

Oil Spill Treatment Product Testing Scheme – Research and development activities from the 2007 scheme review

Authors:

Mark Kirby, Clare Williams, Jon Barry

Issue date: July 2010.

**Cefas Lowestoft Laboratory
Pakefield Road
Lowestoft
Suffolk NR33 0HT**

Cefas Project ME1309

Oil Spill Treatment Product Testing Scheme – Research and development activities from the 2007 scheme review

**Sponsor: Department for Environment, Food
& Rural Affairs, London.**

Table of contents

| | | |
|----------|---|-----------|
| 1 | Summary | 4 |
| 2 | Introduction | 6 |
| | 2.1 The UK scheme and the 2007 review | 6 |
| 3 | Materials, Methods and Results | 8 |
| | 3.1 Testing approach | 8 |
| | 3.2 Test organisms | 8 |
| | 3.3 Heavy Oil Test | 8 |
| | 3.3.1 Heavy oil test development – Test oil selection | 8 |
| | 3.3.2 Heavy oil test method | 9 |
| | 3.3.3 Heavy oil test – Results | 10 |
| | 3.4 Type 3 test assessment | 11 |
| | 3.4.1 Type 3 test development – dispersant selection | 11 |
| | 3.4.2 Type 3 dispersant test method | 12 |
| | 3.4.3 Type 3 tests – Results | 12 |
| 4 | Statistical Analysis | 14 |
| | 4.1 Statistical Test of the Difference in Mortality Rates of the Control and Test Treatments | 14 |
| | 4.2 Experimental Design | 14 |
| | 4.3 Data screening and test of between-tank homogeneity | 14 |
| | 4.4 Comparison of control and test treatment mortality rates | 15 |
| | 4.5 Decision procedure for approval applications for type 2 dispersant tests (current approach) | 15 |
| | 4.6 Choosing procedures for type 3 tests | 15 |
| | 4.7 Power of the D statistical significance test | 15 |
| | 4.8 A decision procedure for heavy oil and type 3 tests | 17 |
| | 4.9 Results for Falmouth oil (the selected heavy oil) | 17 |
| | 4.10 Results for Type 3 tests using Kuwait oil as control | 17 |
| 5 | Discussion & Conclusions | 24 |
| | 5.1 Heavy oil test development | 24 |
| | 5.2 Type 2 and Type 3 testing | 25 |
| | 5.3 Implementation into statutory scheme | 25 |
| | 5.3.1 Setting of pass/fail criteria | 25 |
| | 5.3.2 Need for a Rocky Shore test | 26 |
| | 5.3.3 Putting the new scheme to the industry | 26 |
| | 5.4 The Future | 26 |
| 6 | References | 28 |
| 7 | Acknowledgements | 28 |
| 8 | Appendix 1: Overview of statistical toxicity test | 29 |

1 Summary

Oil spill dispersants are an important option in the toolbox of the UK marine spill response industry. In fact they are, especially for offshore oil spills, the primary response tool used by the UK government response organisation. If used correctly they can help minimise the damage of spills to important marine environment resources. However, it is recognised that the indiscriminate and incorrect use of dispersants also has the ability to have detrimental effects. Therefore, the UK government, through the Marine Management Organisation (and formerly through the Marine Fisheries Agency), regulates the use and approval of oil spill treatment products through a statutory scheme. Under this scheme products can only gain approval for use in UK marine waters after undergoing assessments for both efficacy and toxicity.

In order that the approval scheme is fit for purpose and continues to enable products to be tested and approved appropriately a periodic review including full stakeholder engagement is undertaken. The Marine and Fisheries Agency undertook the last comprehensive scheme review in 2007 and a range of recommendations and research suggestions were reported as a result. However, the two primary recommendations that would require research to implement were:

- i) In recognition of the increased likelihood of fuel oil or heavy oil spills it was recommended that a new test was developed to enable the toxicity testing of products designed for use on these types of oil; and
- ii) In recognition of data that provide strong evidence that dispersant products elicit higher toxicity in the Sea Test when added as a Type 3 (neat) as opposed to a Type 2 (water diluted) it was recommended that the practice to approve products as type 2/3 on the basis of a type 2 test only should cease. Further research will be required to establish the type 2 vs type 3 toxicity differences and to establish new pass/fail criteria for each test type.

