RTP
Regulatory Toxicology and Pharmacology
Official Journal of the International Society of Regulatory Toxicology and Pharmacology
ScienceDirect

An evaluation of the fixed concentration procedure for assessment of acute inhalation toxicity

Fiona Sewell, Ian Ragan, Ian Indans, Tim Marczylo, Nigel Stallard, David Griffiths, Thomas Holmes, Paul Smith, Graham Horgan

PII: S0273-2300(18)30001-1

DOI: 10.1016/j.yrtph.2018.01.001

Reference: YRTPH 4029

To appear in: *Regulatory Toxicology and Pharmacology*

Received Date: 22 August 2017

Revised Date: 1 January 2018

Accepted Date: 3 January 2018

Please cite this article as: Sewell, F., Ragan, I., Indans, I., Marczylo, T., Stallard, N., Griffiths, D., Holmes, T., Smith, P., Horgan, G., An evaluation of the fixed concentration procedure for assessment of acute inhalation toxicity, *Regulatory Toxicology and Pharmacology* (2018), doi: 10.1016/j.yrtph.2018.01.001.

An evaluation of the Fixed Concentration Procedure for assessment of acute inhalation toxicity

Fiona Sewell[a] Ian Ragan[b], Ian Indans[c], Tim Marczylo[d], Nigel Stallard[e], David Griffiths[f], Thomas Holmes[g], Paul Smith[h], Graham Horgan[i]

[a,]*National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs), UK

[b]Board member, NC3Rs

[c]Health and Safety Executive, UK

[d]Public Heath England, UK

[e]University of Warwick, UK

[f]Envigo, UK

[g]Exponent International Ltd, UK

[h]Charles River Laboratories Edinburgh Ltd., UK

[i] Biomathematics & Statistics Scotland (BioSS), UK

*Corresponding author. Email address: Fiona.Sewell@nc3rs.org.uk

Abstract

Acute inhalation studies are conducted in animals as part of chemical hazard identification and for classification and labelling. Current methods employ death as an endpoint (OECD TG403 and TG436) while the recently approved fixed concentration procedure (FCP[1]) (OECD TG433) uses fewer animals and replaces lethality as an endpoint with evident toxicity. Evident toxicity is the presence of clinical signs that predict that exposure to the next highest concentration will cause severe toxicity or death in most animals. Approval of TG433 was the result of an international initiative, led by the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs), which collected data from six laboratories on clinical signs recorded for inhalation studies on 172 substances. This paper summarises previously published data and describes the additional analyses of the dataset that were essential for approval of the TG.

Highlights:

The FCP for acute inhalation toxicity has been accepted by OECD as TG433.

TG433 uses evident toxicity while other approved methods use lethality.

A sighting study with 1 M and 1 F animal reliably identifies the more sensitive sex.

The three methods (LC_{50}, ATC, FCP) showed good agreement in a retrospective analysis.

Keywords:

[1] Abbreviations: FCP, fixed concentration procedure; LC_{50}, concentration causing death in 50% of animals tested; GHS, global harmonised system; ATC, acute toxic class; NC3Rs, National Centre for the 3Rs; PPV, positive predictive value; Cl; confidence limits; MTD, maximum tolerated dose; TC_{50}; concentration causing toxicity in 50% of animals tested

Acute inhalation studies; 3Rs; Evident toxicity; Fixed concentration procedure (FCP); Refinement; Regulatory toxicology; TG403; TG436; TG433.

1. Introduction

Acute inhalation studies are conducted in animals as part of chemical hazard identification and for classification and labelling purposes. There has been considerable work towards refining the existing methods so that 'evident toxicity' rather than death can be used as an endpoint, through the use of the fixed concentration procedure (FCP) (OECD, 2004). This has recently been accepted as OECD test guideline (TG) 433 as an alternative to the currently accepted LC_{50}[2] and the Acute Toxic Class (ATC) methods (OECD TGs 403 and 436 respectively) (OECD, 2009a; OECD, 2009b). The FCP also has the potential to use fewer animals, due to the use of a single sex, and fewer studies overall, as it will obviate the need to test at the next concentration up in some cases. The principles of the three methods are summarised in Table 1 and are described in more detail in Sewell *et al.* (2015). In brief, the LC_{50} method involves testing at three or more concentrations to enable construction of a concentration-mortality curve and a point estimation of the LC_{50} which allows classification into one of five toxic classes using the globally harmonised system (GHS) of classification and labelling of chemicals (OECD, 2001) (Table 2). The ATC method is a refinement of the LC_{50}. Rather than a point estimate of the LC_{50}, this method estimates which toxic class the LC_{50} falls within, so that classification can be assigned. It uses an 'up-and-down' procedure to test up to four fixed concentrations from the boundaries of the categories (or toxic classes) in the GHS classification system. Depending on the number of deaths at each concentration further testing may be required, or a classification can be made. The FCP uses a similar up-and-down approach to the ATC, but instead identifies an exposure concentration that causes evident toxicity rather than death, so that the LC_{50} can be inferred (based on the prediction of death at the next fixed higher concentration). Classification can then be assigned according to the GHS criteria using the predicted LC_{50}. Figures 1, 2 and 3 summarise the possible study outcomes and the resulting classifications for the LC_{50}, ATC and FCP methods

[2] the concentration that is expected to result in the death of 50% of the animals

respectively, using a starting concentration of 5mg/L for dusts and mists as an example (Price et al., 2010).

The FCP was removed from the OECD work plan in 2007 because of three main concerns: the ill-defined and subjective nature of evident toxicity; the lack of evidence for comparable performance to the LC_{50} and ATC methods; and suspected sex differences (the FCP originally proposed the default use of females). Concerns about the definition of ‘evident toxicity’ were raised despite its long use in the Acute Oral Fixed Dose Procedure (OECD TG420) without guidance on what constitutes evident toxicity, nor in the dermal toxicity equivalent of this TG (OECD TG434) which was approved in 2017 without similar guidance. However, all the concerns about the FCP have been resolved through the work of a global initiative led by the UK National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) resulting in its acceptance in April 2017.

Some of the work that led to this decision has already been published (Sewell *et al.*, 2015). This previous paper described analyses of a large data set of acute inhalation studies using the LC_{50} or ATC methods in which signs predictive of death at the next highest concentration (i.e. evident toxicity) were identified. Further analyses were needed to address fully the points noted above and to satisfy concerns raised by the OECD national coordinators during the consultation process, and were therefore vital for the final acceptance of the FCP method by OECD. These included further support for the robustness of the signs previously identified, new statistical calculations to support the value of the sighting study in choosing the most sensitive sex, and retrospective classifications to compare outcomes obtained using the three methods. This paper summarises the previously published data and presents the new analyses that formed the basis for acceptance of the new test guideline.

2. The robustness of evident toxicity as an endpoint

2.1 Definitions

Evident toxicity is an accepted endpoint in the fixed dose procedure for acute oral toxicity studies (OECD TG420) (OECD, 2002a). Here evident toxicity is defined as "*a general term describing clear signs of toxicity following the administration of test substance, such that at the next highest fixed dose either severe pain and enduring signs of severe distress, moribund status or probable mortality in most animals can be expected*." However, for this accepted test guideline, no further guidance has been provided on what constitutes 'evident toxicity', and it is not clear how often this test guideline is being used in practice.

Although evident toxicity was already accepted as an endpoint for this existing test guideline, criticism of this endpoint was a major factor for the withdrawal of the FCP from the OECD work plan in 2007, due to concerns around subjectivity. With the aim of making evident toxicity more objective and transferable between laboratories, the NC3Rs working group collected data on the clinical signs observed in individual animals during acute inhalation studies on 172 substances (Sewell *et al.*, 2015). Because data was collected from a number of laboratories, there was some variation in terminology, requiring retrospective harmonisation by the working group leading to an agreed lexicon of signs (Sewell *et al.*, 2015). These data were analysed to identify signs that could predict lethality would occur if the animals were exposed to the next highest concentration, lethality here being defined as the death, or severe toxicity requiring euthanasia, in two or more animals in a group of five.

There are three important quantities derived from the analysis. The positive predictive value (PPV) is defined as the percentage of times that the presence of a sign correctly predicts lethality at the next highest concentration. A value less than 100% indicates some false positives that would result in over-classification of the substance, undesirable from a business perspective, but erring on the side of caution for human safety. Sensitivity is defined as the proportion of lethality predicted by the presence of the sign at the lower concentration. There is no expectation that a single sign would predict 100% of toxicity at the next higher concentration, but signs with very low levels of sensitivity are less useful because of their rarity and their small contribution to overall evident toxicity. Less than 100% sensitivity indicates some false negatives, that is, lethality occurs at the higher concentration even though the sign was absent at the lower concentration. This does not result in incorrect classification as testing would be carried out at the higher concentration anyway. Specificity is the measure of the percentage of non-lethality at the higher concentration associated with the absence of the sign at the lower concentration. The individual signs focussed upon were those with high PPV and specificity, with appreciable sensitivity.

