Accurate Mass Measurement: Terminology and Treatment of Data

A. Gareth Brenton and A. Ruth Godfrey

Institute of Mass Spectrometry, Grove Building, School of Medicine, Swansea University, Swansea, Wales, United Kingdom

High-resolution mass spectrometry has become ever more accessible with improvements in instrumentation, such as modern FT-ICR and Orbitrap mass spectrometers. This has resulted in an increase in the number of articles submitted for publication quoting accurate mass data. There is a plethora of terms related to accurate mass analysis that are in current usage, many employed incorrectly or inconsistently. This article is based on a set of notes prepared by the authors for research students and staff in our laboratories as a guide to the correct terminology and basic statistical procedures to apply in relation to mass measurement, particularly for accurate mass measurement. It elaborates on the editorial by Gross in 1994 regarding the use of accurate masses for structure confirmation [1]. We have presented and defined the main terms in use with reference to the International Union of Pure and Applied Chemistry (IUPAC) recommendations for nomenclature and symbolism for mass spectrometry. The correct use of statistics and treatment of data is illustrated as a guide to new and existing mass spectrometry users with a series of examples as well as statistical methods to compare different experimental methods and datasets. (J Am Soc Mass Spectrom 2010, 21, 1821–1835) © 2010 Published by Elsevier Inc. on behalf of American Society for Mass Spectrometry

The measurement of ion mass is an important tool for scientists over a wide range of disciplines, and mass is reported either as a

Nominal mass—the mass of an ion or molecule calculated using the mass of the most abundant isotope of each element rounded to the nearest integer value and equivalent to the sum of the mass numbers of all constituent atoms.

or

Accurate mass—the experimentally determined mass of an ion measured to an appropriate degree of accuracy and precision used to determine, or limit the possibilities for, the elemental formula of the ion [1].

A mass spectrum can be annotated with its nominal masses or accurate masses, to an appropriate number of significant figures. The IUPAC unit of mass is the unified mass unit (u) [2] and is also referred to as the Dalton (Da) although this is not an SI unit. The term atomic mass unit (amu) is a redundant unit although it is still in wide use. In this article, discussions are limited to singly charged ions and the terms Da and mDa will be used (rather than the SI units u and mu (i.e., milli u), respectively) to refer to ion mass as these are widely recognized by mass spectrometrists. However, a mass spectrometer measures mass-to-charge, and m/z should normally be used when referring to the mass scale. There are a number of methods for accurate mass measurement, which involve different approaches and instrumentation, but all involve calibration of the mass scale using ions of known exact mass [3, 4].

It is necessary to emphasize the difference between the terms accurate mass and exact mass. Accurate mass is the experimental quantity that is measured and exact mass the calculated quantity.

Exact mass—is the calculated mass of an ion whose elemental formula, isotopic composition and charge state are known, i.e., it is the theoretical mass. The IUPAC definition constricts the definition to using one isotope of each atom involved, usually the lightest isotope, but generalizes the definition to cover an ion or neutral molecule [2]. The charge state is relevant as the mass of the electron (0.00055 Da), or multiple charges, may not be negligible in the context of mass measurement.

There is a need to statistically treat accurate mass measurement data and apply terminology that describes these procedures in a consistent manner. The aim of this article is to clarify and define terms in common usage and to advise which are preferred.

Whilst the terms "accurate mass" and "exact mass" are commonly used for the measured and calculated masses, respectively, Sparkman [5] has suggested using "measured accurate mass" and "calculated exact mass", that leaves no doubt which is the experimental and the

Address reprint requests to Dr. A. G. Brenton, Institute of Mass Spectrometry, Grove Building, School of Medicine, Swansea University, Singleton Park, Swansea, Wales SA2 8PP, UK. E-mail: g.brenton@swansea.ac.uk

calculated mass. In this article the briefer terms accurate mass and exact mass will be used as they are in common usage. It is not our intention to make strong statements or recommendations on which terminology constitutes best practice, this should be the work of IUPAC [2].

The difference between the measured value (accurate mass) and the true value (exact mass) is the "accuracy" of the "accurate mass measurement" (an unfortunate double use of the word) and it is suggested that the term "mass measurement accuracy" should be used to denote this difference.

We have introduced the concept of the accuracy of measurements, which reflects the presence of systematic errors, and it is important to emphasize the difference between accuracy and precision.

Accuracy—the proximity of the experimental measurement to the true value (exact mass). When a measurement is close to the true value we say it is accurate and when it is not we say it is inaccurate.

Normally, *mass measurement error* would be used to describe the accuracy of a single reading.

Precision—the repeatability of the measurement reflecting random errors. Random errors cause measurements to fall on either side of the average experimental measurement and affect the precision of the set of measurements. When a set of mass measurements of one ion species lie close together we say the measurements are precise, and when not we say the measurements are imprecise.

Two other items of terminology which must be clarified are:

Repeatability—this is the short-term precision of multiple replicate experimental measurements made under similar conditions, i.e., the same instrument, operator and over a limited time, normally the same day.

Reproducibility—refers to differences among experimental measurements made under different circumstances i.e., a measurement of the same quantity made by different operators, even different instruments and often with a significant time difference between groups of measurements.

There are appropriate statistical methods to test the repeatability and reproducibility of experimental measurements, e.g., the separation and estimation of precision (measured as "standard deviation" or its square, "variance") of groups of measurements. The most general approach is termed analysis of variance (ANOVA) [6].

The mass spectrometry literature is large, with numerous text books published on the subject. To our knowledge, there is currently no single source of material that adequately describes terminology and statistics of accurate mass measurement. In 1984, Sack et al. described a generalized procedure for evaluating the accuracy and precision of mass measurement with a study of magnetic sector data [7]. IUPAC's standard definitions of terms relating to mass spectrometry [2, 8] has limited definitions in relation to accurate mass measurements, e.g., accurate mass and exact mass, whereas numerous terms are employed within the literature. In 1991, Price published a compilation of terms used in mass spectrometry, which was commissioned by the American Society for Mass Spectrometry [9]. Compendium of mass spectrometry terms have been published in a "*Mass Spec Desk Reference*" [5] and a "*Dictionary of Mass Spectrometry*" [10], however the contents are the views of the authors and are not necessarily consistent with those MS community or IUPAC approved nomenclature. In addition, the textbooks in references [11–14] are good sources for the student. However, none of these sources are comprehensive with regard to accurate mass measurement terminology. Best practice guides relating to the "*Best Practice Guide: Methodology for Accurate Mass Measurement of Small Molecules*" [3] and "*Generating Mass Spectra*" [4] have been prepared and published by the Laboratory of the Government Chemist (LGC Ltd., Middlesex, UK) under their Valid Analytical Measurement (VAM) program.

Introduction

Repeated experimental measurements of the accurate mass of a molecular ion species $M^{+\cdot}$ or protonated molecule $[M + H]^+$, or any other ion, provides an estimation of errors associated with the measurement. Types of errors encountered are random, systematic, or gross. Gross errors are so serious that if they occur, the data should be dismissed and the experiment repeated. Random errors cause mass measurements to fall on both sides of the mean value of the experimentally measured mass, and affect precision of the measurements. In contrast, systematic errors cause mass measurements to tend to values higher, or lower, than the calculated exact mass and affect accuracy, that is, they give evidence of bias in mass measurements.

In the absence of systematic errors, if an increasing number of measurements is made, the mean value of the measured mass should get closer to the calculated exact mass. That is, the mass measurement accuracy will improve and its magnitude will reduce towards zero as the result of (partial) cancellation of positive and negative deviations. In practice, this generally does not happen as there are always systematic errors giving rise to drift of the mass scale arising from small instrumental or electronic instabilities or other effects.