The ME1309 project was commissioned to address these issues and has developed modified procedures for new test protocols and generated the necessary data which can be used to set appropriate pass fail criteria for the new product categories.

The UK already has a well established toxicity testing procedure for oil spill dispersants (and other treatment products) and the primary test, known as the Sea Test, involves the exposure of a test organism (brown shrimp) under controlled mixing conditions to mechanically dispersed oil and the same oil treated with the product under assessment. Pass/fail criteria are based on the comparative mortalities between the control (oil only) and test (oil + product) treatments. Under this project the Sea Test protocol was taken as the starting point and modifications made to enable the assessment for treatment with a representative heavy oil.

A wide range of candidate oils were used in the test system to establish an oil with the correct levels of dispersibility, toxicity and reproducibility. A test was developed using a Falmouth sourced IFO180 fuel oil as the standard test oil

A further phase of the project investigated the toxicity performance of oil spill dispersants in the standard Sea Test when added as a Type 2 (water diluted) or a Type 3 (neat). A dataset was generated by the repetitive running of the Sea Test with Type 3 additions of a wide range of dispersants and it was confirmed that toxicity was higher when a dispersant was added as a Type 3.

The generated datasets for the new heavy oil test protocol and for the standard Sea Test with addition as a Type 3 were statistically analysed in order to establish what options for pass/fail criteria of the new product assessment procedures (to be based on an allowable percentage increase in mortality in the Sea Test for the treatment compared to the control).

The project has successfully developed test methods appropriate for the assessment of oil spill treatment products for use on heavy oils and also for the assessment of dispersants as Type 3's. It has also generated sufficient datasets that will enable appropriate pass/fail criteria to be set. Issues surrounding the implementation of the new procedures into the statutory schemes, including issues such as pass/fail thresholds, timescales and industry liaison, are also discussed in the report. However, the exact implantation strategy will be decided by the regulator, the Marine Management Organisation, with reference to this research.

2 Introduction

The UK is an island nation with a long coastline and as such the UK government recognises the importance of shipping to the economic health and international standing of the country. The levels of marine traffic in UK waters is high and with this comes the inevitable risks associated with potential oil and chemical spills as a result of shipping accidents and the consequent threat to marine resources (biological, physical and amenity). These threats have been realised on many occasions in the past during incidents such as those involving the Sea Empress and the MSC Napoli.

When these inevitable incidents do occur the UK has a responsibility to ensure that response procedures are in place to enable prompt action to be taken to mitigate the likely environmental impacts of, for example, spilled oil or chemicals.

Primary amongst the options available in the event of oil spills are the use of a range of oil spill treatment products (e.g. chemical dispersants, sorbents, surface cleaners etc.). If used appropriately these products offer vital tools to the responders in their efforts to minimise impact and aid clean-up after spills. However, the deliberate addition of chemical products into the marine environment requires regulation under the Food and Environment Protection Act (FEPA 1985) and, therefore, their use in spill response requires products to be 'approved'. In order to ensure that only the most efficient and environmentally acceptable products are used in UK waters the government oversee an oil spill treatment product testing and approval scheme designed to maintain an 'approved list' of products for use in UK waters (and also adopted by many other nations).

2.1 The UK Scheme and the 2007 Review

The UK oil spill product approval scheme has been established for over 30 years and is recognised as one of the most comprehensive and effective schemes of its type. It is designed to enable the assessment of a product on the basis of efficiency (dispersants only) and toxicity (all products). The standard toxicity assessment procedure comprises two tests i) the Sea Test and ii) the Rocky Shore Test.

In order to ensure that the testing and approval process and techniques remain 'fit for purpose' in the light of a changing shipping and response industry the UK government conducts periodic public consultation and review of the process. The most recent of these reviews was conducted and completed in 2007. Its remit was to canvass expert opinion from a broad spectrum of stakeholders on a wide range of testing and approval issues as raised by stakeholders and Defra/MFA's scientific advisors, Cefas.