In the absence of any deaths at the lower concentration, toxicity occurred at the higher concentration in 77% of the studies (95% CI 72-82%), hence this value was used to set a threshold for use of a sign as an indicator of toxicity. Consequently, those signs with PPV's not only in excess of this value, but whose lower value of the 95% confidence limits of the PPV also exceeded 77% were selected.

2.2 Death as a predictor of toxicity at the next highest concentration

In the Sewell *et al.* (2015) dataset, death or euthanasia was found in the majority of studies at one or more concentrations. The PPV of a single death at the lower concentration was 93% (95% CI 84-98%) i.e. a single death is a strong predictor of lethality at the higher concentration. Although evident toxicity is the intended endpoint for the FCP method, and severe toxicity and death are to be avoided where possible, if death does occur this endpoint can therefore also be used to make decisions concerning classifications (Figure 1). But interestingly, since death is used as an objective endpoint for LC_{50} and ATC methods, it should also be noted that when two deaths occurred at the lower concentration this too was only 97% (95% CI 91-99%) predictive of lethality at the next higher concentration. That is to say, for a small number of the studies conducted, fewer deaths occurred at the higher concentration than at the lower. For the ATC method in particular, this could lead to an inaccurate classification.

2.3 Signs observed on day 0

Signs seen on the day of the test cannot unambiguously be ascribed to the chemical and may have resulted from handling, restraint or the inhalation procedure. Some signs such as wet coat and writhing were only observed on day 0, but some of the common and severe signs were seen both on day 0 and on subsequent days. For two such signs, irregular respiration and hypoactivity, the effect of discounting the day 0 observations increased the PPV and specificity (Sewell *et al.,* 2015) showing that signs that persist for more than 24h after exposure are better predictors of toxicity. However, as pointed out in this paper and in the new TG, severe signs seen on day 0 should be a signal to halt the study or possibly euthanize the animals so affected.

2.4 Signs of evident toxicity

In the case of one death at the lower concentration, a number of signs observed in the surviving animals increased the PPV of the single death (Sewell *et al.,* 2015). Some of these also had high sensitivity. Most importantly, a subset of these were also seen to be highly predictive in the absence of death at the lower level. The four signs in this subset were: hypoactivity, tremors, bodyweight loss (>10%), and irregular respiration (Table 3). The data showed that if any of these signs were observed in at least one animal from the day after exposure, animals were highly likely to die if exposed to the next higher concentration. Where any animals experienced tremors or hypoactivity this was 100% predictive of lethality at the next higher concentration. If any animal experienced body weight loss in excess of 10% of their pre-dosing weight, this was predictive of death at the higher concentration in 94% of cases. Similarly, body weight loss has previously been shown to be a reliable and frequent objective marker for the determination of the maximum tolerated dose (MTD) in short term toxicity tests in animals (Chapman *et al.,* 2013). Irregular respiration was also highly predictive, being indicative of lethality in 89% of cases.

These four signs were chosen to represent evident toxicity since they had lower 95% confidence interval limits in excess of the 77% threshold detailed above. However, there were other signs that were also highly predictive of lethality at the next higher concentration, albeit with wider confidence intervals often due to their infrequent occurrence in the dataset. For example, oral discharge occurred rarely (sensitivity 2.4%), but was 100% (95% confidence interval (CI) 54.9 -100%) predictive of lethality at the next highest concentration. Therefore the signs used to guide the decision of evident toxicity should not necessarily be restricted to the four signs named in Table 3. Information on the pred ictivity and sensitivity of each of the clinical signs observed in the dataset has been made available in Supplementary Data File 1. Information on subclasses of the dataset for dusts and mists, males and females is also available. This is intended to complement and add to study director judgement and experience so that a decision can be made on the recognition of evident toxicity in the absence of death or the four named signs.

The definition of 'evident toxicity' used for the purpose of the analysis was conservative when considering the accepted definition of evident toxicity in TG420, since it was based simply on the prediction of actual mortality or euthanasia at the higher concentration (in the absence of death at the lower), and did not also include 'severe distress or moribund status' at the higher concentration. However, this definition was chosen to reflect the different outcomes used for decision making in the protocol, so that 'evident toxicity' could be used to predict 'outcome A' (the death of 2 animals at the higher concentration), and therefore avoid the need for testing at that level (Figure 1). By using evident toxicity, classification can be made based on the *prediction* of death at the higher concentration. The method therefore has the potential to minimise the number of studies (i.e. concentrations tested) that will be required to make a classification and reduce the overall degree of suffering of animals in the study.

2.5 Severity and duration of signs

Severity of signs was not recorded consistently in the dataset, only whether a sign was present or not, and as the data had been generated in a number of different laboratories, the grading of severity may have had a strong subjective element. Therefore in the previous publication, only the severity of bodyweight loss was examined in more detail as it had been recorded as either unspecified, mild (reduced weight gain), moderate (10-20% compared with day 0) or substantial (>20% compared with day 0). In fact, PPV was largely unaffected by dividing body weight loss into these subcategories, but sensitivity declined because of the smaller numbers in each category.

Another way of looking at severity was to examine whether the sign was present in more than one animal. In the previous paper (Sewell *et al.,* 2015), it was shown that for irregular respiration (the sign for which there are the largest number of observations), the impact on PPV and specificity of increasing numbers of animals showing the sign was very small. However, because seeing the sign in a majority of animals was less common, the sensitivity declined accordingly.

2.6 Combinations and co-occurrence of signs (including signs in isolation)

Sewell *et al.,* (2015) considered whether combinations of signs would increase sensitivity, and thereby improve prediction of lethality at the higher concentration. However, the gains in sensitivity of all pairwise combinations were small because of the strong co-occurrence of signs, and inclusion of third or fourth signs had progressively less impact.

At the other extreme, we examined whether misclassification was likely if a sign was the only one reported (i.e. seen in isolation), and occurred only once and in only one animal. Irregular respiration and body staining were the most commonly observed signs in isolation (42% and 28% respectively of those animals that showed the sign) (Table 4). However, of the 268 pairs of studies[3] analysed, there were only 5 in which irregular respiration was recorded in the absence of other signs, and only once in only one animal. In each case, at least two animals died at the next higher concentration showing that the single sign was predictive (Table 5). Admittedly this is a small data set, but the finding supports the general robustness of the sign which is typically seen in more than one animal, and rarely occurs in isolation.

2.7 Varying concentration ratios

An odd feature of the GHS classification system is that the ratios of LC_{50} concentrations defined for each grade 1-5 are not of equal size but vary from 2 to 10. For example, for dusts and mists the concentrations tested are 0.05, 0.5, 1.0 and 5.0 mg/l (Table 2). Sewell *et al.* (2015) considered how this would affect classifications by the FCP method. It seemed possible that lethality at the higher concentration would be more likely if the concentration ratio was larger and that conversely, a smaller change in concentration might lead to a greater number of false positives i.e. lethality not seen at the higher concentration despite evident toxicity at the lower. This has now been looked at in two ways. Sewell *et al.* (2015) found that, for a small number of signs, the average concentration ratio for false positives was smaller than for true positives, in agreement with this idea. However, of the four signs selected as markers of evident toxicity, two were never associated with false positives (PPVs of 100%) and in the other two cases, the effect of concentration ratio did not reach statistical significance.

A further analysis was undertaken to look at the effect of the ratio of the higher to lower concentration on the PPV. In Table 6, PPVs are shown for a number of signs with >2 to <5, >5 to <10 or >10-fold ratios between the lower and higher concentrations. As anticipated, PPVs are higher for the larger concentration ratios, but since the majority of the studies used

[3] A pair of studies indicates a set of data from five animals, either all male or all female, exposed at two concentrations differing by at least a factor of two and in which no deaths occurred at the lower concentration.

the >2 to <5 fold ratio, the lower numbers in the remaining studies resulted in wider 95% confidence limits of the PPV values. The conclusion is that the main signs of evident toxicity were equally predictive regardless of the ratio of the higher to lower concentration.

3. Default sex and sighting studies

For the LC_{50} procedure, since males and females are treated identically and classifications are based on the sex that is most sensitive, sex differences generally do not have any impact on classification. For the ATC procedure, since males and females are not treated separately and the endpoints are based on the total number of deaths, irrespective of sex, differences in sensitivity have more of an impact and make the test less stringent. For example, where there is a 10-fold difference in sex sensitivity, simulations (Price *et al.,* 2011) showed that substances where the LC_{50} value of the most sensitive sex falls within GHS class 3 (the narrowest GHS classification band), these are almost always incorrectly classified as GHS class 4 (i.e. as less toxic). However, the guideline suggests that testing should be conducted in the more sensitive sex alone if a sex difference is indicated, which may mitigate this if sex differences are correctly identified in practice.