Assuming the applicability of "normal" (or Gaussian) statistics (see later), when n measurements of a mass, m_i, are taken, the error (difference between the measured mean value and the "true" exact mass) will vary inversely with the square root of the number of readings, i.e., as $n^{-1/2}$. Thus, the mass measurement

accuracy will never be guaranteed to reach zero even in the absence of systematic errors (measurement bias). So a pragmatic choice of the number of measurements has to be made, i.e., "*what value of n should I choose*?" It will not always be the same and will depend on the strength of the signal (strictly the signal:noise ratio), ionization technique employed, and other factors. Often the operator has little or no choice, e.g., an on-line chromatographic separation may allow only a limited number of attempts at mass measurement. Repeated measurements also assume there is adequate sample and that the sample is of sufficient purity to warrant repeated measurements. Measurements taken over long periods of time will tend to introduce systematic errors and, thus, a smaller number of measurements may be more appropriate in such circumstances. In other words, increase of precision by taking a larger number of measurements (n) can not improve the accuracy if significant systematic errors are present.

Estimation of Mass Measurement Statistics

"*Should I analyze the list of masses measured or the mass errors, and if so will it affect the results if I use mass errors in different units such as Da, mDa, ppm, or m/z?*"

Let m_i be the measurement of a single accurate mass in Da, n the number of measurements, and m_a the calculated mass in Da. [The symbol m_a has been used here to represent exact mass as the symbol m_e represents the electron mass and the symbol m_{ex} could be interpreted as experimental mass and not exact mass.]

Then the *mass measurement error* (or accuracy) of a single reading will be

$$\Delta m_i = (m_i - m_a) \text{ in Da}$$

$$= (m_i - m_a) \times 10^3 \text{ in mDa}$$

$$= \frac{(m_i - m_a)}{m_a} \times 10^6 \text{ in ppm (parts per million)}$$

The term "error" is defined as the difference between an individual measurement and the true value [15], where Δm_i can be positive or negative. It is important to remember that the term "error" should only be used in connection with the result of a *single* measurement and not used to represent an *averaged* measurement (see later discussion). Errors in relation to the exact mass measurement are only apparent when the compound's composition is known.

The term "mass deviation" has been proposed to the proteomics community for reporting a single mass measurement error [16]. These authors define mass deviation as "the measured mass minus the calculated mass," and recommend this instead of using "mass

accuracy". Whilst it is correct to point out that the term accuracy should not be used for a single measurement, there is no necessity to redefine this as deviation. In fact, IUPAC defines deviation as "the difference between an observed value and the arithmetic mean of the set to which it belongs." This is not the same definition as that for the error of a measurement, which IUPAC defines as "the result of a measurement minus the true value." Thus, the term mass deviation is not appropriate to describe a mass measurement error. They further define the term maximum mass deviation (MMD), "this is the cutoff value used in database search. Only peptide sequences with a calculated mass within this tolerance are reported as hits."

For readers unfamiliar with statistical terminology, the discussions here represent a particular case of univariate statistics (one variable) and does not deal with procedures involving two or more variables, e.g., in the cases of regression and correlation.

An example (using data tabulated in Table 1).

Consider a molecular ion $M^{+\cdot}$ of known composition and thus known mass. For simplicity (and illustration of statistical procedures here) its exact mass is assumed to be 400.0000 Da, and accounts for the mass of the missing electron. Nine mass measurements (n = 9) have been taken (Col. 2, tabulated in Da) and mass errors calculated in Da (Col. 3), mDa (Col. 4), and parts per million (ppm) (Col. 5), and descriptive statistics calculated.

Note: ppm error is a quantity frequently used to report mass errors and varies along the mass scale.

Significant Figure Notation and the Number of Decimal Places to Use When Quoting Accurate Mass Measurement Data

A simple way to indicate the accuracy of measurement is the use of the significant figure convention. For example, if experience and observation of mass measurements showed measurements to give ~0.001 Da accuracy, then the mass would be reported to one more significant figure. In our example, the first mass measured (Table 1, data point i = 1) may be stored in computer memory as 400.0013415 but would be reported in significant figure notation as 400.0013 (four decimal places), for 0.001 Da mass accuracy. Intermediate values used in formulae and calculations should be generally written to at least one extra decimal place, in this case five decimal places, to ensure there are no rounding errors before the presentation of the final result to four decimal places.

If an even better mass measurement technique was used capable of routine mass measurement to a further decimal place of accuracy, i.e., 0.0001 Da, eight significant figures (five decimal places) is required and the mass would be quoted as 400.00134 Da.

Modern mass spectrometers generally report accurate mass measurements to four decimal places (seven significant figures for masses between 100 and 999 Da) and sometimes more. For the lower end of this mass

Table 1. Example of the calculation using MS Excel *Descriptive Statistics* for a set of nine mass measurements of an ion on mass 400 Da; m_i is the measured mass (to four decimal places), Δm_i the mass errors for each measurement and are given in Da, mDa, and ppm, and written in significant figure notation. To reduce computational errors any derived value, i.e., in this example ppm data, may be written with an extra significant figure and used in computations (given in brackets in the final column). The results of calculations with derived units (ppm) should then be round down one place. In this example the sample variance is the only parameter to have a small computational error, of 0.2, giving a value of 56.3 instead of 56.1

i	m_i	Δm_i (Da)	Δm_i (mDa)	$\Delta m/m_i \times 10^6$ (ppm)
1	400.0013	0.0013	1.3	3.3 (3.25)
2	399.9987	−0.0013	−1.3	−3.3 (−3.25)
3	400.0072	0.0072	7.2	18 (18.00)
4	400.0024	0.0024	2.4	6 (6.00)
5	400.0015	0.0015	1.5	3.8 (3.75)
6	399.9969	−0.0031	−3.1	−7.8 (−7.75)
7	400.0027	0.0027	2.7	6.8 (6.75)
8	400.0013	0.0013	1.3	3.3 (3.25)
9	399.9985	−0.0015	−1.5	−3.8 (−3.75)
Mean	400.0012	0.0012	1.2	2.9 (2.92)
Standard error of the mean†	0.0010	0.0011	1.0	2.5 (2.50)
Median†	400.0013	0.0013	1.3	3.3 (3.25)
Mode†	400.0013	0.0013	1.3	3.3 (3.25)
Standard deviation*	0.0030	0.003	3.0	7.5 (7.49)
Sample variance*	9.0E-06	9.0E-06	9.0	56.3 (56.10)
Count	9	9	9	9
Confidence limits of the mean (95%)	0.0023	0.0023	2.3	5.8

*Excel calculates the sample standard deviation (s) and the sample variance (s²) using, $s = \sqrt{\dfrac{\sum_i (m_i - \overline{m_i})^2}{n-1}}$, which is the unbiased sample standard deviation.

†Standard error of the mean is defined as $s/\sqrt{n}$, which may be compared with the "true" standard deviation of the mean $\sigma/\sqrt{n}$ from the assumed underlying Gaussian distribution. The median is the middle value of the list when it is sorted (for a list of n values it will be the middle value when n is odd and is taken as the average of the $(n/2)^{th}$ and $(n/2 + 1)^{th}$ values when n is even) and is used for non-parametric statistical tests. Mode is the value that occurs most frequently in the list of measurements (twice in this example, for $i = 1$ and 8).

range (masses below ~400 Da) it seems sensible to the authors to report data to eight significant figures to eliminate any chance of rounding errors arising. Internally, data will be stored in a computer to many more significant figures; data are usually stored as a "double type" floating point number to 15 digits, sufficient to eliminate computationally related errors. In the literature and reports generated by mass spectrometer software, lists of mass data can be seen given to insufficient significant figures and rounding errors will arise, especially at lower masses where the effect will be more pronounced. Mass measurement error (difference between the accurate mass and exact mass) is normally reported on a suitable scale, i.e., mDa, or mu (milli u), or ppm and should be given to at least one decimal place (typically two or three significant figures).

Accuracy and Precision of the Measurement Sample Distribution

The mean value of a set of replicate measurements (statisticians call this experimental set the sample distribution) of the mass of a single ion species is the mean experimental mass or average experimental mass. Sometimes mass spectrometrists wrongly refer to this as the average mass. The average mass of a molecule (or an ion) is based on its molecular (or empirical) formula and is calculated using the relative atomic mass ("atomic weight") of each element weighted for its natural isotopic composition, i.e., C = 12.0107, H = 1.00794, O = 15.9994 etc. [17].