The review conducted in 2007 made a number of recommendations which can be viewed via the MFA website. However, the two primary recommendations that would require research to implement were:

i) In recognition of the increased likelihood of fuel oil or heavy oil spills in respect of changing traffic through UK waters it was recommended that a new test was developed to enable the toxicity testing of products designed for use on these types of oil; and

ii) In recognition of data that provide strong evidence that dispersant products elicit higher toxicity in the Sea Test when added as a Type 3 (neat) as opposed to a Type 2 (water diluted) it was recommended that the practice to approve products as type 2/3 on the basis of a type 2 test only should cease. Further research will be required to establish the type 2 vs type 3 toxicity differences and to establish new pass/fail criteria for each test type.

In light of these recommendations this research was commissioned to:

i) Develop and validate a new test procedure (based on the Sea Test premise) for the approval of products for use on Heavy Oils.

ii) Investigate the relationship of toxicity in the Sea Test whether dispersants are added as type 2 or 3. Establish new separate pass/fail assessment criteria for dispersant approval for these two usage types.

3 Materials, Methods and Results

3.1 Testing Approach

The starting position in terms of methodology for the testing procedure for both the heavy oil test and the Type 3 test was the current standard Sea Test protocol. This protocol (as set out in Kirby et al. 1996) remains the core test of the current approval scheme and offered a proven approach from which the necessary amendments could be made.

In principle, this meant a change of standard test oil for the heavy fuel oil test and a change to the method of dispersant addition for the Type 3 test. The appropriate pass/fail criteria for these new, but consistent, approaches is then decided through the adoption of an appropriate statistical approach dependent on the nature of the results generated under the new tests.

3.2 Test Organisms

For both the heavy oil tests and the Type 3 dispersant tests Brown shrimp (*Crangon crangon*) were obtained from the River Crouch in Burnham-on-Crouch, Essex, UK. A sizing guide was used to ensure all the shrimp were between 5 and 7cm in length and 20 were added to each tank randomly.

3.3 Heavy Oil Test

3.3.1 Heavy Oil Test Development - Test Oil Selection

The initial phase of the project involved the selection of a suitable test oil to represent a typical 'heavy oil'. Candidate oils were tested to see how they dispersed in the standard sea test (without dispersant) and what level of toxicity this produced. The criteria for candidate oil selection were:

1. The oil must be considered to be a representative heavy crude or fuel oil
2. The oil should elicit sufficient toxicity to the brown shrimp, *C. crangon*, when physically dispersed under standard test conditions
3. Mortality must reproducibly fall within an acceptable range
4. The oil must be available in the required quantities from the supplier
5. The oil should be relevant to UK policy (i.e. is transported or used as bunker fuel in UK waters)

Initially 2 Heavy fuel oil samples (HFO 380) were provided by OSR/EARL and 2 Intermediate fuel oil samples (IFO 180) were provided by Conoco Phillips. The Heavy fuel oils proved very difficult to disperse and were not toxic enough under normal test conditions. The first intermediate fuel oil from Conoco Phillips "Immingham IFO 180 1" was too toxic but the second sample 'Immingham IFO 180 2' seemed to provide an acceptable level of dispersability and toxicity for test purposes but, unfortunately, a sufficiently large supply of the oil was not available.

Two more IFO 180 oils were sourced from Inver Energy in Cardiff and WFS in Falmouth. These oils were both classed as IFO180 and were obtained in 40L quantities.

A summary of the candidate oils considered for testing can be found in Table 1.

Table 1. Summary of oils considered as candidate ‘heavy oils’ for the new toxicity test procedure

| Oil | ID number | Date | Supplier | Notes | kinematic viscosity (cSt) |
|------------------|-----------|------------|------------------|---|---------------------------|
| Immingham IFO 1 | C3159-01 | 21/02/2009 | Conocco Phillips | Too toxic | 153.2 |
| Beatrice Crude | C3159-02 | 03/03/2009 | OSRLEARL | Not toxic enough, too thick to disperse | unknown |
| Heavy Bunker | C3159-03 | 03/03/2009 | OSRLEARL | Very heavy | unknown |
| Immingham IFO 2 | C3159-04 | 12/03/2009 | Conocco Phillips | 1st potential oil | unknown |
| Cardiff IFO 180 | C3159-05 | 01/10/2009 | Inver Energy | 2nd potential oil | 76.44 |
| Falmouth IFO 180 | C3159-06 | 27/10/2009 | WFS Falmouth | 3rd potential oil | 89.14 |

3.3.2 Heavy Oil test method

The existing sea test protocol was adapted for use with the selected candidate heavy oils (see above). The exposure time of 100 minutes was also retained.