The original FCP method proposed the use of females as the default, as these were thought to be the more sensitive sex, and males only used if they were known to be more sensitive. In practice, significant differences in sensitivities between the sexes are fairly uncommon. Price *et al.,* (2011) showed a significant statistical difference between the LC_{50} values of males and females for 16 out of 56 substances examined (29%), females being the more sensitive in 11 of these. The dataset in Sewell *et al.* (2015) revealed little difference in sensitivity between the sexes. There was no difference in the prevalence of death or animals requiring euthanasia between the sexes, though some clinical signs were more prevalent in one sex than the other (ano-genital staining was more prevalent in females than males ($p = 0.0002$), whereas facial staining and gasping were marginally more common in males ($p = 0.028$ and 0.044 respectively). However, the predictivity of these signs did not differ between males and females, but the smaller numbers of studies in this analysis led to wider confidence intervals.

The statistical simulations carried out by Price *et al.* (2011) showed that where there was an unanticipated sex difference and testing was carried out in the less sensitive sex, this would usually result in misclassification, regardless of the method used. Consequently, the new test guideline proposes that a sighting study should be performed not only to determine a suitable starting concentration for the main study but to also identify whether there is a more sensitive sex. The sighting study is not recommended if there is existing information on which to base these two decisions. Despite the earlier proposal that females should be the default sex, the more recent data that failed to show any difference, and the general view of the OECD coordinators, and their nominated inhalation experts, that males were potentially more sensitive for inhaled substances, led to the proposal that males should be used in preference.

The new sighting study uses a single male and a single female at one or more of the fixed concentrations, depending on the outcome at each concentration as described by Stallard *et al*. (2011) (Figure 4). If there is no difference in sensitivity between the sexes, then the choice of sex for single sex studies for the FCP is irrelevant, and will not affect the classification. Since males are now the default sex, if they are the more sensitive, correct classification will still be made, since this is correctly based on the more sensitive sex. It is only if females are the more sensitive sex and this is *<u>not</u>* correctly identified, that there is potential for incorrect classification.

Though the risk of a sex difference is low, the new sighting study must be robust enough despite using only one male and one female to identify the large differences in sensitivities that might risk misclassification. To demonstrate this, we have carried out statistical calculations of the probability of choosing the most sensitive sex, with varying ratios of male and female sensitivity (i.e. LC_{50} values) (Figure 5). The methods are similar to those described by Stallard *et al.* (2011). Figure 5 shows the classification probabilities using the new sighting study for dusts and mists with a concentration-response curve slope of 4 and *R* (the ratio of the LC_{50} and $TC_{50,}$ the concentration expected to cause death or evident toxicity) of 5 for both sexes, assuming a sighting study starting at 0.05mg/L. The heavy solid line gives the probability of the correct classification given the LC_{50}. The heavy dashed line gives the probability that the main study is conducted in females rather than males.

The first plot of Figure 5 corresponds to the case of no difference between the sexes (i.e. males and females have identical LC_{50} values). In this case, the probability of the main study being carried out in females varies around 0.25, and since there is no difference in sensitivity this will not affect the classification. The other plots show what happens with increasingly large sex differences, with the females becoming more susceptible. In these cases the LC_{50} on the *x*-axis is that for the females, as this is the true value on which classification should be based (since females are more sensitive), and the dashed line gives the probability that the main study is conducted in the females. When the sex difference is small, there is quite a high chance of erroneously testing in the males when the females are marginally more sensitive. For example, for a LC_{50} ratio 1.5 the probability of incorrectly testing in the males is more than 0.5 in many cases. However, since the sex difference is small this is unlikely to impact the classification. As the sex difference increases, the chance of seeing the sex difference in the sighting study and doing the main test in the females correctly also increases. For a ratio of LC_{50} values of 10 or more the probability of choosing females for the main test exceeds 0.9 except for the least toxic substances, when no effects are seen in either sex even at the highest test concentration, or extremely toxic substances, when deaths are seen in both sexes at the lowest test concentration. The probability of misclassification is higher therefore for GHS classes 3 and 4.

These simulations show that the use of a single male and a single female in the sighting study should be sufficient to identify broad differences in sensitivities. Since the effect of sex differences is less when the concentration-response curve is steeper, these simulations represent a worst-case scenario when based on a slope of 4, as it is estimated that only 1% of substances have a concentration-response curve slope of less than this (Greiner, 2008). Again, it is important to note that sex differences are relatively uncommon and only unanticipated greater sensitivity in females is likely to influence classification. Furthermore, for many substances prior knowledge may be also available (e.g. from the oral route) which can be used to verify or indicate any suspected or apparent differences in sensitivity.

For the FCP method, the purpose of the sighting study is also to identify the starting concentration for the main study where existing information is insufficient to make an informed decision. A starting concentration should be chosen that is expected to cause evident toxicity in some animals, and the use of two animals, one male and one female, should be sufficient to determine whether this estimation is too high and allow a lower dose to be used in the main study, particularly if existing data are available. The ATC method does not include a sighting study and the choice of starting concentration is based on prior knowledge or experience, or use of the suggested default starting concentrations of 10 mg/L, 1 mg/L or 2500 ppm for vapours, dusts/mists and gases, respectively. This is also an option for the FCP method, since the sighting study is not compulsory. However, without the aid of a sighting study, it is possible that an inappropriate starting concentration may be chosen, which could result in testing at more concentrations and using more animals.

4. Comparability to existing methods and retrospective analyses

A number of publications have addressed the comparability of the three methods using statistical calculations or simulations to compare the classifications made by each of the three methods and the likelihood of misclassification (under or over) (Price *et al.* 2011; Stallard *et al.* 2011; Stallard *et al*. 2003). The calculations described above were based on hypothetical mortality concentration curves (with varying steepness) for a range of LC_{50} values covering all five toxic classes to represent a wide range of hypothetical substances. These include substances that clearly fall within a specific toxic class, (i.e. LC_{50} within the mid-range of the class bracket) as well as those on the class border (i.e. the most or least toxic substances in each class) where there is greater potential for misclassification. The simulations also took into account the potential for variation between the actual concentration tested and the intended fixed concentration. For the calculations, a variation of +/- 25% was used although this is greater than that permitted in the TG (+/- 20%) so these represent worst-case examples.

The statistical calculations showed that the three methods were comparable, although each of the methods did have the potential to misclassify even though the risk of this was low overall (Price *et al.,* 2011). If anything, the FCP tended to over-classify and the other two methods to under-classify. The impact of misclassification (over or under) and the choice of inhalation test method may raise some diversity of opinion depending on safety, commercial and 3Rs (Replacement, Refinement and Reduction) perspectives. The tendency of the LC_{50} and ATC methods to *under*-classify is more of a concern to human health than the FCP tendency for *over*-classification. However, it is worth highlighting that the statistical models that these conclusions were based on used a conservative ‘worst-case’ scenario, with a low concentration-response slope of four, and the potential to over-classify becomes less with a steeper concentration–response curve. Moreover, the models used a greater than permitted variation of the actual concentration from that intended.

The statistical calculations described above show that the three methods are comparable, particularly in the absence of sex differences, or where these have been taken into account with the use of the sighting study. However, all these methods rely on the assumption of correct identification or prediction of the LC_{50} value and the corresponding GHS class and are not based on real data. We have therefore undertaken further analysis of the data set of 178 dusts and mists to make retrospective classifications by all three methods and to compare their performance. For each method, the classifications were established using the protocols and flow charts in their corresponding test guidelines, based on the order the studies were carried out in practice (i.e. using the default or otherwise determined starting concentration). Supplementary Data File 2 contains information on the 'classification rules' for each method. For the LC_{50} method, rather than establish an LC_{50} value from the data, a flowchart method was used based on whether more or less than 50% animals died at each concentration (as in Figure 1 in Price *et al*. 2011). Only 'valid' concentrations corresponding to within ±20% of the four fixed concentrations for dusts and mists in the ATC and FCP protocols (0.05, 0.5, 1 and 5 mg/L) were included, to comply with the guidelines. Retrospective classifications could only be made for substances where all the necessary and valid concentrations were available. For example, in the FCP method, where testing started at 1mg/L and there was no death or evident toxicity in any animal, further testing would be required at 5mg/L. If this concentration had not been tested or fell outside of the ±20% criterion, then this substance could not be classified.