The average experimental mass ($\overline{m_i}$) of n measurements of a single ion species is defined

$$\overline{m_i} = \frac{\sum_i m_i}{n}$$

When reporting accurate mass measurements, it is important that the researcher quotes whether the results are an average of replicate measurements or are single measurements, and it should be made clear in any write-up. If an average value is reported, the number of measurements and accompanying standard deviation should be quoted. If it is a single measurement, then it should be stated that the data is a result of single measurement(s) and the error(s) should be referred to as the "error(s) of the mass measurement."

Accuracy refers to how close mass measurements are to the expected result m_a, the exact mass in this case. The average experimental mass is used to assess accuracy of a set of measurements made for the same ion.

Our example

$\overline{m_i} = 400.0012$ Da for $m_a = 400$, the mass accuracy is 0.0012 Da or 1.2 mDa or 2.9 ppm (Figure 1).

Precision refers to the distribution or spread of individual results in the sample distribution, and when plotted as a histogram should form or be closely modeled by the so-called normal distribution (also called Gaussian distribution). Although it is recognized by mathematicians that experimental measurements can often lead to a data distribution that is slightly different from the classical normal distribution, usually exhibiting longer tails. The normal distribution is widely used in probability theory and statistics as it is a simple model and, under certain conditions, can be generally applied as a good approximation to many different types of data. The standard deviation(s) is a useful measure of precision for analytical measurements that follow a normal distribution.

The standard deviation of the sample (s) of n repeated measurements obtained for a particular mass is:

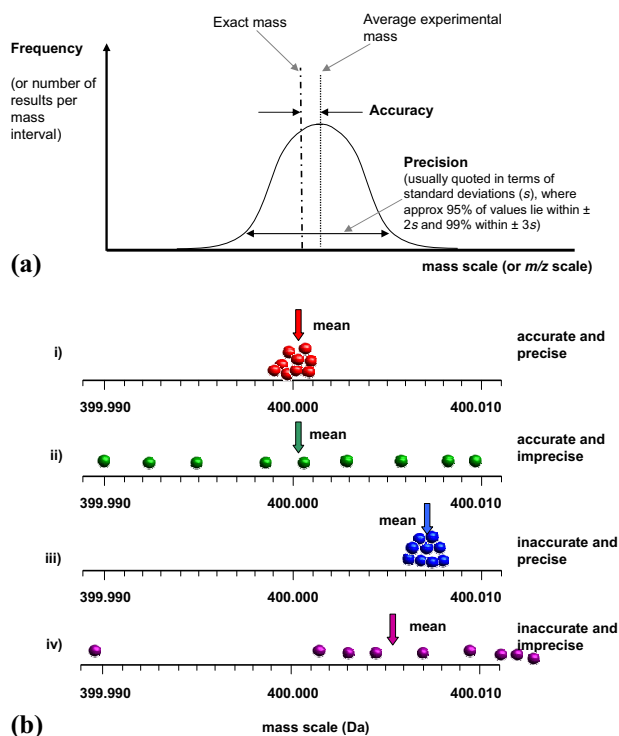$$ s = \sqrt{\frac{\sum_{i} \left(m_i - \overline{m_i}\right)^2}{n - 1}} $$

Figure 1. Illustration of the meaning of the terms accuracy, precision, average experimental mass and calculated exact mass of a dataset. (**a**) Histogram (plotted for convenience as a continuous probability curve) showing accuracy i.e., the difference between the calculated exact mass (reference quantity) and the measured average mass (experimental quantity). Precision is a measure of the spread of mass measurements of the dataset, and relate to the repeatability of the measurements undertaken. (**b**) Plot of nine repeat accurate mass measurements at mass 400 Da, in the range 399.990 to 400.010 Da, illustrating four types of statistical outcomes of data that is, (i) accurate and precise, (ii) accurate but imprecise, (iii) inaccurate but precise, and (iv) inaccurate and imprecise.
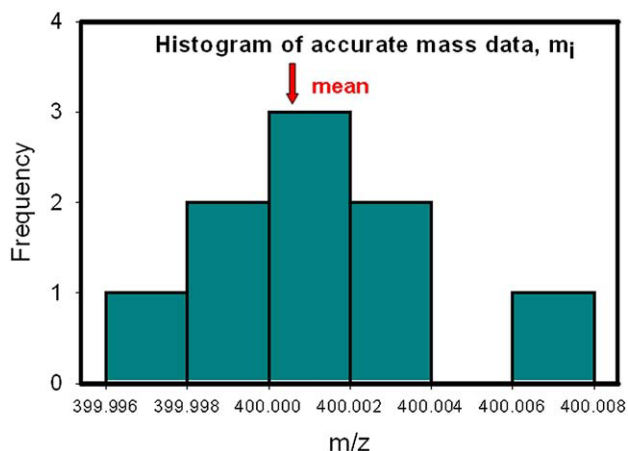

Figure 2. Histogram of data given in Table 1. The frequency plot is calculated for a bin width of 0.002 Da.

When estimating the precision of the sample distribution it is important to use the average experimental mass $(\overline{m_i})$, not the theoretical value m_a. The square of the standard deviation is called the "variance" of the distribution.

Our example

s = 0.0030 Da.

Sample distribution—this is the distribution, e.g., a frequency (histogram) plot, of the measured dataset (see Figure 2).

Population distribution—this is the theoretical set of all possible measurements (the sample distribution is assumed to be a sub-set of this) and is usually modeled mathematically by the normal (Gaussian) distribution

$$ y = \frac{\exp(-(x - \mu)^2/2\sigma^2)}{\sigma\sqrt{2\pi}} $$

where x is the measurement value (mass), y the frequency or intensity of that measurement, μ is the population mean, and σ the standard deviation of the population.

The application of standard statistical tests to mass spectrometry often implicitly assume the data is modeled by the normal distribution. However, this is not necessarily true and, depending on the data and application, you may need to verify the underlying data distribution yourself, for example, by the application of a distribution test such as the Kolmogorov and Smirnov test [18, 19]. In this case, other statistical methods, including nonparametric tests, may need to be applied (for further reading see chapter 7 in [18]).

Mass Errors

Some common terms used in conjunction with mass errors are mass measurement accuracy (MMA), average absolute mass measurement accuracy (AAMMA), root mean square (RMS) error, and precision.

When results are presented as mass errors Δm_i including the sign, i.e., the difference between the measured experimental values and the expected or calculated exact mass, the following definitions are commonly used:

Mass measurement accuracy (MMA) is a commonly used term to describe the average of the mass measurement errors and is defined as:

$$\overline{m_i} = \frac{\sum_i \Delta m_i}{n}$$

By including the sign in Δm_i the positive and negative errors partially cancel. Thus, MMA provides an indication of the accuracy, i.e., the closeness of the mean value to the "true" value, but gives no information about the width of the sample distribution (that defines the precision).

Our example

MMA is 0.0012 Da or 1.2 mDa or 2.9 ppm.

Note: this term is also referred to as "mean mass measurement accuracy," "average mass measurement accuracy," "mass accuracy," "average error" [20], and "average mass error" [21]. The first two terms would appear mathematically appropriate to use, and *mass accuracy* is a common abbreviated term. All these terms have been used although "*average mass error*" is normally used for the *average of the absolute values of mass errors* (the term AAMMA, given below) and use of this term could be confusing.

How to define the magnitude of mass errors: there are several ways in which the magnitudes of mass errors are quoted. If the experimental data can reasonably be represented by a normal distribution then the standard deviation can be quoted and its meaning is well understood. However, the term "*average mass error*" [21] is commonly used for mass measurements. As the term "average mass error" is confused with mass measurement accuracy, it is best to state explicitly if absolute values of errors are used to create an average. The literature cites the following terms for the same quantity: "average mass error (absolute values)" [21], "average absolute error" [22], "average absolute mass error", "absolute exact-mass errors" [23] and "average absolute mass measurement accuracy" [24]. The term average absolute mass measurement accuracy (AAMMA) is widely used, see below. This term clarifies that "absolute" error values are used, and is defined by the following expression:

Average absolute mass measurement accuracy (AAMMA).