The only test parameters adjusted as part of the early assessment process, in comparison to the standard conditions, were the motor speed (increased to a maximum of 1000rpm in order to disperse the heaviest oils) and the test oil concentration (on only two occasions).

There was also an issue to be overcome in terms of the accurate introduction of the test oil to the tank. In the usual sea test Kuwait crude oil is used, which can be easily and accurately dispensed from glass measuring cylinders. Due to the higher viscosity of the heavier test oils they were difficult to pour from the measuring cylinders and too much residue was left. Any attempt to scrape out the measuring cylinder left too much oil behind and on the spatula to be certain of the amount actually added to the tank.

A method was developed using positive displacement pipettes to dispense the heavier oils. This enabled an accurate volume of oil to be dispensed quickly and easily.

The dispersants were added to the heavy fuel oil test tanks as type 3 dispersants i.e. 1.8ml of neat dispersant was added dropwise to the surface of the oil in a containment ring. As soon as the dispersant was added the test motors were started.

A summary of the oils tested with the test conditions and the mortalities of *C. crangon* is displayed in table 2.

Table 2. Summary of heavy oils tested and their mortalities under different sea test conditions

| Oil | Source | Test ref | Test conditions | Sea test Result (Mean % Mortality) |
|------------------------|-----------------|----------|--------------------|------------------------------------|
| Immingham IFO 180 A | Conoco Phillips | C663 | 1000 RPM, 18ml oil | 65 |
| Immingham IFO 180 A | Conoco Phillips | C664 | 1000 RPM, 18ml oil | 82 |
| Immingham IFO 180 A | Conoco Phillips | C664 | 800 RPM, 18ml oil | 81 |
| Beatrice Crude HFO 380 | OSRLEARL | C670 | 1000 RPM, 18ml oil | 6 |
| Heavy Bunker HFO 380 | OSRLEARL | C667 | 1000 RPM, 18ml oil | 20 |
| Immingham IFO 180 B | Conoco Phillips | C668 | 1000 RPM, 18ml oil | 46 |
| Immingham IFO 180 B | Conoco Phillips | C669 | 1000 RPM, 18ml oil | 40 |
| Immingham IFO 180 B | Conoco Phillips | C670 | 800 RPM, 18ml oil | 61 |
| Immingham IFO 180 B | Conoco Phillips | C671 | 800 RPM, 18ml oil | 39 |
| Immingham IFO 180 B | Conoco Phillips | C673 | 800 RPM, 18ml oil | 75 |
| Immingham IFO 180 B | Conoco Phillips | C674 | 800 RPM, 18ml oil | 52 |
| Immingham IFO 180 B | Conoco Phillips | C675 | 800 RPM, 18ml oil | 65 |
| Immingham IFO 180 B | Conoco Phillips | C675 | 800 RPM, 18ml oil | 69 |
| Immingham IFO 180 B | Conoco Phillips | C675 | 800 RPM, 9ml oil | 54 |
| Immingham IFO 180 B | Conoco Phillips | C676 | 800 RPM, 18ml oil | 26 |
| Cardiff IFO 180 | Inver Energy | C681 | 800 RPM, 18ml oil | 11 |
| Cardiff IFO 180 | Inver Energy | C682 | 800 RPM, 18ml oil | 19 |
| Cardiff IFO 180 | Inver Energy | C682 | 1000 RPM, 18ml oil | 38 |
| Cardiff IFO 180 | Inver Energy | C682 | 800 RPM, 24ml oil | 32 |
| Cardiff IFO 180 | Inver Energy | C683 | 800 RPM, 18ml oil | 25 |
| Cardiff IFO 180 | Inver Energy | C684 | 800 RPM, 18ml oil | 16 |
| Falmouth IFO 180 | WFS Falmouth | C677 | 800 RPM, 18ml oil | 42 |
| Falmouth IFO 180 | WFS Falmouth | C678 | 800 RPM, 18ml oil | 24 |
| Falmouth IFO 180 | WFS Falmouth | C679 | 800 RPM, 18ml oil | 50 |
| Falmouth IFO 180 | WFS Falmouth | C680 | 800 RPM, 18ml oil | 48 |
| Falmouth IFO 180 | WFS Falmouth | C686 | 800 RPM, 18ml oil | 35 |
| Falmouth IFO 180 | WFS Falmouth | C687 | 800 RPM, 18ml oil | 26 |
| Falmouth IFO 180 | WFS Falmouth | C688 | 800 RPM, 18ml oil | 28 |
| Falmouth IFO 180 | WFS Falmouth | C689 | 800 RPM, 18ml oil | 22 |
| Falmouth IFO 180 | WFS Falmouth | C690 | 800 RPM, 18ml oil | 31 |
| Falmouth IFO 180 | WFS Falmouth | C691 | 800 RPM, 18ml oil | 9 |