Retrospective classifications were made for 77 substances via the LC_{50} method, 57 substances via ATC, and 124 substances for FCP. For the FCP, classifications were generally able to be made using one or two concentrations requiring five to ten animals (Table 7). For the ATC and LC_{50} methods, classifications were generally made after two concentrations, requiring 12 animals and 20 animals respectively.

There were 42 substances for which a retrospective classification was made via all three methods (including based on females and males separately), and for 35 of these (83.3%) all classifications were in agreement (Table 8). If using the LC_{50} as the 'reference' method (though as described above there are limitations for this method and potential for misclassification), the ATC method under-classified by one class on three occasions. For the FCP method, when conducted in males only, there was one occasion of over-classification, and one of under-classification, both by one class. When the FCP was conducted in females only, there was also one occasion of over-classification, in the adjacent more stringent class, but three occasions of under-classification, one of these by two classes (class 4 *vs.* class 2). The reasons for these differences could be because the retrospective classification method was not able to take sex differences into account, or because the LC_{50} value falls near a class border where there is greater potential for misclassification. Table 9 shows that for 6 of these 7 substances there appears to be a more sensitive sex. If for the FCP, the classification is made according to the most sensitive sex, there are fewer disagreements with the classifications from the LC_{50} method. For example, instead there are now three occasions where classification made via FCP differs from LC_{50}, and these are all over-classifications into the adjacent more stringent class. Whereas the three occasions where the ATC method differed from the LC_{50} method were under-classifications into the less stringent adjacent class. This supports the conclusions from the statistical calculations that show the FCP is comparable to the existing methods if sex differences are taken into account.

Often it was not possible to make a retrospective classification using all three methods (e.g. due to a missing concentration), and there are more examples of the classifications made by two of the methods. Table 10 shows the agreement between any two of the methods. With the exception of the male and female comparisons, which had an agreement of 76.5% and 87.0% for the FCP and LC_{50} methods respectively, there was over 90% agreement with all combinations of the other methods. Supplementary Tables S1 -S7 compare the classifications made by each of these methods. The difference between the male and female comparisons may reflect differences in sensitivities between sexes and the fact that for the other comparisons the same animals will have been used to make the classification, which could not be done for the male and female comparisons. It is vital for the acceptance of the new TG that there is strong agreement between the classifications made by the FCP and the two accepted methods, irrespective of the sex used by the FCP.

However, as previously pointed out, a major difference between the three methods is the number of studies required to make a classification and consequently the numbers of animals used (Table 7).

5. Summary and conclusions

The new work described here strengthens and clarifies the conclusions of earlier publications on the FCP method. In particular we have shown that evident toxicity can reliably predict death or moribund status at the next highest fixed concentration irrespective of the fold-change in concentration or the number of animals showing the sign of evident toxicity, so demonstrating the robustness of the method.

As part of the OECD approval process, the simplicity of the definition of evident toxicity was questioned (i.e. that evident toxicity is said to have been reached if only one of the four signs is observed at least once in at least one animal). However, the dataset had been extensively interrogated to look at multiple scenarios, including the effect of combinations of signs, the duration of signs, and/or the number of animals displaying the sign(s) (see sections 2.5 and 2.6 and Sewell et al., 2015). Whilst predictivity did increase to some extent for some of these, these were associated with wider confidence intervals, since the pool of data also decreased. Clearly, if other data sets become available, it might be possible to confirm these trends more precisely. Therefore, increases in severity and/or the number of animals displaying the sign may increase confidence in the decision, but the statistical analysis of the dataset supports the simple definition regardless of any of such additional information.

The change of the default sex from female to male was an unexpected outcome from the consultation with the OECD national coordinators, but there was no evidence from the analysis of Sewell *et al.* (2015) for a consistent bias one way or the other. The decision therefore to adopt males as the default sex was based on the experience of the national coordinators and their nominated inhalation experts. However, since use of the less sensitive sex could result in misclassification, it was important to establish that the proposed sighting study with one male and one female would have the power to identify the more sensitive sex, at least under those circumstances where the difference in sensitivity was large enough that it might have led to wrong classification and in the absence of existing information on sex differences. The results of the statistical analysis confirms that a sighting study with one male and one female has the power to identify the more sensitive sex.

The retrospective analysis of the dataset to classify the chemicals by all three methods (LC_{50}, ATC and FCP) was especially important in gaining acceptance of TG 433 by OECD. Agreement between the three methods is very good as only 7 out of 42 substances showed any disagreement between the three methods and then by only one class if the most sensitive sex was selected for the FCP method. All three methods have the potential to misclassify so it is important that the advantages and limitations of each test method are understood so that users can select the most appropriate test method for their needs. However in the absence of any other considerations, the FCP method is to be preferred since it offers animal welfare benefits through the avoidance of death as an endpoint, and other 3Rs benefits through the use of fewer animals and fewer studies when compared to the ATC and LC_{50} methods. We hope that these factors will encourage wide uptake and use of the method in the future.

We attribute the reluctance to use the equivalent method for oral toxicity studies (TG 420) to lack of guidance on evident toxicity and the absence of the detailed analyses described here, that were needed to convince the OECD national coordinators that TG 433 was fit for purpose. A similar exercise is therefore planned in collaboration with the European Partnership for Alternatives to Animal Testing to examine clinical signs observed during acute oral toxicity studies and to provide guidance that will encourage the use of TG 420.

The experience of gaining acceptance of the FCP method for acute inhalation has been both positive and negative. The positive is the agreement to accept extensive retrospective analysis as sufficient justification for a new test guideline without the need for prospective validation studies which would have required further use of animals. This approach could no doubt be used on other occasions. The negative is the inordinately long time it has taken to get this method accepted even though the principle of evident toxicity had already been accepted by OECD, and the cumbersome process of consultation and submission which was required. Even now, the experience with the oral toxicity guideline TG 420 suggests that there will still be work needed to ensure that TG 433 becomes the preferred method for assessment of inhalation toxicity, and it is to be hoped that this will not take a further 13 years.

Acknowledgements

We would like to thank everyone who was involved in the Test Guideline Development Process, including the OECD secretariat, the OECD national co-ordinators and their nominated experts.

References

Chapman, K., F. Sewell, L. Allais, J.L. Delongeas, E. Donald, M. Festag, S. Kervyn, D. Ockert, V. Nogues, H. Palmer, M. Popovic, W. Roosen, A. Schoenmakers, K. Somers, C. Stark, P. Stei, and S. Robinson. 2013. A global pharmaceutical company initiative: an evidence-based approach to define the upper limit of body weight loss in short term toxicity studies. *Regulatory toxicology and pharmacology : RTP*. 67:27-38.

Greiner. 2008. *'Report on biostatistical performance assessment of draft TG436 acute toxic class method for acute inhalation toxicity,'* 2008.

OECD. 1987. Deleted Test Guidline 401: Acute Oral Toxicity. LD50 method. http://ntp.niehs.nih.gov/iccvam/docs/acutetox_docs/udpproc/udpfin01/append/appi.pdf.

OECD. 2001. Harmonized integrated hazard classification system for human health and environmental effects of chemical substances. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono%282001%296.

OECD. 2002a. Test Guideline 420: Acute Oral Toxicity - Fixed Dose Procedure. http://www.oecd-ilibrary.org/environment/test-no-420-acute-oral-toxicity-fixed-dose-procedure_9789264070943-en.

OECD. 2002b. Test guideline 423: Acute Oral toxicity - Acute Toxic Class Method. http://www.oecd-ilibrary.org/environment/test-no-423-acute-oral-toxicity-acute-toxic-class-method_9789264071001-en.

OECD. 2004. Draft proposal for a new guideline:433. http://www.oecd.org/chemicalsafety/testing/32035886.pdf.

OECD. 2008. Test Guideline 425: Acute Oral Toxicity: Up-and-Down Procedure. http://www.oecd-ilibrary.org/environment/test-no-425-acute-oral-toxicity-up-and-down-procedure_9789264071049-en.

OECD. 2009a. OECD Test Guideline 403: Acute Inhalation Toxicity. http://www.oecd-ilibrary.org/docserver/download/9740301e.pdf?expires=1433247636&id=id&accname=guest&checksum=7012504CE687B2E5614DB637989CE606.

OECD. 2009b. Test Guideline 436: Acute Inhalation Toxicity - Acute Toxic Class Method. http://www.oecd-ilibrary.org/docserver/download/9743601e.pdf?expires=1433247696&id=id&accname=guest&checksum=DF8991A8B9D88D13E75E914D1A2D5022.

Price, C., N. Stallard, S. Creton, I. Indans, R. Guest, D. Griffiths, and P. Edwards. 2011. A statistical evaluation of the effects of sex differences in assessment of acute inhalation toxicity. *Human & experimental toxicology*. 30:217-238.