AAMMA (also known average absolute MMA)

$$= \frac{\sum_i |\Delta m_i|}{n}$$

The "average absolute MMA" or "average absolute mass error" is widely quoted to describe this term.

Our example

AAMMA = 0.0025 Da or 2.5 mDa or 6.2 ppm.

Note: there is considerable inconsistency and confusion in the literature on how to describe the magnitude of mass accuracy measurements. In addition "error" is often used in place of "accuracy." Any terminology that does not note that absolute values of mass measurement accuracy (or mass errors) are involved will be confusing.

Root mean square (RMS) error

$$RMS\ error = \sqrt{\frac{\sum_i (\Delta m_i)^2}{n}}$$

The root mean square error is also widely quoted. This is a term commonly used in engineering and physics to denote the magnitude of a varying quantity, especially when the quantity intrinsically varies between positive and negative values, e.g., random noise superimposed on an electrical signal.

Our example

RMS $_{error}$ = 0.0031 Da or 3.1 mDa or 7.7 ppm.

Precision—standard deviation of the mass error sample (s)

$$s = \sqrt{\frac{\sum_i (\Delta m_i - \overline{\Delta m_i})^2}{n-1}}$$

in this case we are dealing with mass errors rather than actual masses (for the analogous form see the earlier equation for 5), however both methods for calculating the standard deviation lead to the same numerical value.

Our example

s = 0.0030 Da or 3.0 mDa or 7.5 ppm.

Note: This term may be called "s" or "SD" and is the experimental estimate of the underlying normal population value, σ.

Conclusions on describing the magnitude of mass measurement accuracy.

There are three different parameters that are used to assess the uncertainty of a mass measurement:

(i)   average absolute mass measurement accuracy (AAMMA)
(ii) root mean square (RMS) error and
(iii) standard deviation (SD)

where SD $\approx$ RMS error > AAMMA, which arises by comparing the algebraic forms of these quantities. Standard deviation should be used to quote the precision of the dataset because it plays a key role in

statistical tests described later, but note that SD ≈ RMS error.

Note: If the dataset is not described by a normal distribution, the AAMMA value can still be used since it is an absolute measure and can be reported with or without also reporting the mid-point of the dataset (the median value) For more details you will need to read about nonparametric statistical treatment of data (e.g., this is treated comprehensively in ref [18]).

Our example

s = 3.0 mDa ≈ RMS = 3.1 mDa > AAMMA = 2.5 mDa.

Confidence Limits of the Average Mass (Mean)

A statistical problem is to relate the experimental distribution of measurements, typically a few data points, to the underlying population distribution (Gaussian, characterized with appropriate values of the mean, σ, and precision, μ) represented by a large number of points. The Student-t model (a distribution function that is lower and wider than the Gaussian distribution [6, 18]) is a way of relating Gaussian theory to the experimental distribution. The fundamental equation that relates the best possible experimental value of the average experimental mass, $\overline{m}_i$, to the population mean, μ, inevitably includes an estimate of uncertainty in that relationship. This uncertainty, represented by confidence limits is described by the following equation, which involves the experimental estimate of the precision s, the number of data points n, and a special statistic called the Student-t statistic, t.

For a set of measurements we can define a range where the true value is likely to lie within and is described by the "confidence limits of the mean." Analytically, we usually give the confidence interval for a 95% confidence level for the mean, i.e., the range within which we are 95% confident that the true value lies. Note this also means there is a 5% chance that the true value lies outside of the confidence limits (the extremities of the confidence interval). One reason for the common use of a 95% confidence level is that for a Gaussian distribution, this closely corresponds to confidence limits given as the mean ± 2 s). In other words, if n measurements of the same quantity (mass) are made, given the measurement precision estimated as s, one in 20 of the mean values measured is expected to fall outside of the confidence interval. This interval is written as

$$\overline{m}_i \pm t\left(\frac{s}{\sqrt{n}}\right)$$

where $\overline{m}_i$ is the mean of the sample distribution (experimental measured dataset),

s is the standard deviation of the sample distribution and

t is Student t-statistic, defined later on.

The mean of the theoretical population distribution (μ) will lie within this confidence interval, to a given probability. The number of degrees of freedom (n-1) defines the value to the Student-t statistic to use from statistical tables. Tabulated t-values decrease with increasing number of measurements (and degrees of freedom). Typical confidence level (CL) values in use are 90%, 95%, and 99%, but the most commonly used for analytical purposes is the 95% level. The so-called "P value" is simply (100-CL)/100. The higher the confidence level required, the larger t is and the wider the confidence limits, $\overline{m}_i \pm t(s/\sqrt{n})$, will be. In terms of the mass measurement accuracy, Δm_i, its confidence limits will be determined similarly and is $\pm t(s/\sqrt{n})$.

To use the tabulated values of t to apply the statistical test you will have to specify the number of degrees of freedom (dof). In the case of calculating average experimental mass values from measurements of a single quantity (e.g., ion mass) the dof is n-1. Degrees of freedom represent the number of independent pieces of information, i.e., in this case $\Delta m_i = (m_i - m_a)$ individual mass errors, before the average mass is fully defined (i.e., is able to be determined) and this is one less than the total number of measurements. This corresponds to the fact that if you know, e.g., the m_a and n-1 values of Δm_i, you can immediately calculate the value of the nth. In other statistical tests you will need to consider and check what the appropriate value of dof is, as it can be n-2 or something else [6].

Our example

The confidence limits of the average experimental mass at a 95% level of confidence is calculated as follows. Using

Table 2. t-Values at different degrees of freedom*

Confidence level	90%	95%	99%
P values	0.10	0.05	0.01
No. of degrees of freedom			
1	6.31	12.71	63.66
2	2.92	4.30	9.92
3	2.35	3.18	5.84
4	2.13	2.78	4.60
5	2.02	2.57	4.03
6	1.94	2.45	3.71
7	1.89	2.36	3.50
8	1.86	2.31	3.36
9	1.83	2.26	3.25
10	1.81	2.23	3.17
20	1.72	2.09	2.85
30	1.70	2.04	2.75
40	1.68	2.02	2.70
50	1.68	2.01	2.68
100	1.66	1.98	2.63
∞	1.64	1.96	2.58

*The critical values of |t| are appropriate to use for a two-tailed test (i.e., a test where there is no prior idea that the result, where the average mass will be higher or lower than the exact mass, in other words the average mass can fall either side of the expected result and there is no expected bias). For one-tailed tests take a t-value from the column that has a P value twice the desired P value.

$$\overline{m}_i \pm t\left(\frac{s}{\sqrt{n}}\right)$$

and gives $400.0012 \pm 2.31 \times (0.0030/\sqrt{9}) = 400.0012 \pm 0.0023$ Da, where $t = 2.31$ for $n = 9$ (8 degrees of freedom) at a 95% confidence level. Selected t-values are given in Table 2. This is an example of a two-tailed test where the difference between the experimental and the exact mass could be either positive or negative in value, that is we do not have any notion which direction the result will take.

