CHAPTER 3: NULL HYPOTHESES

3.1 INTRODUCTION

The following sections describe the concept of the null hypothesis and associated errors. Specific null hypotheses are proposed for the NTMP that are compatible with a possible resource classification system. The implications of making errors are also described. This provides guidance on the choice of monitoring variables and the guidelines (that are used to assess monitoring data), particularly in the context of sensitivity of tests and possible bias introduced by the guidelines.

3.2 INTERPRETING HYPOTHESES

3.2.1 Terminology

The following sections show how to interpret various null hypotheses (H_0) and the corresponding test or alternative hypotheses (H_a). Usually the alternative hypothesis states what we hope to be true. The null hypothesis states the opposite of this. For example, in respect of toxicity, we would hope that there is not a toxicity problem in a given water resource.

When reporting the results of monitoring, two kinds of errors can be made. These are most usefully referred to as "false negative" and "false positive" results. The word "positive" refers to reporting that the null hypothesis (whatever it might be) is true.

- A "false negative" refers to reporting a negative result (*i.e.* that the null hypothesis is false) when it should have been reported as positive.
- A "false positive" result refers to incorrectly reporting a positive result (*i.e.* the result was actually negative).

These terms are fairly intuitively obvious and therefore should be used in communication. Statisticians refer to these respectively as "Type I errors" (or the Greek letter α) and "Type II errors" (or the Greek letter β). However, these are not useful terms to use as they do not convey any meaning in themselves (except to those who use them frequently) and so will not be used further here.

Statisticians also relate the "confidence level" to the probability of a false negative as follows:

$$\text{Confidence level (\%)} = 100 - \text{probability of false negative (\%)}$$

This means the higher the required confidence, the lower the probability of a false negative.

Statisticians also talk about "power". This is related to the probability of a false positive as follows:

$$\text{Power (\%)} = 100 - \text{probability of a false positive (\%)}$$

This means the higher the power, the lower the probability of a false positive.

3.2.2 Causes of errors

A number of factors cause false negative and false positive errors. The following table summarises some of them. It assumes that the null hypothesis is simply "there is a toxicity problem".

Table 3.1. Possible causes of false negative and false positive results.

	Causes of false negative results	Causes of false positive results
Sampling method	Snapshot sampling that may miss toxicant peaks or spikes.	
Sampling site		Sampling in an effluent plume or mixing zone (giving an unrepresentative sample).
Sensitivity	Insensitive tests.	Overly sensitive tests.
Bias	Using toxic criteria that are too lenient.	Using toxic criteria that are too strict.
Variability	Highly variable toxicity test results in the vicinity of the toxic criterion can report no problem when there is actually a problem (see figure).	Highly variable toxicity results in the vicinity of the toxic criterion can report a problem when there is actually no problem (see figure).

The *sampling method* may contribute to false negatives if water samples are taken at times that miss peaks or spikes of toxicant concentrations. In such cases, the measurement (of either toxicants or toxicity) may indicate little or no problem while organisms in the water resource have indeed been impacted by previous peaks.

The choice of *sampling site* is particularly important in our seasonal rivers and when point source effluent discharges create local plumes of high concentrations in the mixing zone.

Sensitivity relates directly to the choice of monitoring variable, particularly toxicity tests. A test organism may be chosen that is very sensitive to a stressor (*i.e.* a toxicant) relative to organisms present in a water resource. This inherently means an increased likelihood of false positive results being reported. They are "false" because although the response of the test organism indicates a problem, the others organisms may indicate no response because they are less sensitive to the stressor.

Similarly, choosing a test organism that is less sensitive to a stressor than other organisms increases the likelihood of a false negative result.

Bias can be interpreted similarly. Using a toxic criterion that is very strict (*e.g.* has been made conservative by application of the precautionary principle) increases the likelihood of a false positive. If it is lenient, then the likelihood of false negatives increases.

Variability can be caused by either of the following:

- Inherent variability of the sampling and analytical procedures. These involve random errors over which one has little or no control.
- Natural variability of the water resource being sampled.

The following figure illustrates how variability in a measurement causes false negative and false positive results. (Hypothetical values have been chosen although such a high variability is possible with certain toxicity tests. The "true value" is also not always in the centre of the range.)

