Data Presentation in Biology

<u>Units</u>

The international system of units should be used wherever possible, although the main consideration is that units should be fit for purpose. It is, for example, preferable to use minutes rather than seconds in some instances such as when assessing the effect of exercise on heart rate or the rate of transpiration, or cm^3 rather than m^3 for depicting the volume of carbon dioxide produced by respiring yeast cells. Non metric units such as inches or cups should not be used.

<u>Tables</u>

Tables are designed to lay out the data ready for analysis. The table should have an explanatory title. "Table of results" is not an explanatory title, whereas "Table to show the time taken to produce 1 cm^3 of oxygen at different concentrations of carbon dioxide by *Elodea*" describes the nature of the data collected. Other points to note are:

- units should only appear in cell headings rather than in the body of the table
- error for the instrument used or the accuracy of the reading should appear in the cell heading if relevant
- the independent variable should be in the first column
- subsequent columns should show the results for the dependent variable
- decimal places should be consistent throughout a column
- mean values should not have more decimal places than the raw data used to produce them.

The methods used to process the data should be easy to follow and the processed data may be included in the same table as the raw data, there is no need to separate them.

<u>Graphs</u>

Graphs should be clear, easy to read and interpret with an explanatory title. If IT software is used, the graph should have clearly identifiable data points and demarcated and labelled axes of a suitable scale.

Adjacent data points should be joined by a straight line and the line should start with the first data point and end with the last one, as there should be no extrapolation beyond these points. Lines of best fit are only useful if there is good reason to believe that intermediate points fall on the line between two data points. The usual reason for this is the collection of a large amount of data, which is often not possible given the time constraints of investigations at this level. Likewise, extrapolation of the line will only make sense if there is a large amount of data and a line of best fit is predicted or there is reference made to the literature values. Students should exercise caution when making assumptions.

Finally, the type of graph chosen should be appropriate to the nature of the data collected.

Data Presentation in Biology

Error

There are sources of error at a number of stages of any investigation. The chosen method should try to address as many as possible by considering the control of variables, but despite this, many will remain. Students should not be discouraged by this, experimental results are only samples (see NOS section 3, "The objectivity of science" in the *Biology guide*), but rather take them into consideration when analysing the data and drawing conclusions. A thorough evaluation of the sources of uncertainty and error will also help to gain perspective on the investigation in general and to suggest potential improvements and extensions.

Random variation and normal variation

In biological investigations, errors can be caused by changes in the material used or by changes in the conditions under which the experiment is carried out. Biological materials are particularly variable. For example, the water potential of potato tissue may be calculated by soaking pieces of tissue in a range of concentrations of sucrose solutions. However, the pieces of tissue will vary in their water potential, especially if they have been taken from different potatoes. Pieces of tissue taken from the same potato will also show variations in water potential, but they will probably show a normal variation that is less than that from samples taken from different potatoes. Random errors can, therefore, be kept to a minimum by careful selection of material and by careful control of variables. For example, use of a water bath to reduce the random fluctuations in ambient temperature.

Human errors

Making mistakes is not an acceptable source of error if they could have been easily avoided with more due care and attention. Data loggers can be used if a large number of measurements need to be made, to avoid errors arising due to loss of concentration. Careful planning can help reduce this risk.

The act of measuring

When a measurement is taken, this can affect the environment of the experiment. For example, when a cold thermometer is put into a test tube with only a small volume of warm water in it, the water will be cooled by the presence of the thermometer so it would be sensible to scale up the volume or have the thermometer in the solution from the start. If the behaviour of animals is being recorded, the presence of the experimenter may influence the animals' behaviour. Although there are ways to reduce the impact of observer influences, it may have to be something that is taken into account later.

Data Presentation in Biology

Systematic errors

Systematic errors can be reduced if equipment is regularly checked or calibrated to ensure that it is functioning correctly. For example, a thermometer should be placed in an electronic water bath to check that the thermostat of the water bath is correctly adjusted. A blank should be used to calibrate a colorimeter to compensate for the drift of the instrument.

Degrees of precision and uncertainty in data

Students must choose an appropriate instrument for measuring such things as length, volume, pH and light intensity. This does not mean that every piece of equipment needs to be justified, and it can be appreciated that, in a normal science laboratory, the most appropriate instrument may not be available.

For the degrees of precision, the simplest rule is that the degree of precision is plus or minus ($\pm$) the smallest division on the instrument (the least count). This is true for rulers and instruments with digital displays.

The instrument limit of error is usually no greater than the least count and is often a fraction of the least count value. For example, a burette or a mercury thermometer is often read to half of the least count division. This would mean that a burette value of 34.1 cm^3 becomes 34.10 cm^3 (±0.05 cm^3). Note that the volume value is now cited to one extra decimal place so as to be consistent with the uncertainty.

The estimated uncertainty takes into account the concepts of least count and instrument limit of error, but also, where relevant, higher levels of uncertainty as indicated by an instrument manufacturer which is usually obtainable online, or qualitative considerations such as parallax problems in reading a thermometer scale, reaction time in starting and stopping a timer, or random fluctuation in an electronic balance read-out. Students should do their best to quantify these observations into the estimated uncertainty.

Other protocols exist and no specific protocol is preferred as long as it is clear that recording of uncertainties has been undertaken and the uncertainties are of a sensible and consistent magnitude.

Propagating errors

Propagating errors during data processing is not expected but it is accepted provided the basis of the experimental error is explained.

Data Presentation in Biology

Replicates and samples

Biological systems, because of their complexity and normal variability, require replicate observations and multiple samples of material. As a rule of thumb, the lower limit is five measurements within the independent variable, with three runs for each. This will produce five data points for analysis. So in an investigation into the effect of temperature on the rate of reaction of an enzyme, temperature is the independent variable (IV) and the rate of reaction the dependent variable (DV). The IV would need to be assessed three times at five different temperatures at the very least. Obviously, this will vary within the limits of the time available for an investigation. Some simple investigations permit a large number of measurements, or a large number of runs. It is also possible to use class data to generate sufficient replicates to permit adequate processing of the data in class, non-assessed practical work.

The standard deviation is the spread of the data around the mean. The larger the standard deviation the wider the spread of data is. Standard deviation is used for normally distributed data. This makes it useful for showing the general variation/uncertainty around a point on a line graph, but it is less helpful for identifying potential anomalies.

Error bars that plot the highest and the lowest value for a test, joined up through the mean that will form the data point plotted on the graph with a vertical line, will allow the variation/uncertainty for each data set to be assessed. If the error bars are particularly large, then it may show that the readings taken are unreliable (although reference to the scale might be needed to determine what large actually is). If the error bars overlap with the error bar of a previous or subsequent point, then it would show that the spread of data is too wide to allow for effective discrimination. If trend lines are possible, then adding the coefficient of determination (R^2) can be helpful as an indication of how well the trend line fits the data.

Statistics

An effective presentation of the data goes a long way to assessing whether or not a trend is emerging. This is, however, not the same as using statistics to assess the nature of such a trend and whether it is significant—in other words, whether a trend, judged subjectively from a graph, is actually valid. Students are encouraged to use a statistical test to assess their data, but should briefly explain their choice of test, outline the working hypothesis and put the results of the test into the context of their investigation. For statistical tests the correct protocol should be presented including null and alternative hypotheses, degrees of freedom, critical values and probability levels.