Uncertainties in Elemental Composition Deduced from Accurate Mass Measurements

The purpose of the statistical methods outlined so far is to help deduce elemental compositions, (molecular formulae) that are consistent with given confidence levels, usually 95% and 99% confidence levels (corresponding to confidence limits defined by approximately the mean $\pm$ 2 or $\pm$ 3 times the standard deviation, respectively). It must be noted that even at these confidence levels, there

Table 3. Elemental composition listings for two compounds, A and B, with measured masses of 309.1128 and 574.2335 Da, respectively. The ions were established to be singly protonated molecular species and mass measured on a magnetic sector instrument using electrospray ionization in positive ion mode. From analysis of historic mass measurements, with the instrument in this ionization mode, it has a precision measured to be 0.81ppm (one standard deviation) [25]. Elemental composition listing is given to ±3 standard deviations (2.43 ppm) of the measured mass and encompasses candidates $^{12}C_{0-40}$ $^1H_{5-60}$ $^{14}N_{0-8}$ $^{16}O_{0-8}$ $^{31}P_{0-2}$ $^{32}S_{0-2}$. The known formulae are indicated by *

No	Mass (m/z)	Δm_i (mDa)	Δm_i (ppm)	RDB	Elemental composition	Std deviations
Compound A						
1	309.1128	0.05	0.16	7.5	$^{12}C_{12}H_{17}\,O_2\,N_6\,^{32}S$	
2	309.1128	0.11	0.36	2.5	$^{12}C_{11}H_{23}\,O_4\,N_2\,P_2$	
3	309.1126	0.23	0.74	−1.0	$^{12}C_5\,H_{19}O_{10}\,N_5$	1
4	309.1131	−0.26	−0.84	3.0	$^{12}C_8\,H_{20}\,O_2\,N_7\quad P^{32}S$	
5	309.1124	0.42	1.36	7.0	$^{12}C_{15}H_{20}\,O_4\,N\quad P$	2
6	309.1135	−0.60	−1.94	16.5	$^{12}C_{20}H_{13}\,N_4$	
7*	309.1121	0.73	2.36	11.5	$^{12}C_{19}H_{17}\,O_4$	3
Compound B						
1	574.2335	−0.01	−0.02	0.0	$^{12}C_{13}H_{38}O_{13}N_{10}\,^{32}S$	
2	574.2335	0.01	0.02	11.0	$^{12}C_{31}H_{43}\,O_4\quad P^{32}S_2$	
3	574.2336	−0.12	−0.21	3.5	$^{12}C_{19}H_{42}\,O_7\,N_7\quad P_2{}^{32}S$	
4	574.2336	−0.14	−0.24	27.0	$^{12}C_{34}H_{26}N_{10}$	
5	574.2336	−0.15	−0.26	21.5	$^{12}C_{35}H_{32}\,O_5\,N_3$	
6	574.2337	−0.19	−0.33	3.0	$^{12}C_{21}H_{42}O_{10}\,N_4{}^{32}S_2$	
7	574.2333	0.19	0.33	8.0	$^{12}C_{23}H_{39}\,O_7\,N_6\quad P^{32}S$	
8	574.2338	−0.31	−0.54	6.5	$^{12}C_{27}H_{46}\,O_4\,N\quad P^{32}S_2$	
9	574.2331	0.37	0.64	−0.5	$^{12}C_{16}H_{41}O_{15}\,N_5\quad P$	
10	574.2331	0.39	0.68	16.0	$^{12}C_{33}H_{40}\;\,O\,N_2\quad P_2{}^{32}S$	
11	574.2340	−0.46	−0.80	17.0	$^{12}C_{31}H_{35}\,O_5\,N_4\quad P$	1
12*	574.2330	0.50	0.87	12.5	$^{12}C_{27}H_{36}\,O_7\,N_5\,^{32}S$	
13	574.2329	0.56	0.98	7.5	$^{12}C_{26}H_{42}\,O_9\,N\quad P_2$	
14	574.2329	0.57	0.99	13.0	$^{12}C_{25}H_{36}\,O_4\,N_8\quad P_2$	
15	574.2342	−0.66	−1.15	9.0	$^{12}C_{21}H_{34}O_{11}\,N_8$	
16	574.2328	0.68	1.18	4.0	$^{12}C_{20}H_{38}O_{15}\,N_4$	
17	574.2328	0.70	1.22	20.5	$^{12}C_{37}H_{37}\;\,O\;\,N\quad P^{32}S$	
18	574.2343	−0.77	−1.34	12.5	$^{12}C_{27}H_{38}\,O_5\,N_5\quad P_2$	
19	574.2343	−0.83	−1.45	17.5	$^{12}C_{28}H_{32}\,O_3\,N_9\,^{32}S$	
20	574.2343	−0.84	−1.46	12.0	$^{12}C_{29}H_{38}\,O_8\,N_2\,^{32}S$	
21	574.2327	0.84	1.46	−1.0	$^{12}C_{15}H_{43}\,O_9\,N_8\quad P^{32}S_2$	
22	574.2326	0.88	1.53	12.0	$^{12}C_{30}H_{39}\,O_9\quad P$	
23	574.2326	0.88	1.53	17.5	$^{12}C_{29}H_{33}\,O_4\,N_7\quad P$	2
24	574.2345	−0.97	−1.69	4.5	$^{12}C_{17}H_{37}O_{11}\,N_9\quad P$	
25	574.2345	−1.02	−1.78	20.5	$^{12}C_{36}H_{36}\,N_3\quad P^{32}S_2$	
26	574.2325	1.04	1.81	7.0	$^{12}C_{25}H_{44}\,O_3\,N_4\quad P_2{}^{32}S_2$	
27	574.2346	−1.15	−2.00	13.0	$^{12}C_{24}H_{35}\,O_3N_{10}\quad P^{32}S$	
28	574.2347	−1.15	−2.00	7.5	$^{12}C_{25}H_{41}\,O_8\,N_3\quad P^{32}S$	
29	574.2323	1.15	2.00	3.5	$^{12}C_{19}H_{40}\,O_9\,N_7{}^{32}S_2$	
30	574.2323	1.19	2.07	22.0	$^{12}C_{33}H_{30}\,O_4\,N_6$	
31	574.2323	1.22	2.12	4.0	$^{12}C_{17}H_{40}\,O_6N_{10}\quad P_2{}^{32}S$	
32	574.2348	−1.28	−2.23	0.0	$^{12}C_{13}H_{40}O_{11}N_{10}\quad P_2$	
33	574.2348	−1.33	−2.32	16.0	$^{12}C_{32}H_{39}\,N_4\quad P^{32}S_2$	
34	574.2348	−1.35	−2.35	−0.5	$^{12}C_{15}H_{40}O_{14}\,N_7\;\,^{32}S$	
35	574.2322	1.35	2.35	11.5	$^{12}C_{29}H_{41}\,O_3\,N_3\quad P^{32}S_2$	3

Table 4. Terminology used in connection with accurate mass measurement

Term	Definition and comments	Other terms used*
Accurate mass	The experimentally determined mass of an ion measured to an appropriate degree of accuracy. It is often used to determine an elemental formula.	Measured accurate mass
Accuracy	The proximity of the experimental measurement to the true value (exact mass). When a measurement is close to the true value we say it is accurate and when it is not we say it is inaccurate.	
Analysis of variance (ANOVA)	ANOVA can be used to test differences between two or more average masses, e.g., due to systematic error (a fixed change) between two or more different sets of measurements. ANOVA can also be used to estimate where the sources of variation arise i.e. so called 'between-samples' e.g., for experiments taken in different laboratories or on different days and 'within-samples' e.g., for experiments taken sequentially in one laboratory and on one apparatus.	
Atomic mass unit (amu)	Although still widely used to refer to the mass of an ion, <u>this unit is incorrect</u> as it refers to the previous definition of the atomic mass scale based on one sixteenth of the mass of ^{16}O atom (^{16}O was defined as a mass of 16 amu). This unit was replaced in 1959/60 when both IUPAC (International Union of Pure and Applied Chemistry) and IUPAP (International Union of Pure and Applied Physics) agreed to define "the mole is the amount of substance of a system, which contains as many elementary entities as there are atoms in 0.012 kg of carbon 12; its symbol is "mol." The unit of mass was changed to unified atomic mass (u) [15].	Unified atomic mass unit (u).
Atomic mass	Mass of an atom, units are the unified atomic mass unit (u). 1 u $\approx$ 1.66054002 $\times$ 10^{-27} kg.	
Atomic weight [17,28]	The ratio of the average mass of the atom, based on stable isotope distribution, to the unified atomic mass e.g., C = 12.0107, H = 1.00794, O = 15.9994 etc. [17].	Relative atomic mass (is the recommended term to use [28,29]), Average molecular mass; molecular mass.
Average absolute mass measurement accuracy	$\dfrac{\sum_i (|\Delta m_i|)}{n}$ The average experimental mass is calculated using the absolute values of individual mass errors ($|\Delta m_i|$).	Average mass error (absolute values); average absolute error; average absolute mass error; absolute exact-mass errors.
Average experimental mass	$\overline{m_i} = \dfrac{\sum_i (m_i)}{n}$ The average experimental mass calculated from individual masses (m_i)	Mean mass; sample mean (sample – refers to the set of mass measurements).
Average mass	See atomic weight.	
Bias	Used to denote a "biased sample" of measurements in which a systematic error is present so that individual measurements are not equally distributed about the "true" value. For a fuller discussion refer to the description of "measurement result" in the IPUAC Gold Book [15].	
Confidence level	Probability that, e.g., an experimental mean value will lie within a confidence interval defined by the standard deviation and other statistical parameters	
Confidence limits of the average mass (mean)	$\overline{m_i} = \mu \pm t\left(\dfrac{s}{\sqrt{n}}\right)$ The extremities of the confidence interval, e.g., a mass range in which we are confident, to a given degree usually 99%, 95%, or 90% confidence level, that the true mass value lies.	