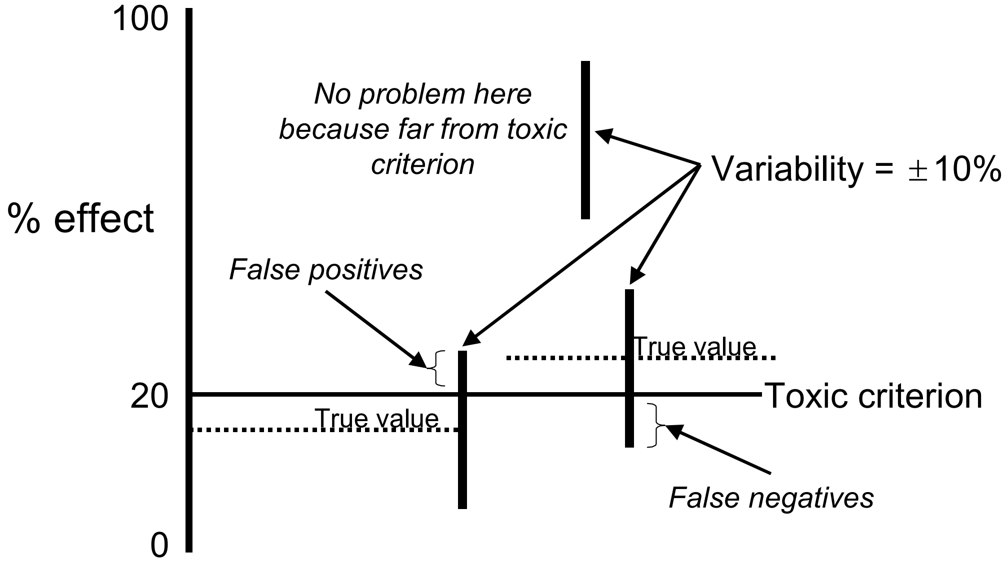

Figure 3.1. Illustration of how variability can cause false negative and false positive results.

3.2.3 Managing errors

Irrespective of which null hypothesis is being tested, false negative and false positive results are errors whose implications should be carefully considered. The following sub-sections describe such implications for some specific hypotheses. However, some generic statements can be made about errors irrespective of any specific hypothesis.

- *Design phase*: In general, the more severe the implications of either a false negative or false positive result, the greater the effort should be to reduce the probability of making such errors in the first place. This requires (a) identifying when specific errors might have unacceptable consequences and (b) designing the monitoring programme to ensure that, under these circumstances, the probability of making the errors is reduced to acceptable levels. This means possible errors need to be addressed before any monitoring takes place, that is, in the design phase.

- *Implementation phase*: When monitoring results begin to be reported, the Department should be aware of the likely uncertainty in any particular result. Specifically it should be aware of the probability of either a false negative or a false positive result. It is therefore in a position to react in a manner appropriate to this known level of uncertainty. The less the uncertainty (*i.e.* the greater the confidence in the result) the more decisive a response by the Department can be. For example, if the probability of a false positive result is 5% or less (and a "positive" result indicates possible problems) then more costly responses would be justified than if the probability was 20%. Equivalently, a high probability of a false positive would suggest greater caution in the response. The same applies to a result that suggests that no action need be taken. This result may be in error. That is, actually a response is required. Again the Department needs to recognise situations when not responding may have serious consequences. As with false positives, the degree to which the Department can be confident that not responding is acceptable depends on the probability of a false negative result.

3.3 SPECIFIC NULL HYPOTHESES

3.3.1 Introduction

Linking the NTMP to the classification system automatically identifies appropriate null hypotheses. It is assumed that it is appropriate to separate the contexts of ecosystem integrity and fitness for use, as indicated in the previous chapter, and identify two null hypotheses addressing status and one hypothesis addressing trends. These are addressed in the following sections.

3.3.2 Null hypothesis for status of ecosystem integrity

3.3.2.1 Definition

The following table defines a possible null hypothesis that allows statements to be made about the status of the ecosystem integrity of a specific water resource. It also defines what is meant by a false negative result and a false positive result.

Table 3.2. Ecosystem integrity null hypothesis.

Null hypothesis	**H_0 = Resource is not in an acceptable ecological category**
Alternative hypothesis	H_a = Resource is in an acceptable ecological category
False negative result	Report resource is in an acceptable category when it is actually in a worse category
False positive result	Report resource is in a worse category when it is actually in an acceptable category

When considering random variability in data only (*i.e.* excluding the sensitivity and bias of measurements and associated toxic criteria), then the following table indicates how different probabilities of the errors can be interpreted.

Table 3.3. Interpretation of different probabilities of a false negative or a false positive result.

	Interpretation
Probability of a false negative	The Department is willing to accept that …
5%	… 1 in 20 times …
10%	… 1 in 10 times …
20%	… 1 in 5 times …
	… it will be reported that the resource is in an acceptable category when it is actually in a worse category
Probability of a false positive	The Department is willing to accept that …
5%	… 1 in 20 times …
10%	… 1 in 10 times …
20%	… 1 in 5 times …
	… it will be reported that the resource is in a worse category when it is actually in an acceptable category

3.3.2.2 Consequences of "false negatives"

The higher the probability of a false negative result, the greater the chances are of reporting that the resource is in an acceptable category when it is actually in a worse category. The NTMP is concerned with providing a degree of protection to selected target organisms and aquatic ecosystem integrity. A high probability of a false negative result increases the chances of such protection not being provided because actual problems are not being detected. The protection is not provided simply because corrective action is not taken because there is apparently no problem. (A "red flag" was simply not raised.)