Sewell, F., I. Ragan, T. Marczylo, B. Anderson, A. Braun, W. Casey, N. Dennison, D. Griffiths, R. Guest, T. Holmes, T. van Huygevoort, I. Indans, T. Kenny, H. Kojima, K. Lee, P. Prieto, P. Smith, J. Smedley, W.S. Stokes, G. Wnorowski, and G. Horgan. 2015. A global initiative to refine acute inhalation studies through the use of 'evident toxicity' as an endpoint: Towards adoption of the fixed concentration procedure. *Regulatory toxicology and pharmacology.* 73:770-779.

Stallard, N., C. Price, S. Creton, I. Indans, R. Guest, D. Griffiths, and P. Edwards. 2011. A new sighting study for the fixed concentration procedure to allow for sex differences. *Human & experimental toxicology*. 30:239-249.

Stallard, N., A. Whitehead, and I. Indans. 2003. Statistical evaluation of the fixed concentration procedure for acute inhalation toxicity assessment. *Human & experimental toxicology*. 22:575-585.

Figure Legends

Figure 1: LC_{50} test (OECD test guideline 403) for dusts and mists, using example concentrations, starting at 5 mg/L (Price et al., 2010). Please note the LC_{50} test method does not require fixed concentrations, but specifies that 10 animals (5 males and 5 females) should be exposed at three different concentration levels. The concentration levels should be sufficiently spaced to enable construction of a mortality curve so that an estimation of the LC_{50} can be obtained.

Figure 2: Acute toxic class (ATC) method for dusts and mists for an example starting concentration of 5 mg/L (Price et al., 2010). Please note, the ATC method specifies that 6 animals (3 males and 3 females) are tested at fixed concentrations that form the upper limit of the GHS categories. The starting concentration is either the highest concentration, or that which is expected to lead to mortality in some of the exposed animals, based on prior information.

Figure 3: Fixed concentration procedure (FCP) method for dusts and mists for an example starting concentration of 5 mg/L (Price et al., 2010). Please note, the draft test guideline specifies that substances are tested at fixed concentrations that form the upper limit of the GHS categories. The starting concentration is chosen to be the fixed concentration level that is most likely to lead to evident toxicity but not death.

Figure 4: FCP sighting study for dusts and mists.

Figure 5: Classification probabilities for the fixed concentration procedure (FCP) with the new sighting study for dusts and mists with concentration-response curve slope of 4 and R (LC50/TC50) of 5 assuming sighting study starting at 0.05 mg/L. The different plots show varying sex differences, to assess the impact of increased female sensitivity compared to male (i.e. female LC_{50} increasingly lower than male LC_{50}). The vertical dotted line in each plot indicates the classification boundary concentrations and the light solid line indicates the cumulative probabilities of classification (on left-hand axis scale) into each toxic class for LC_{50} values shown. The heavy solid line gives the probability of the correct classification given the LC_{50}. The heavy dashed line gives the probability that the main study is conducted in females rather than males. For more information on these plots please refer to Stallard *et al*. (2011).

Supplementary data

Supplementary Data File 1: Information on the predictivity and sensitivity of each of the clinical signs observed in the dataset.

Supplementary Data File 2: ‘Classification rules’ for each method.

Table 1: Comparison of LC_{50}, ATC and FCP methods.

Parameter	**LC_{50} (concentration causing 50% lethality)**	**ATC (acute toxic class)**	**FCP (fixed concentration procedure)**
OECD Test Guideline	403	436	433
Endpoint	Death	Death	Evident toxicity
Sighting study	No sighting study required.	No sighting study required.	A sighting study may be carried out to help inform the starting concentration and choice of sex, if deemed necessary. This is not compulsory. 1M+1F at one to four concentrations (usually only one or two concentrations required). The starting concentration should be that which is most expected to produce evident toxicity in some animals. If no prior information is available this should be 10 mg/L, 1 mg/L or 2500 ppm for vapours, dusts/mists and gases, respectively.
Number of animals	5M+5F per study. Usually three studies required. Min 10 – max 40 animals.	3M+3F per study. Usually at least two studies required (12 animals), though classification can sometimes be made based on one study, if testing at the lowest or highest concentrations (depending on the outcome). Numbers of animals range from 6 to max 24 (depending on the number of studies). An inappropriate starting concentration (causing too much or too little toxicity) may require testing at additional concentrations and may therefore result in higher numbers of animals being used. Where a marked sex difference is observed additional	Single (most sensitive) sex, or males only as default. 5 animals per study. Classification can often be made after a single study (5 animals). Numbers of animals range from 5 to max 20 (depending on the number of studies). Plus 2-8 in the sighting study (though the use of 8 animals in the sighting study would be very unusual, and only if the highest or lowest concentrations were chosen inappropriately as the starting concentration). An inappropriate starting concentration (causing too much or too little toxicity) may require testing at additional concentrations and may therefore result in

		animals may be required.	higher numbers of animals being used. However, a sighting study should avoid this.
Number of concentrations	At least three concentrations (to enable production of a concentration-mortality curve and estimation of LC_{50}).	An 'up and down method' is used, requiring 1 to 4 fixed concentrations (based on the upper limit of the GHS classification system) depending on the outcome at each concentration. Generally at least two concentrations are required to make a classification. Sometimes a classification can be made based on only one study if starting at the highest or lowest fixed concentration, and depending on the outcome.	An 'up and down method' is used, requiring 1 to 4 fixed concentrations (based on the upper limit of the GHS classification system) depending on the outcome at each concentration. A classification can often be made based on one study only.
Starting concentration	n/a This is not a sequential method. At least three concentrations are required to enable production of a concentration-mortality curve and estimation of $LC_{50.}$	Starting concentration level should be that which is most likely to produce toxicity in some animals. If no prior information is available the starting concentration will be 10 mg/L, 1 mg/L or 2500 ppm for vapours, dusts/mists and gases, respectively. An inappropriate starting concentration (causing too much or too little toxicity) may require testing at more concentrations than if a more appropriate concentration had been chosen.	Starting concentration level should be that which is most expected to produce evident toxicity in some animals. The sighting study may inform this choice, or prior information if available. If a sighting study has not been conducted or is inconclusive, or if no prior information is available the starting concentration will be 10 mg/L, 1 mg/L or 2500 ppm for vapours, dusts/mists and gases, respectively. An inappropriate starting concentration (causing too much or too little toxicity) may require testing at more concentrations than if a more appropriate concentration had been chosen. The use of a sighting study should avoid this.
Classification Method	Based on a point estimate of LC_{50} which allows classification according to the GHS classification system.	Based on an interval estimate of LC_{50}, so that classification is based on the toxic class that the estimated LC_{50} falls within, using the GHS classification system.	LC_{50} is inferred through the use of evident toxicity to *predict* death at a higher dose, and classification made according to the inferred LC_{50} using the GHS classification system.

Table 2: GHS classification system for inhalation. For the LC_{50} method, a point estimate of the LC_{50} allows classification into the relevant GHS class according to the table. The ATC method estimates which class the LC_{50} falls within and makes classification on that basis, whereas classifications made by FCP are based on the *inferred* LC_{50}.

GHS category	Vapours (mg/L)	Dusts and mist (mg/L)	Gases (ppm)
1 (most toxic)	≤0.5	≤0.05	≤100
2	>0.5 and ≤2	>0.05 and ≤0.5	>100 and ≤500
3	>2 and ≤10	>0.5 and ≤1	>500 and ≤2,500
4	>10 and ≤20	>1 and ≤5	>2,500 and ≤20,000
5	>20	>5	>20,000

GHS, Globally Harmonised System; LC50, median concentration; ppm, parts per million.

Table 3: Clinical signs indicating evident toxicity (PPV, sensitivity and specificity)

Clinical signs	PPV (95% CI)		Sensitivity (95% CI)		Specificity (95% CI)	
Hypoactivity	100.0	(92.4 - 100.0)	18.4	(13.6 - 24.1)	100.0	(95.2 - 100.0)
Tremors	100.0	(68.8 - 100.0)	3.90	(1.90 - 7.20)	100.0	(95.2 - 100.0)
Bodyweight loss	94.0	(84.6 - 98.4)	22.7	(17.4 - 28.8)	95.1	(87.2 - 98.7)
Irregular respiration	89.0	(80.9 - 94.5)	35.3	(29.0 - 42.0)	85.2	(74.7 - 92.5)

CI, Confidence Interval; PPV, positive predictive value.

Table 4: Number of animals displaying a clinical sign in isolation, and the total number of animals displaying the sign.