Table 4. Continued

Term	Definition and comments	Other terms used*
Degrees of freedom	Number of independent measurements needed to specify a statistical parameter (a more thorough definition can be obtained from references [6] or [18]), e.g., analysis of variance or confidence limits, it is usually n-1 or n-2, where n is the total number of measurements.	
Descriptive statistics	A summary of statistics to describe data in a quantitative manner. Various software packages, (e.g., MS Excel and SPSS) [30] provide standardized analysis reports described as Descriptive Statistics.	
Deviation (mass)	IUPAC [15] defines deviation as "The difference between an observed value and the arithmetic mean of the set to which it belongs."	Mass deviation has been proposed to the proteomics community [16] to describe a single mass measurement error, this usage is incorrect. These authors also define a term maximum mass deviation (MMD) to use in database searching as "This is the cut-off value used in database search. Only peptide sequences with a calculated mass within this tolerance are reported as hits".
Electron mass (m_e)	Experimentally determined as 0.000548579903 u.	
Error (of measurement)	IUPAC [15] definition is "The result of a measurement minus the true value of the measurement."	Mass error is often incorrectly referred to as mass accuracy or even mass deviation.
Exact mass	The calculated mass of an ion whose elemental formula, isotopic composition, and charge state are known. The IUPAC definition constricts the definition to using one specified isotope of each atom involved, usually the lightest isotope [2].	Measured accurate mass; calculated exact mass, theoretical mass.
F-test	A test which can be used to compare the precision of mass measurements of two sets of data. The F-test is a broad ranging test where the test statistic follows an F-distribution, named after RA Fisher.	
Gross errors	Describes (mass) errors so large the data should not be considered.	
IUPAC	International Union of Pure and Applied Chemistry (http://www.iupac.org/), an international organization that provides recommendations on chemical nomenclature and up to date values of atomic masses and other data relevant to mass spectrometry [15, 29, 31].	
Kolmogorov and Smirnov test	A test which can compare one set of data to another set (or a model distribution such as the normal distribution), and provide a measure of the goodness of fit between the two. It can be used to test which type of statistical distribution best fits a set of experimental measurements.	
Mass error	Δm_i. The difference between the accurate measured mass and the expected or calculated exact mass. The term "mass errors" refers to the set of errors arising from individual mass measurements.	
Mass accuracy	See mass measurement accuracy below.	

Table 4. Continued

Term	Definition and comments	Other terms used*
Mass measurement accuracy	$\overline{\Delta m_i} = \dfrac{\sum_i \Delta m_i}{n}$ The average of n individual mass errors (Δm_i), including the sign so that partial cancellation of individual mass errors occurs in the use of this formula.	Mean mass measurement accuracy; average mass measurement accuracy; mass accuracy; average error; average mass error.
Median	The median is the middle value of a list of measurements when it is sorted. For a list of n values it will be the middle value when n is odd, and is taken as the average of the $(n/2)^{th}$ and $(n/2 + 1)^{th}$ values when n is even. It is used in nonparametric statistical tests.	
Milli mass unit (mmu)	An incorrect term for a "milli u" or in SI units "mu" used in the Literature and found in software packages.	The term "δ mmu" is sometimes used to denote mass measurement error of a single reading.
Mode	The mode is the value that occurs most frequently in the measured data set and is used in non-parametric statistical tests.	
Monoisotopic mass	The exact mass calculated using the mass of the most abundant isotope of each element.	
m/z	A three character symbol defined as a dimensionless quantity, formed by dividing the mass of an ion in unified atomic mass units and its charge number. It is written in lower case italics with no spaces, i.e. m/z. Like any SI unit it can have a prefix to signify decimal submultiples (or multiples) e.g., mm/z, this is a milli m/z or 0.001 m/z.	The unit Thomson was used with m/z but is now deprecated.
Nominal mass	The "mass of an ion or molecule calculated using the mass of the most abundant isotope of each element rounded to the nearest integer value and equivalent to the sum of the mass numbers of all constituent atoms" [2]. Nominal masses or masses to one decimal place should be quoted where there is any uncertainly regarding the accuracy of the mass scale. e.g., resulting from a poor or failed mass calibration or if an instrument/software change occurs that could affect the mass scale.	Sometimes referred to as an integer mass, when the value is written without any decimal places.
Null hypothesis	The null hypothesis is a statement, of no difference or no effect, concerning the mass measurement analysis (and its parameters) that we want to test. For example, on comparing the accuracy between different methods, the null hypothesis would be that the accuracies are statistically indistinguishable.	
Normal distribution	$y = \dfrac{\exp(-(x-\mu)^2/2\sigma^2)}{\sigma\sqrt{2\pi}}$ A probability distribution widely used in statistics and also named the Gaussian distribution, after Gauss. It takes on the mathematical form shown, where x is the measurement value (mass), y the frequency or intensity of that measurement, μ is the population mean, and σ the standard deviation of the population.	Gaussian distribution
ppm error	Parts per million mass error.	
Population distribution	This is the mathematical model of the distribution of mass measurements. It may be a normal distribution, although this would require to be established by an appropriate test such as the Kolmogorov Smirnov test or χ^2 (or Chi-squared) test.	
P value	The probability of finding a test statistic at least as large as that observed. P values of 0.05 and 0.01, which are widely used, correspond to confidence levels of 95% and 99%, respectively.	

Table 4. Continued

Term	Definition and comments	Other terms used*
Precision	The repeatability of the measurement reflecting random errors. Random errors $s = \sqrt{\dfrac{\sum_i \left(\Delta m_i - \overline{\Delta m_i}\right)^2}{n-1}}$ cause mass measurements to fall on either side of the average experimental measurement and affect precision of the set of measurements. When a set of mass measurements, of one ion species, lie close together we say the measurements are precise and when not we say the measurements are imprecise.	
R	A computer language developed for statistical computing http://www.r-project.org	
Robust statistics [32]	Statistical methods used in circumstances where the data depart from ideal statistical assumptions, (e.g., normal distribution of data), and where outliers can be treated. Robust statistics are effective in these circumstances, and these statistical methods are referred to as nonparametric methods, but are beyond the scope of this article.	
Root mean square error	An estimator of the magnitude of mass measurement based on the sum of the squares of individual mass errors (Δm_i).	RMS error; mean squared error; root mean square deviation
Relative atomic mass	The IUPAC approved term superceding atomic weight.	
Repeatability	This is the short-term precision of experimental measurements made under similar conditions, i.e., the same instrument, operator, and over a limited time, normally the same day.	
Reproducibility	Refers to differences among experimental measurements made under different circumstances, i.e., a measurement of the same quantity made by different operators, possibly on different instruments, and often with a significant time difference between groups of measurements, e.g., next day, or week, or month.	
Sample distribution	This is the distribution of the actual data measured and is usually plotted as a frequency versus mass measurement (histogram plot).	
Significant figures	The number of digits (or figures) used to report an accurate mass value. This should be one more than the level of mass accuracy of the measurements, e.g., an ion mass measured as 524.26469 u would be quoted, in significant figure notation, as 524.2647 u (four decimal places, seven significant figures), for a mass accuracy of 0.001 u.	
Standard deviation	σ – standard deviation of the theoretical population distribution and s – standard deviation of the experimental sample distribution.	
Standard error of the mean	$s/\sqrt{n}$ For a (large) set of measurements one might divide the data into groups, (e.g., columns), and calculate the mass measurement accuracy for each group. These mass measurement accuracies form a sampling distribution of the mass measurement accuracy. Its mean is the same as the mean value of the whole population with a standard deviation called the standard error of the mean (SEM). SEM gives an estimate of the uncertainly in estimating the population mean (μ) from the sample mean of individual mass measurements $\left(\overline{m_i}\right)$ or for mass errors $\left(\overline{\Delta m_i}\right)$	