In essence, an excessive number of false negative results can result in the degree of water resource protection being inadequate to sustain the desired level of ecosystem integrity.

3.3.2.3 Consequences of "false positives"

The higher the probability of a false positive result, the greater the chances are of reporting the resource is in a worse category when it is actually in an acceptable category. The consequences of such an error depend on how (and when) the Department reacts to such information. Many possible scenarios exist.

There may be a response mechanism required for a single sample for which a positive result has just been obtained. The response may be to confirm the result. However, there are again a number of issues that determine whether confirmation is possible and, if so, how it might be done.

- If the result was based on a water sample, should the result be confirmed on the same water sample? Does the water sample still exist? Is it acceptable to use the same sample so long after the sample was taken?
- If the result was based on a biomarker or a bioaccumulation measurement, is the original specimen still available and appropriate to use?
- If the result was based on an active (in stream) measurement, how appropriate is it to repeat the measurement?
- Should the same test be performed or a different one?

An alternative may be to choose not to attempt to confirm individual results. The rationale for this might be that the NTMP takes a strategic perspective with an emphasis on long-term trends and large spatial scale reporting. However, the question still remains that if the individual result is, for example, reported back to the local area it is conceivable that the local stakeholder may decide to take some action. The costs of this action, although now possibly regarded as outside the scope of the NTMP, remain a real consequence of the false positive result. The least that can be done in these circumstances is to report the probability of the result being false to the local stakeholder so that they can use this to make their own judgements on whether to act upon it or not.

An alternative might be to only react to results in an annual report and not to individual measurements. In this case, the results can probably not be confirmed in any way so long after the event. So the individual results would simply be accepted and reported "as is". However, again, their associated probability of being false can be reported.

A possible response might be to act upon an apparent problem by designing and implementing a local monitoring programme with the objectives to (a) confirm there is still a problem and possibly its extent and (b) possibly establish the most likely source. (Note: Such local monitoring programmes are not the direct responsibility of the NTMP.) However, such exercises require considerable resources, both financial and human. The greater the probability of a false positive the greater the chances are that such resources will be wasted because no problem actually existed.

A specific pollution source (say an industrial or agricultural source) may be strongly suspected because such a source exists upstream of the monitoring site at which the apparent problem was reported. Another response may be to confront the suspected polluter with the monitoring results in order to get their involvement in confirming the problem and possibly its extent (even though identification of pollution sources is explicitly NOT regarded as being within the mandate of the NTMP). The greater the probability of a false positive result the greater the chances that such a confrontation would be totally unnecessary since no problem actually existed. Such confrontations would not only again waste resources (this time possibly including that of the suspected polluter) but would almost certainly sour relations somewhat between the suspected polluter and the Department.

There is therefore a wide range of possible scenarios relating to the implications of false positive results. In essence, **an excessive number of false positive results could directly decrease the cost-effectiveness and financial sustainability of the NTMP** because unnecessary resources may be allocated to confirm or follow up apparent problems that are, in reality, not problems.

3.3.3 Null hypothesis for status of fitness for use

3.3.3.1 Definition

The following table defines a possible null hypothesis that allows statements to be made about the status of the fitness for use of a specific water resource.

Table 3.4. Fitness for use null hypothesis.

Null hypothesis	**H_0 = the resource is not in an acceptable water use class**
Alternative hypothesis	H_a = the resource is in an acceptable water use class
False negative result	Report the resource is in an acceptable water use class when it is actually not
False positive result	Report the resource is not in an acceptable water use class when it is actually acceptable

When considering random variability in data only (*i.e.* excluding the sensitivity and bias of measurements and associated toxic criteria), then the following table indicates how different probabilities of the errors can be interpreted.

Table 3.5. Interpretation of different probabilities of a false negative or a false positive result.

	Interpretation
Probability of a false negative	The Department is willing to accept that …
5%	… 1 in 20 times …
10%	… 1 in 10 times …
20%	… 1 in 5 times …
	… it will be reported that the resource is in an acceptable water use class when it is actually not
Probability of a false positive	The Department is willing to accept that …
5%	… 1 in 20 times …
10%	… 1 in 10 times …
20%	… 1 in 5 times …
	… it will be reported that the resource is not in an acceptable water use class when it is actually acceptable

3.3.3.2 Consequences of "false negatives"

A false negative result would report that the resource is fit for use when it is actually not fit for use. This would mean that adequate protection may not be provided to the water users of concern simply because there is apparently no problem. **An excessive number of false negative results increases the likelihood of negative impacts on the water users of concern and hence socio-economic enhancement and optimal water use.**

3.3.3.3 Consequences of "false positives"

A false positive result would report that the resource is not fit for use when it is actually fit for use. The consequences of this depend on how the Department reacts to such a "red flag". However, the same issues as noted above for false positives for ecosystem integrity are likely to apply equally well in this context.