Clinical sign	**No. animals displaying sign ONLY (%)**		**Total no. animals displaying the sign**
Irregular respiration	137	(42%)	325
Body staining	27	(27%)	99
Hypoactivity	12	(16%)	77
Laboured respiration	12	(16%)	77
Faeces reduced	13	(12%)	107
Hunched posture	18	(8%)	227
Ano-genital staining	4	(8%)	51
Naso-ocular discharge	6	(7%)	89
Congested respiration	4	(5%)	87
Facial staining	3	(5%)	65
>10% bodyweight loss	2	(2%)	93
Noisy respiration	1	(0.4%)	267

Table 5: Studies where irregular respiration was observed only once in one animal at the lower concentration in females, with no other signs.

Study	Concentration tested	Female observations		Male observations	
		Number of Deaths	Number with evident toxicity	Number of Deaths	Number with evident toxicity
1	0.05 mg/L	0	1	0	4
	0.5 mg/L	5	-	3	2
	2 mg/L	5	-	5	0
2	0.06 mg/L	0	1	0	5
	0.5 mg/L	2	3	3	2
	2 mg/L	4	1	5	-
3	0.5 mg/L	0	1	0	4
	2 mg/L	2	3	2	3
4	0.05 mg/L	0	1	0	2
	0.2 mg/L	5	-	5	-
	2 mg/L	5	-	5	-
	5 mg/L	5	-	5	-
5	0.06 mg/L	0	1	n/a	n/a
	0.5 mg/L	2	3	0	5
	2 mg/L	5	-	5	0

Table 6: PPV (95% confidence interval) for highly predictive signs with 2, 5 or 10-fold concentration change between the lower and higher concentration.

Clinical sign	≥2-fold	(95% CI)	≥5-fold	(95% CI)	≥10-fold	(95% CI)
Tremors	100.0	(68.8 - 100.0)	100.0	(5.0 - 100.0)	100.0	(5.0 - 100.0)
Hypoactivity	100.0	(92.0 - 100.0)	100.0	(47.3 - 100.0)	100.0	(47.3 - 100.0)
>10% bodyweight loss	91.7	(79.0 - 97.8)	85.7	(47.0 - 99.3)	100.0	(36.8 - 100.0)
Irregular respiration	89.0	(80.9 - 94.5)	95.8	(81.2 - 99.8)	100.0	(86.1 - 100.0)
Body staining	88.5	(71.8 - 97.0)	100.0	(60.7 - 100.0)	100.0	(22.4 - 100.0)
Ano-genital staining	86.4	(67.3 - 96.4)	0.0	(0.0 - 95.0)	100.0	(5.0 - 100.0)
Faeces reduced	85.3	(70.4 - 94.4)	100.0	(47.3 - 100.0)	100.0	(47.3 - 100.0)
Naso-ocular discharge	84.2	(70.1 - 93.3)	100.0	(74.1 - 100.0)	100.0	(65.2 - 100.0)
Noisy respiration	80.5	(70.9 - 88.0)	94.1	(74.3 - 99.7)	100.0	(68.8 - 100.0)
Hunched posture	78.0	(65.0 – 87.8)	87.5	(64.5 - 97.8)	100.0	(54.9 - 100.0)
Gasping	76.5	(52.5 – 92.0)	100.0	(22.4 - 100.0)	100.0	(22.4 - 100.0)

Table 7: Number of studies required to make a classification, and the associated number of animals.

No. studies to make a classification	FCP			ATC		LC$_{50}$	
	No. animals involved	No. studies		No. animals involved	No. studies	No. animals involved	No. Studies
		FCP-F	FCP-M				
1 study	5	54	64	6	18	10	32
2 studies	10	46	41	12	37	20	41
3 studies	15	1	3	18	2	30	3
4 studies	20	0	1	24	0	40	1

Table 8: Classifications made by all three methods, showing the number of substances classified into each class and the number of substances where there was a disagreement between the three methods (which is expanded on in Table 9).

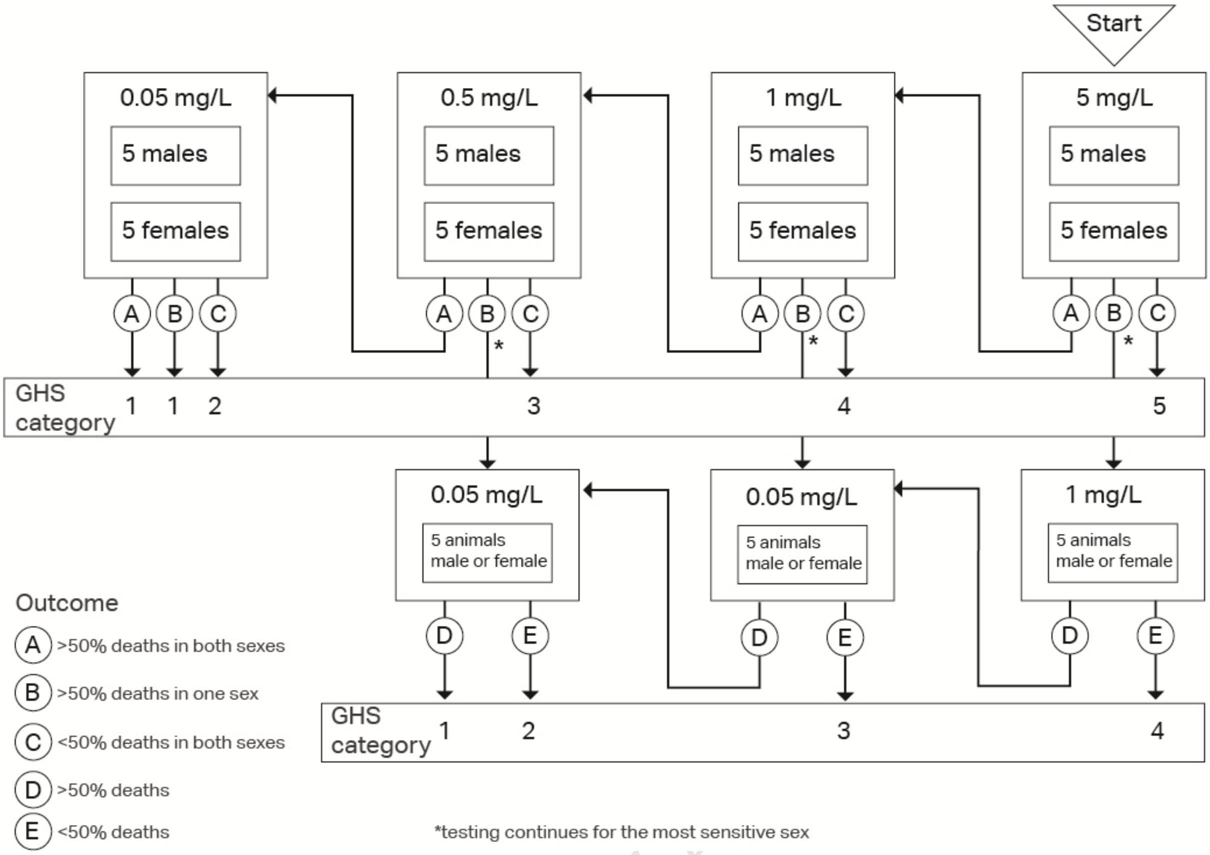
Classification	**No. substances**
Class 1	1
Class 2	11
Class 3	3
Class 4	14
Class 5	6
Disagreements	7

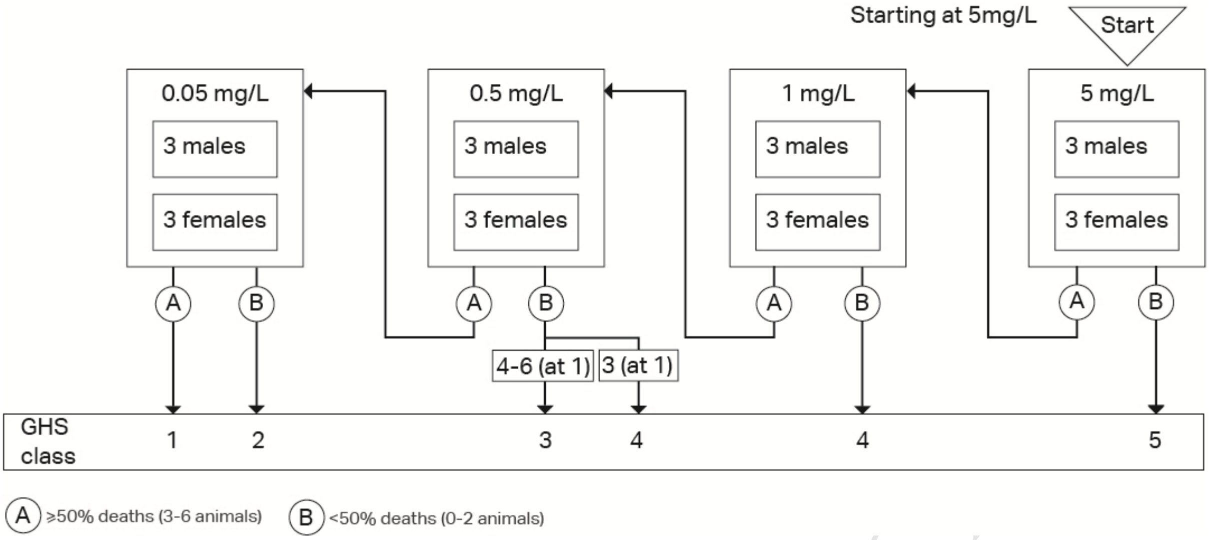
Table 9: Substances where there were differences in retrospective classifications made via the LC_{50}, ATC and FCP methods. FCP retrospective classifications were made for both females (F) and males (M) only. For each substance the concentrations tested, the number of deaths and/or animals with evident toxicity are indicated.