Table 4. Continued

Term	Definition and comments	Other terms used*
Student t-statistic	This is the statistic derived from the Student t-distribution; it arises when estimating means or confidence intervals for a normally distributed population when the sample size is small. The Student t-distribution gets closer to the normal distribution as the number of degrees of freedom increase. 'Student' was a pseudonym of W. Gossett who worked at a Guinness brewery in Dublin and was not allowed to publish his statistical work under his own name in order to preserve trade secrets that statistics were used in quality control processes for production of food products (including stout!).	
Systematic errors	Describes (mass) errors that affect the accuracy of the data set.	
Test statistic	This is calculated from the data (mass measurements) and is compared with a critical value of t, i.e., t_c obtained from appropriate statistical tables for the given test, as proof it either contradicts (or proves) the null hypothesis.	
Unified atomic mass unit (u)	One-twelfth of the mass of a ^{12}C atom (^{12}C is defined as a mass of 12 u), where 1 u $\approx$ $1.66054002 \times 10^{-27}$ kg.	Dalton (Da)
Variance	Square of the standard deviation.	Can be qualified as either the population variance (σ^2) or the sample variance (s^2).

*Includes other terms in use, although the reader needs to be aware that some terms may not be appropriate to use.

is still a 5% and 1% chance, respectively, that the correct answer lies outside of the corresponding confidence limits and that there is a corresponding non-zero probability that other elemental compositions (molecular formulae) are the correct ones. Table 3 shows two examples of elemental composition determination for ions with masses measured as 309.1128 and 574.2335 u, both thought to be protonated molecule species [M + H]$^+$. These compounds were mass measured by electrospray ionization using a magnetic sector double-focusing mass spectrometer, whose precision has been previously determined to be 0.81 ppm (1 standard deviation) for ESI positive ion mode, determined from recent archive records [25]. Table 3 shows elemental compositions calculated for the two measured masses, with candidate compositions encompassing $^{12}C_{0-40}$ $^{1}H_{5-60}$ $^{14}N_{0-8}$ $^{16}O_{0-8}$ $^{31}P_{0-2}$ $^{32}S_{0-2}$. The elemental composition output listing was limited to $\pm 3\times$ standard deviation (3s = 2.43 ppm). The correct elemental compositions for Compounds A and B were ranked 7th and 12th, respectively.

Five ppm is often used as a limit for candidate formulae selection, however there are 16 and 81 candidate formulae for Compounds A and B, respectively, using a 5ppm selection criterion. In contrast, using a $\pm$ 3 s criterion reduces the choices whilst retaining a high level of confidence in the correct formulae being present. Application of chemical knowledge of the compounds, i.e., nitrogen rule and the rings and double bonds (RDB) rule, further significantly reduces the number of possible choices. The known element compositions are indicated in Table 3, and both lie within $\pm$ 3 s of the measured mass. It must be noted that there is always a non-zero probability that the correct elemental composition may lie outside of any arbitrarily defined limit; however the chances of this at 95% and 99% confidence levels are 5% and 1%, respectively. The choice of limiting the elemental composition list to ± 3 standard deviations is a rigorous option.

For instruments whose mass measurement precision is well defined and reproducible, the use of the above criteria should be a good method for elemental composition determinations. It is well known that the number of possible element compositions rises "exponentially" with mass, and further information is required to improve the selection of elemental composition, for example, for masses $\leq$ 500 Da a mass accuracy of ~0.1 mDa is required for unique determination of elemental composition even when limited to C, H, N, O, and S [26]. Coupling of mass measurements to isotope pattern recognition software is currently employed to reduce the number of possible elemental compositions for identification of unknown compounds [27].

Testing for Evidence of Systematic Errors

The standard statistical method to test for evidence of systematic errors is to calculate the value of |t| (no regard to sign) and determine whether it exceeds the

expected critical value (obtained from the t-distribution, see Table 2). Using the above formula rewritten

$$t = (\bar{x} - \mu)\frac{\sqrt{n}}{s}$$

Our example

The average experimental mass error ($\bar{x}$ or $\overline{\Delta m_i}$) is 0.0012 Da, the expected mass error $\mu = 0$, and s = 0.0030, thus

$$|t| = |(0.0012 - 0) \times \sqrt{9}/0.003| = 1.2$$

For 8 (n-1) degrees of freedom the critical value of $|t|$, i.e., t_c is 2.31 (95% confidence level or $P = 5\% = 0.05$). As the observed value of $t \ll t_c$ there is no evidence of systematic error, i.e., only random errors affect the measurement.

Comparison of Average Experimental Masses (Means) Obtained by Two Different Methods (i.e., Representing Two Sample Sets)

This is referred to as the comparison of means of two datasets. In this case, the two means are $\bar{x}_1$ and $\bar{x}_2$, the null hypothesis (the "safe" or "no effect" hypothesis) adopted is that $(\bar{x}_1 - \bar{x}_2)$ does not differ significantly from zero, i.e., the average experimental masses given by the two methods are statistically indistinguishable. A pooled standard deviation, s_p, is calculated from the two individual values s_1 and s_2 as follows

$$s_p^2 = ((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2)/(n_2 + n_1 - 2)$$

the value for t is calculated by

$$t = (\bar{x}_1 - \bar{x}_2)/(s_p\sqrt{(1/n_1 + 1/n_2)})$$

and t has $(n_2 + n_1 - 2)$ degrees of freedom (essentially n-1 for each sample set representing the two means).

An example

An established method is compared to a new method
Established method: MMA = 0.0012 Da; s = 0.0030 Da

New method: MMA = 0.0006 Da; s = 0.0014 Da

For each method nine measurements were made. The pooled standard deviation is

$s_p{}^2 = (8 \times 0.0030^2 + 8 \times 0.0014^2) /16$, thus $s_p = 0.0023$ Da

From the above equation $|t| = |(0.0012 - 0.0006)/(0.0023\sqrt{(1/9 + 1/9)})| = 0.55$

The critical value t_c, for $(n_2 + n_1 - 2) = 16$ dof is 2.12. Because this value of t_c is greater than the experimental value ($t = 0.54$), the two methods have no significant difference (do not have statistically distinguishable mean values), and it can be concluded that the methods give similar mass accuracy. Hence, the new method can not be considered more (or less) accurate than the existing method and there is no reason to replace the existing method on the basis of only accuracy improvement.

F-test Comparison of the Precision (Standard Deviations) of Two Sample Sets

The F-test uses the ratio of the sample variances (s^2) to consider if the precision of two methods is different, i.e., to test if a new method is more precise

$$F = \frac{s_1^2}{s_2^2}$$

An example

Taking the example from a section above, the F-statistic value is written as $F_{8,8}$, the two subscripts denote the degrees of freedom of the two sets of experimental measurements, and they do not have to be equal in value.

$F_{8,8} = 0.003^2/0.0014^2 = 4.59$, this is a one-tailed test (as we are only concerned if the new method is more precise a one-tailed test is appropriate). However, a two-tailed test would be required if it was required to test if the two methods differ in precision. From tables of the F-statistical test (one-tailed) $F_{8,8} = 3.44$ at 95% confidence level [6].

The calculated statistic for the experimental data is greater than the F-statistic and the variance of the second method is significantly greater then the first so the proposed method is more precise. Hence, the existing method could be replaced with the new method, although the difference between the calculated means is not significant so one would not expect an increase in accuracy only an improvement in precision.