Therefore, as above, an excessive number of false positive results could directly decrease the cost-effectiveness and financial sustainability of the NTMP.

3.3.4 Null hypothesis for trends

3.3.4.1 Definition

This hypothesis requires the current year's statistic being compared with the same statistic from the previous year. The statistic could be:

- The annual mean toxicant concentration at a given monitoring site.
- The annual mean toxicity (*e.g.* % lethality) at a given monitoring site.

These could apply to either ecosystem integrity or fitness for use, depending on which variables are used. The variables chosen for ecosystem integrity necessarily give information on trends in ecosystem integrity. The same applies to fitness for use.

The purpose of comparing one year's statistic with the previous year's statistic could simply be to determine whether there has been a general improvement or worsening in the monitoring site's status over a one-year period.

Table 3.6. Possible trend null hypothesis.

Null hypothesis	**H_0 = the current status is worse than last year**
Alternative hypothesis	H_a = the current status is the same/better as last year
False negative result	Report same/better status when it is actually worse
False positive result	Report status is worse when it is actually the same/better

When considering random variability in data only, then the following table indicates how different probabilities of the errors can be interpreted.

Table 3.7. Interpretation of different probabilities of a false negative or a false positive result.

	Interpretation
Probability of a false negative	The Department is willing to accept that …
5%	… 1 in 20 times …
10%	… 1 in 10 times …
20%	… 1 in 5 times …
	… it will be reported that the current status is the same/better when it is actually worse
Probability of a false positive	The Department is willing to accept that …
5%	… 1 in 20 times …
10%	… 1 in 10 times …
20%	… 1 in 5 times …
	… it will be reported that the current status is worse when it is actually the same/better

3.3.4.2 Consequences of "false negatives"

A false negative result would report that the current status is the same/better as the previous year when it is actually worse. **An excessive number of false negative results can directly:**

- **Result in the degree of water resource protection being inadequate to sustain the desired level of ecosystem integrity, or**
- **Increase the likelihood of negative impacts on the water users of concern and hence socio-economic enhancement and optimal water use.**

3.3.4.3 Consequences of "false positives"

A false positive result would report that the current status is worse than the previous year when it is actually the same/better. The consequences of this again depend on how the Department reacts to such information. Two specific scenarios are envisaged:

- The apparent deterioration in status does not indicate a change to an unacceptable category or class. Although an apparent deterioration has taken place, since the resource remains in the same category or class, this is less cause for concern than the following scenario. No action is therefore somewhat justified.
- The apparent deterioration in status indicates that either the present ecological state or water use class changes to an unacceptable category or class. Since the borderlines between categories and class are, by their very nature, "thresholds of concern", a change from one category/class to a worse one is a significant "red flag". A decisive response by water resource managers would therefore be expected. These may include designing and implementing local monitoring programmes or possibly even approaching suspected polluters. In either case, significant costs may be incurred.

Since the current situation would arise only after an annual report is produced, confirmation of the result would typically not be possible.

As above, an excessive number of false positive results could directly decrease the cost-effectiveness and financial sustainability of the NTMP.

3.4 RECOMMENDATIONS

It is strongly recommended that the implications of false positives and false negatives be carefully considered for each monitoring variable that is chosen for the NTMP. Although it may be difficult, if not impossible, to quantify the real probabilities of these errors occurring, at least some qualitative consideration should be given to them. Furthermore, the above concepts also provide a framework for assessing the ramifications (in terms of sustainable development and sustainability of the NTMP itself) when choosing the sensitivity of test species and the strictness of toxic criteria. It is therefore also strongly recommended that the ramifications are borne in mind when these issues are being considered.

3.1 INTRODUCTION 1

3.2 INTERPRETING HYPOTHESES 1

3.2.1 Terminology 1
3.2.2 Causes of errors 2
3.2.3 Managing errors 3

3.3 SPECIFIC NULL HYPOTHESES 4

3.3.1 Introduction 4
3.3.2 Null hypothesis for status of ecosystem integrity 4
3.3.2.1 Definition 4
3.3.2.2 Consequences of "false negatives" 4
3.3.2.3 Consequences of "false positives" 5
3.3.3 Null hypothesis for status of fitness for use 6
3.3.3.1 Definition 6
3.3.3.2 Consequences of "false negatives" 6
3.3.3.3 Consequences of "false positives" 7
3.3.4 Null hypothesis for trends 7
3.3.4.1 Definition 7
3.3.4.2 Consequences of "false negatives" 8
3.3.4.3 Consequences of "false positives" 8

3.4 RECOMMENDATIONS 8