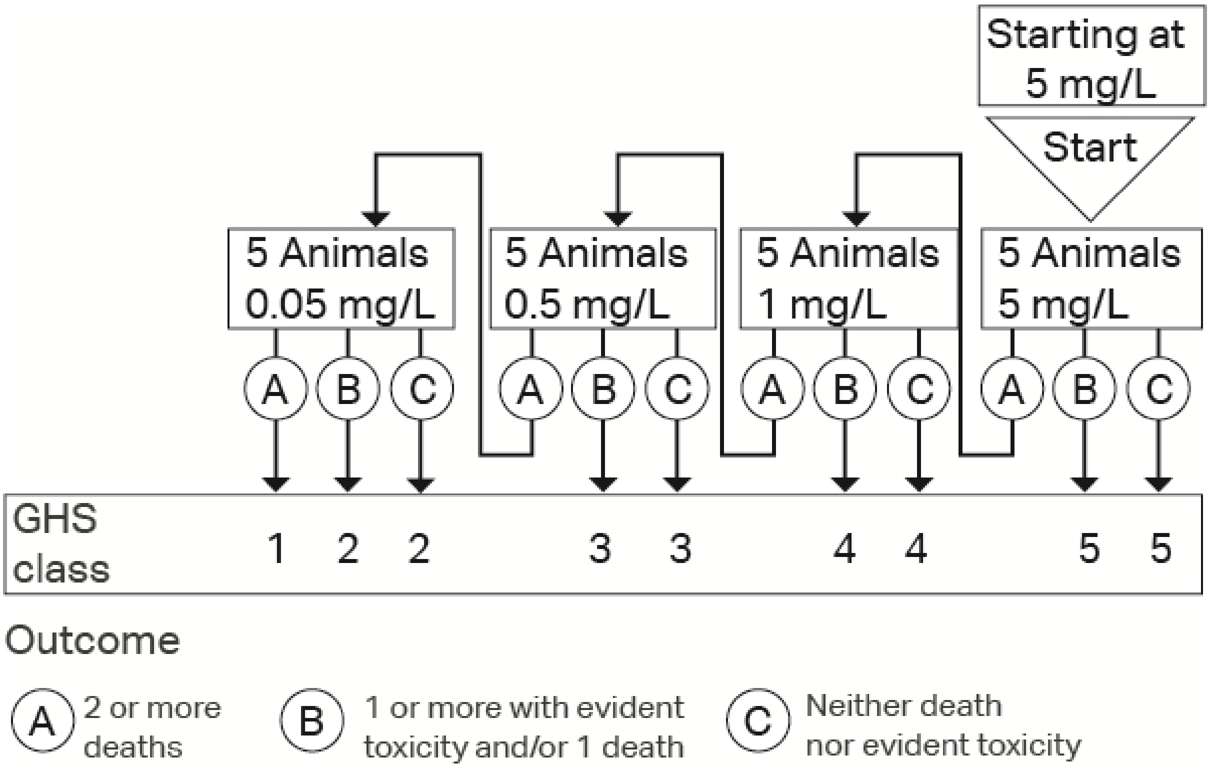
Substance	Concentrations tested	No. deaths		No. evident toxicity		Classification					
		F	M	F	M	LC_{50}	ATC	FCP(F)	FCP(M)		
1	0.5mg/L	0	0	0	0	3	4	3	4		
	1 mg/L	4	1	1	0						
2	1 mg/L	0	0	4	4	5	5	4	5		
	5 mg/L	2	1	3	4						
3	1 mg/L – males	-	0	-	0	5	5	5	4		
	5 mg/L	0	2	5	3						
4	1 mg/L – males	-	0	-	5	4	5	5	4		
	5 mg/L	0	3	5	2						
5	0.05 mg/L	0	0	0	0	2	2	4	2		
	0.5 mg/L	3	4	0	0						
	1 mg/L	1	4	2	0						
6	1 mg/L	0	0	0	0	5	5	4	5		
	5 mg/L	2	0	3	5						
7	0.5 mg/L	0	0	0	0	3	4	4	3		
	1 mg/L	1	3	4	2						
	5 mg/L	5	5	-	-						

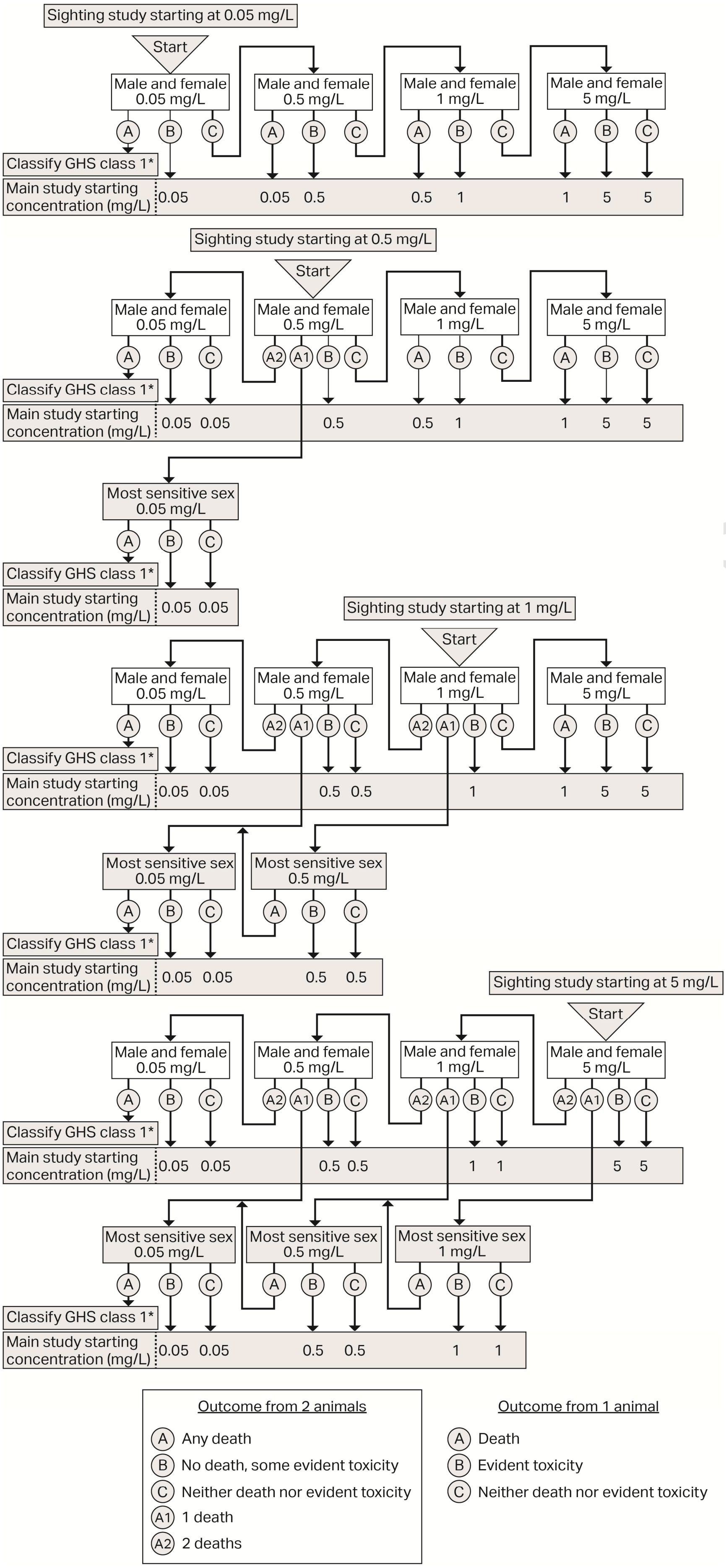
Table 10: Differences in classifications between the three methods, showing the numbers of substances for which pairwise comparisons were made, and the number for which there was agreement between the two methods.