Conclusions

The basic terminology relating to accurate mass measurement has been described and defined (Table 4). Both familiar terms and less common terms, which still appear in the literature and text books are included. It has been emphasized that the terms *accurate mass* and *exact mass* may be written as **measured** *accurate mass* and **calculated** *exact mass*, although the adoption of correct terminology is not an aim of this article and could be dealt by an expansion of the recommended terminology defined by IUPAC. Significant figures used in reporting of mass spectrometry data are often inadequate, and mass lists in Da (or u) or m/z should be given to at least four decimal places (seven significant figures for masses between 100 and 999 Da) for mass uncertainties of 0.001 Da. It is suggested that a further decimal place is included for masses below mass 400 Da to ensure no rounding errors arise in any subsequent calculations

using that accurate mass dataset. Similarly, corresponding uncertainties in mass measurement errors quoted in mDa, or mu (milli u) or ppm should be given to one decimal place (typically two or three significant digits). Some basic significance tests for normally distributed data, and their application to mass spectrometry datasets, are described in detail with appropriate examples.

Acknowledgments

References

1. Gross, M. L. Accurate Masses for Structure Confirmation. *J. Am. Soc. Mass Spectrom.* **1994,** *5,* 57.
2. Murray, K. K.; Boyd, R. K.; Eberlin, M. N.; Langley, G. J.; Li, L.; Naito, Y. *Standard Definitions of Terms Relating to Mass Spectrometry*; International Union of Pure and Applied Chemistry (IUPAC), 2006; 3rd draft document. Online http://www.msterms.com/.
3. Webb, K.; Bristow, A. T. W.; Sargent, M.; Stein, B. K. *Best Practice Guide: Methodology for Accurate Mass Measurement of Small Molecules*; LGC Ltd.: London, 2004.
4. Barwick, V.; Langley, J. G.; Mallet, A. I.; Stein, B. K.; Webb, K. *Best Practice Guide for Generating Mass Spectra*; LGC Ltd.: London, 2006.
5. Sparkman, O. D. *Mass Spec Desk Reference*, 2nd ed.; Global View Publishing: Pittsburgh, PA, 2006.
6. Miller, J. C.; Miller, J. N. *Statistics and Chemometrics for Analytical Chemistry*, 5th ed.; Pearson: Harlow, U.K., 2005.
7. Sack, T. M.; Lapp, R. L.; Gross, M. L.; Kimble, B. J. A Method for the Statistical Evaluation of Accurate Mass Measurement Quality. *Int. J. Mass Spectrom. Ion Processes* **1984,** *61,* 191–210.
8. Todd, J. F. J. Recommendations for Nomenclature and Symbolism for Mass Spectroscopy, International Union of Pure and Applied Chemistry (IUPAC), *Pure Appl. Chem.* **1991,** *63,* 1541–1566.
9. Price, P. Standard Definitions of Terms Relating to Mass Spectrometry: A Report from the Committee on Measurements and Standards of the American Society for Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1991,** *2,* 336–348.
10. Mallet, A. I.; Down, S. *Dictionary of Mass Spectrometry*; John Wiley: Chichester, UK, 2009.
11. de Hoffmann, E.; Stroobant, V. *Mass Spectrometry: Principles and Applications*, 2nd ed.; John Wiley: Chichester, UK, 2007.
12. McLafferty, F. W.; Tureček, F. *Interpretation of Mass Spectra*, 4th ed.; University Science Books: Sausalito, CA, 1993.
13. Boyd, R. K.; Basic, C.; Bethem, R. A. *Trace Quantitative Analysis by Mass Spectrometry*; John Wiley: Chichester, UK, 2008.
14. Watson, J. T.; Sparkman, O. D. *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation*, 4th ed.; John Wiley: Chichester, UK, 2007.
15. *IUPAC Compendium of Chemical Terminology—the Gold Book* http://goldbook.iupac.org/ accessed 18 Dec **2009.**
16. Zubarev, R.; Mann, M. On the Proper Use of Mass Accuracy in Proteomics. *Mol. Cell. Proteom.* **2007,** *6,* 377–381.
17. Wieser, M. E. Atomic Weights of the Elements 2005: An IUPAC Technical Report of the Latest Evaluation of Atomic Weight Determinations. *Pure Appl. Chem.* **2006,** *78,* 2051–2066.
18. Sheskin, D. *Handbook of parametric and nonparametric statistical procedures*; CRC Press: Boca Raton, 2004; Chap XII.
19. Lewis, P. D. Describes the Application of Normality Tests in Detail, Using R. In *R for Medicine and Biology*, 1st ed; Jones and Bartlett Publishers: Sudbury, MA, 2010. A Wide Range of Statistical Tests are Available Using the R Computer Language. This is a Free System Developed by the R Foundation for Statistical Computing http://www.r-project.org.
20. Wong, R. L.; Amster, I. J. Sub Part-Per-Million Mass Accuracy by Using Stepwise-External Calibration in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2006,** *17,* 1681–1691.
21. Mihaleva, V. V.; Vorst, O.; Maliepaard, C.; Verhoeven, H. A.; de Vos, R. C. H.; Hall, R. D.; van Ham, R. C. H. J. *Metabolomics* **2008,** *4,* 171–182.
22. Huang, N.; Siegel, M. M.; Muenster, H.; Weissenberg, K. On-Line Acquisition, Analysis, and E-Mailing of High-Resolution Exact-Mass Electron Impact/Chemical Ionization Mass Spectrometry Data Acquired Using an Automated Direct Probe. *J. Am. Soc. Mass Spectrom.* **1999,** *10,* 1212–1216.
23. van der Heeft, E.; Block, Y. J. C.; Beumer, B.; Nijrolder, A. W. J. M.; Stolker, A. A. M.; Nielen, M. W. F. Full-Scan Accurate Mass Selectivity of Ultra-Performance Liquid Chromatography Combined with Time-of-Flight and Orbitrap Mass Spectrometry in Hormone and Veterinary Drug Residue Analysis. *J. Am. Soc. Mass Spectrom.* **2009,** *20,* 451–463.
24. Bereman, M. S.; Lyndon, M. M.; Dixon R. B.; Muddiman, D. C. Mass Measurement Accuracy Comparisons Between a Double-Focusing Magnetic Sector and a Time-of-Flight Mass Analyzer. *Rapid Commun. Mass Spectrom.* **2008,** *22,* 1563–1566.
25. Brenton, A, G; Godfrey, A. R.; Alamri, M.; Stein, B. K.; Williams, C. M.; Hunter, A. P.; Wyatt, M. F. Analysis of Large Historical Accurate Mass Data Sets on Sector Mass Spectrometers. *Rapid Commun. Mass Spectrom.* **2009,** *23,* 3484–3487.
26. Kim, S.; Rodgers, R. P.; Marshall, A. G. Truly "Exact" Mass: Elemental Composition Can Be Determined Uniquely from Molecular Mass Measurement at ~0.1 mDa Accuracy for Molecules up to ~500 Da. *Int. J. Mass Spectrom.* **2006,** *251,* 260–265.
27. Ojanpera, S.; Pelander, A.; Pelzing, M.; Krebs, I.; Vuori, E.; Ojanpera, I. Isotopic Pattern and Accurate Mass Determination in Urine Drug Screening by Liquid Chromatography/Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2006,** *20,* 1161–1167.
28. De Bièvre, P.; Peiser, H. S. An IUPAC Sponsored *Article on Atomic Weight—the Name, Its History, Definition, and Units*. Pure Applied Chem. **1992,** *64,* 1535–1543.
29. Mills, I.; Cvitas, T.; Homann, K.; Kallay, N.; Kuchhitsu, K. *Quantities and Symbols Used in Physical Chemistry*; Blackwell Scientific Publications: Oxford, 1989.
30. MS Excel. Microsoft Corp., Redmond, WA, SPSS software ver. 13.0.1. SPSS Inc., Chicago, IL.
31. McNaught, A. D.; Wilkinson, A. *IUPAC Compendium of Chemical Terminology*, 2nd ed; Blackwell Science Publications: Oxford, U.K., 1997.
32. Ripley, B. D. Robust Statistics—How not to reject outliers. Part 1. Basic Concepts. Analyst **1989**, *114,* 1693–1697. (For example, see the recommendations by the Analytical Methods Committee of the Royal Society of Chemistry, Burlington House, Piccadilly, London).