glucose tolerance
336 test glucose peak time is most predictive of prediabetes and hepatic steatosis in obese girls.
337 *Journal of the Endocrine Society, 2*(6), 547-62.
338 de Souza, R. J., Eisen, R. B., Perera, S., Bantoto, B., Bawor, M., Dennis, B. B., Samaan, Z., &
339 Thabane, L. (2016). Best (but oft-forgotten) practices: sensitivity analyses in randomized
340 controlled trials. *American Journal of Clinical Nutrition, 103*(10), 5-17.
341 Di Nicolao, G., Liberati, D., & Sartorio, A. (2000). Stimulated secretion of pituitary hormones
342 in normal humans: a novel direct assessment from blood concentrations. *Annals of Biomedical*
343 *Engineering, 28*(9), 1136-45.
344 Gannon, M. C., Nuttall, F. Q., Westphal, S. A., Neil, B. J., & Seaquist, E. R. (1989). Effects of
345 dose of ingested glucose on plasma metabolite and hormone responses in type II diabetic
346 subjects. *Diabetes Care, 12*(8), 544-52.
347 Garlick, P. J., Wernerman, J., McNurlan, M. A., Essen, P., Lobley, G. E., Milne, E., Calder, G.
348 A., & Vinnars, E. (1989). Measurement of the rate of protein synthesis in muscle of
349 postabsorptive young men by injection of a 'flooding dose' of [1-^{13}C] leucine. *Clinical Science*
350 *(London, England: 1979), 77*(3), 329-36.
351 Gonzalez, J. T., Veasey, R. C., Rumbold, P. L. S., & Stevenson, E. J. (2013). Breakfast and
352 exercise contingently affect postprandial metabolism and energy balance in physically active
353 males. *British Journal of Nutrition, 110*(4), 721-32.

354 Gore, C. J., & Withers, R. T. (1990). The effect of exercise intensity and duration on the oxygen
355 deficit and excess post-exercise oxygen consumption. *European Journal of Applied Physiology,*
356 *60*(3), 169-74.
357 Grinspon, R. P., Ropelato, M. G., Gottlieb, S., Keselman, A., Martínez, A., Ballerini, M. G.,
358 Domené, H. M., & Rey, R. A. (2010). Basal follicle-stimulating hormone and peak
359 gonadotropin levels after gonadotropin-releasing hormone infusion show high diagnostic
360 accuracy in boys with suspicion of hypogonadotropic hypogonadism. *The Journal of Clinical*
361 *Endocrinology & Metabolism, 95*(6), 2811-18.
362 Ha, M. A., Mann, J. I., Melton, L. D., & Lewisbarned, N. J. (1992). Calculation of the glycemic
363 index. *Diabetes, Nutrition & Metabolism, 5*(2), 137-9.
364 Hofman, Z., van Drunen, J. D. E., de Later, C., & Kuipers, H. (2004). The effect of different
365 nutritional feeds on the postprandial glucose response in healthy volunteers and patients with
366 type II diabetes. *European Journal of Clinical Nutrition, 58*(11), 1553-6.
367 Huck, S. W., & Cormier, W. H. (1995). *Reading statistics and research* (2nd ed.). New York,
368 USA: HarperCollins.
369 Koppeschaar, H. P. F., Popovic, V., Leal, A., Otero, X. L., Torres, E., Paramo, C., Micic, D.,
370 Garcia-Mayor, R. V., Sartorio, A., Dieguez, C., & Casanueva, F. F. (2004). Growth hormone
371 (GH) peaks versus areas under the curve in the diagnosis of adult GH deficiency: analysis of
372 the variables provided by the GHRH + GHRP-6 test. *Pituitary, 7*(1), 15-20.
373 Landler, L., Ruxton, G. D., & Malkemper, E. P. (2018). Circular data in biology: advice for
374 effectively implementing statistical procedures. *Behavioral Ecology and Sociobiology, 72*, 128.
375 le Roux, C. W., Patterson, M., Vincent, R. P., Hunt, C., Ghatei, M. A., & Bloom, S. R. (2005).
376 Postprandial plasma ghrelin is suppressed proportional to meal calorie content in normal-
377 weight but not obese subjects. *The Journal of Clinical Endocrinology & Metabolism, 90*(2),
378 1068-71.