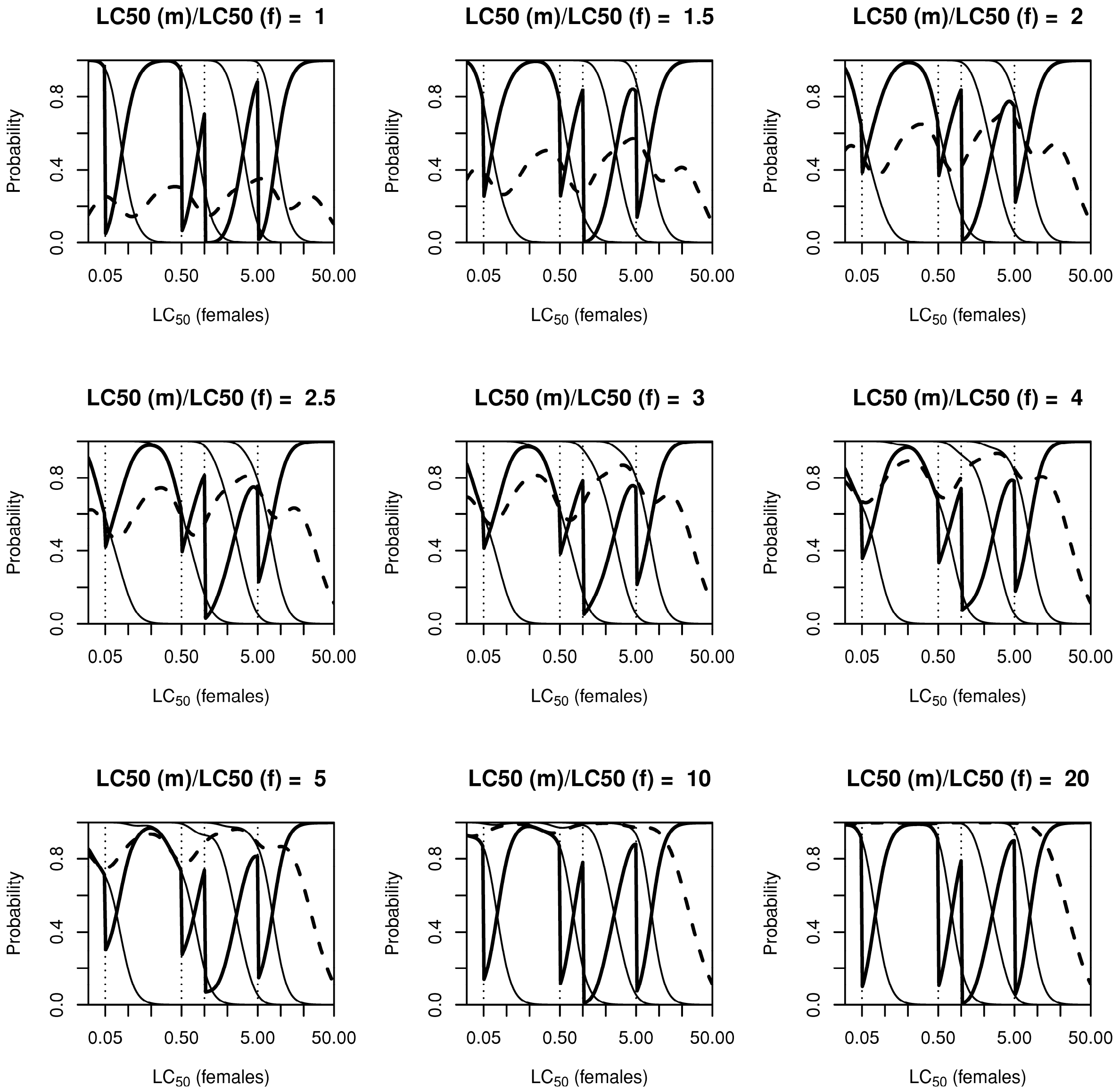
Comparison			No. classified	No. substances in agreement	% agreement
FCP-M	*vs.*	FCP-F	85	65	76.5%
LC_{50}-M	*vs.*	LC_{50}-F	46	40	87.0%
ATC	*vs.*	FCP-F	46	42	91.3%
LC_{50}	*vs.*	FCP-F	43	40	93.0%
LC_{50}	*vs.*	FCP-M	44	41	93.2%
ATC	*vs.*	FCP-M	51	48	94.1%
LC_{50}	*vs.*	ATC	46	44	95.7%


Start
0.05 mg/L
5 males
5 females
0.5 mg/L
5 males
5 females
1 mg/L
5 males
5 females
5 mg/L
5 males
5 females
A
B
C
*
GHS category
1
1
2
3
4
5
0.05 mg/L
5 animals male or female
0.05 mg/L
5 animals male or female
1 mg/L
5 animals male or female
D
E
GHS category
1
2
3
4
Outcome
A >50% deaths in both sexes
B >50% deaths in one sex
C <50% deaths in both sexes
D >50% deaths
E <50% deaths
*testing continues for the most sensitive sex


Starting at 5mg/L
Start
0.05 mg/L
0.5 mg/L
1 mg/L
5 mg/L
3 males
3 females
A
B
4-6 (at 1)
3 (at 1)
GHS class
1
2
3
4
4
5
A ≥50% deaths (3-6 animals)
B <50% deaths (0-2 animals)


Starting at 5 mg/L
Start
5 Animals 0.05 mg/L
5 Animals 0.5 mg/L
5 Animals 1 mg/L
5 Animals 5 mg/L
A
B
C
GHS class
1 2 2
3 3
4 4
5 5
Outcome
A 2 or more deaths
B 1 or more with evident toxicity and/or 1 death
C Neither death nor evident toxicity


Sighting study starting at 0.05 mg/L
Start
Male and female 0.05 mg/L
Male and female 0.5 mg/L
Male and female 1 mg/L
Male and female 5 mg/L
A B C
Classify GHS class 1*
Main study starting concentration (mg/L)
0.05 0.05 0.5 0.5 1 1 5 5
Sighting study starting at 0.5 mg/L
Start
Male and female 0.05 mg/L
Male and female 0.5 mg/L
Male and female 1 mg/L
Male and female 5 mg/L
A B C A2 A1 B C A B C A B C
Classify GHS class 1*
Main study starting concentration (mg/L)
0.05 0.05 0.5 0.5 1 1 5 5
Most sensitive sex 0.05 mg/L
A B C
Classify GHS class 1*
Main study starting concentration (mg/L)
0.05 0.05
Sighting study starting at 1 mg/L
Start
Male and female 0.05 mg/L
Male and female 0.5 mg/L
Male and female 1 mg/L
Male and female 5 mg/L
A B C A2 A1 B C A2 A1 B C A B C
Classify GHS class 1*
Main study starting concentration (mg/L)
0.05 0.05 0.5 0.5 1 1 5 5
Most sensitive sex 0.05 mg/L
Most sensitive sex 0.5 mg/L
A B C A B C
Classify GHS class 1*
Main study starting concentration (mg/L)
0.05 0.05 0.5 0.5
Sighting study starting at 5 mg/L
Start
Male and female 0.05 mg/L
Male and female 0.5 mg/L
Male and female 1 mg/L
Male and female 5 mg/L
A B C A2 A1 B C A2 A1 B C A2 A1 B C
Classify GHS class 1*
Main study starting concentration (mg/L)
0.05 0.05 0.5 0.5 1 1 5 5
Most sensitive sex 0.05 mg/L
Most sensitive sex 0.5 mg/L
Most sensitive sex 1 mg/L
A B C A B C A B C
Classify GHS class 1*
Main study starting concentration (mg/L)
0.05 0.05 0.5 0.5 1 1
Outcome from 2 animals
A Any death
B No death, some evident toxicity
C Neither death nor evident toxicity
A1 1 death
A2 2 deaths
Outcome from 1 animal
A Death
B Evident toxicity
C Neither death nor evident toxicity


LC50 (m)/LC50 (f) = 1
LC50 (m)/LC50 (f) = 1.5
LC50 (m)/LC50 (f) = 2
LC50 (m)/LC50 (f) = 2.5
LC50 (m)/LC50 (f) = 3
LC50 (m)/LC50 (f) = 4
LC50 (m)/LC50 (f) = 5
LC50 (m)/LC50 (f) = 10
LC50 (m)/LC50 (f) = 20
Probability
0.0
0.4
0.8
0.05
0.50
5.00
50.00
LC_{50} (females)

Highlights:

The FCP for acute inhalation toxicity has been accepted by OECD as TG433.

TG433 uses evident toxicity while other approved methods use lethality.

A sighting study with 1 M and 1 F animal reliably identifies the more sensitive sex.

The three methods (LC_{50}, ATC, FCP) showed good agreement in a retrospective analysis.