379 Li, P., & Stuart, E. A. (2019). Best (but oft-forgotten) practices: missing data methods in
380 randomized controlled nutrition trials. *American Journal of Clinical Nutrition, 109*(3), 504-8.
381 Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY:
382 John Wiley.
383 Matthan, N. R., Ausman, L. M., Meng, H., Tighiouart, H., Lichtenstein A. H. (2016).
384 Estimating the reliability of glycemic index values and potential sources of methodological and
385 biological variability. *American Journal of Clinical Nutrition, 104*(4), 1004-13.
386 Matthews, J. N., Altman, D. G., Campbell, M. J., & Royston, P. (1990). Analysis of serial
387 measurements in medical research. *British Medical Journal, 300*(6719), 230-5.
388 Monnier, L., Lapinski, H., & Colette, C. (2003). Contributions of fasting and postprandial
389 plasma glucose increments to the overall diurnal hyperglycemia of type 2 diabetic patients.
390 *Diabetes Care, 26*(3), 881-5.
391 Rodbard, D. (2009). New and improved methods to characterize glycemic variability using
392 continuous glucose monitoring. *Diabetes Technology & Therapeutics, 11*(9), 551-65.
393 Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA:
394 Allyn and Bacon.
395 Tentolouris, N., Kanellos, P. T., Siami, E., Athanasopoulou, E., Chaviaras, N., Kolovou, G.,
396 Sfikakis, P. P., & Katsilambros, N. (2017). Assessment of the validity and reproducibility of a
397 novel standardised test meal for the study of postprandial triacylglycerol concentrations. *Lipids,*
398 *52*(8), 675-86.
399 Vinson, J., & Bose, P. (1988). Comparative bioavailability to humans of ascorbic acids alone
400 of in Citrus extract. *American Journal of Clinical Nutrition, 48*(3), 601-4.
401 Vorster, H. H., Venter, C. S., & Silvis, N. (1990). The glycaemic index of foods: a critical
402 evaluation. *South African Journal of Clinical Nutrition, 2*(1), 13-17.

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biology, 13*(4), e1002128.

Wijsman, C. A., van Heemst, D., Hoogeveen, E. S., Slagboom, P. E., Maier, A. B., de Craen, A. J. M., van der Ouderaa, F., Pijl, H., Westendorp, R. G. J., & Mooijaart, S. P. (2013). Ambulant 24-h glucose rhythms mark calendar and biological age in apparently healthy individuals. *Aging cell, 12*(2), 207-13.

Wolever, T. M. (2004). Effect of blood sampling schedule and method of calculating the area under the curve on validity and precision of glycaemic index values. *British Journal of Nutrition, 91*(2), 295-301.

Wolever, T. M., & Jenkins, D. J. A. (1986). The use of the glycemic index in predicting the blood glucose response to mixed meals. *American Journal of Clinical Nutrition, 43*(1), 167-72.

Table 1. Summary of the various summary statistics available in the output of the TSRA.

Summary Statistic	Definition/Inference	Advantages	Limitations	Examples in Nutrition and Exercise Science
Area under the curve	A value representative of the magnitude of the total response to a stimulus across a given time period, calculated using the trapezoidal rule.	A single value that takes into account the two-dimensionality of time-series data (e.g. both the magnitude and the duration of the response are accounted for)	Inconsistent definitions throughout the literature Mathematical complexity increases probability of human/computational error	Blood glucose and insulin concentration responses to an oral glucose tolerance test Appetite hormone responses to certain meals
Peak	The maximum measured value attained in response to the stimulus.	Simple identification of the highest measured value Clearly indicative of the maximum instantaneous exposure to the stimulus	Validity dependent on measurement frequency relative to true peak, and error associated with the measurement technique	Diagnosis of diabetes during an oral glucose tolerance test Exogenous glucose oxidation rates during exercise, when comparing carbohydrate-based sports drinks
Time to Peak	The time taken to reach the maximum measured value. The onset of a given exposure.	Simple identification of the time at which the highest measured value was sampled May provide insight into the early-phase response to a stimulus	Validity dependent on measurement frequency relative to true peak, and error associated with the measurement technique Mechanistic inference may be confounded by contributing rates of appearance and disappearance	Early-phase insulin response to an oral glucose tolerance test Oxygen uptake kinetics at the onset of steady-state exercise Enhancing post-exercise glycogen resynthesis rates
Minimum	The minimum value attained in response to a stimulus.	Simple identification of the lowest measured value	Validity dependent on measurement frequency relative to true nadir, and error associated with the measurement technique	Analysis of variables that are known to decrease in response to a stimulus, such as plasma non-esterified fatty acid or glucagon-like peptide-1 responses to carbohydrate ingestion
Variability Statistics	The degree to which a measured marker varies throughout a given period of time.	Calculations can be relatively straightforward (e.g. standard deviation, coefficient of variation etc.) Provides insight into holistic homeostatic control mechanisms	Wide range of variability statistics available Susceptible to confounding by the existence of outliers	Glycemic variability with continuous glucose monitoring data Exercise intensity variability during endurance events (e.g. heart rate or perceived exertion during a cycling road race)

425

426

427 **Figure Legends**

428

429 **Figure 1.** 90-minute blood glucose concentration response to milkshake ingestion under two
430 conditions (breakfast-rest vs. breakfast-exercise). Data are presented as individual measured
431 responses across time (A), and using the incremental area under the curve (AUC) summary
432 statistic displayed as mean ± 95% confidence intervals with individual measured responses (B).
433 Real experimental data for nine participants extracted from Gonzalez et al. (2013).

434

435 **Figure 2.** Illustrations of the range of area under the curve definitions used throughout the
436 literature. See text for descriptions and examples for each.

437

438 **Figure 3.** Hypothetical illustration of an individual measured response to a stimulus across
439 time. The alternative measured responses on each panel demonstrate when area under the curve,
440 peak and time-to-peak summary statistics all provide different inferences, requiring cautious
441 and contextual interpretation

442

443 **Figure 4.** Simple representation of linear interpolation to impute missing data (A), and a
444 hypothetical time series response demonstrating a key limitation of linear interpolation (B).