3.3.3 Heavy Fuel Oil Test - Results

Once the candidate test oil and test conditions were selected, based on the above investigative studies, the aim of the project moved to generating a sufficiently large dataset to enable statistical analyses. These analyses would allow an assessment of the affects of a range of pass/fail criteria (see below) and to enable this the new test procedure was conducted on numerous occasions with a wide range of dispersant formulations. The results dataset for the new heavy fuel oil test is shown in table 3.

Table 3. Difference in mortality when dispersants are added to Falmouth Heavy oil IFO 180 under new heavy oil test procedure

| Test ref | Dispersant added | Control | Treatment | Difference (treatment - control) |
|----------|----------------------------------|---------|-----------|----------------------------------|
| C677 | oil only | 42 | | |
| C678 | oil only | 24 | | |
| C679 | Enersperse type 2 | 50 | 78 | 28 |
| C679 | Slickgone type 2 | 50 | 73 | 23 |
| C679 | AGMA type 2 | 50 | 75 | 25 |
| C679 | Superdispersant 25 Type 2 | 50 | 63 | 13 |
| C680 | oil only | 48 | | |
| C680 | Enersperse type 3 | 48 | 56 | 8 |
| C680 | Slickgone type 3 | 48 | 48 | 0 |
| C680 | AGMA type 3 | 48 | 57 | 9 |
| C680 | Superdispersant 25 Type 3 | 48 | 58 | 10 |
| C686 | oil only | 35 | | |
| C686 | Enersperse type 3 | 35 | 57 | 22 |
| C686 | Slickgone type 3 | 35 | 53 | 18 |
| C687 | oil only | 26 | | |
| C687 | AGMA type 3 | 26 | 73 | 47 |
| C687 | Finasol type 3 | 26 | 48 | 22 |
| C688 | oil only | 28 | | |
| C688 | Corexit 9500 type 3 | 28 | 64 | 36 |
| C688 | Slickgone NS type 3 | 28 | 39 | 11 |
| C689 | oil only | 22 | | |
| C689 | Superdispersant 25 type 3 | 22 | 63 | 41 |
| C689 | Radiagreen type 3 | 22 | 56 | 34 |
| C690 | oil only | 31 | | |
| C690 | Gamlen OD 4000 type 3 | 31 | 43 | 12 |
| C690 | Total fluides dev 2006-34 type 3 | 31 | 49 | 18 |
| C691 | oil only | 9 | | |
| C691 | Dispolene type 3 | 9 | 48 | 39 |
| C691 | Caflon type 3 | 9 | 29 | 20 |

3.4 Type 3 Test Assessment

3.4.1 Type 3 Test Development– Dispersant Selection

For the Type 3 dispersant tests Kuwait crude oil was used as per the current standard Sea Test protocol. The aim here was to test as many different dispersants as possible to enable an assessment of how toxicity changes when the dispersant is added as a Type 3 product and to assess what this would mean in terms of a change to the pass/fail criteria as compared to the current situation for addition of dispersants as a Type 2 product.

A range of dispersant samples were obtained from the Maritime and Coastguard Agency (MCA) national stockpile to represent the current stocks held in the UK. Some dispersants were also used from sample stocks held at Cefas.

3.4.2 Type 3 dispersant test method

For the Type 3 test data generation the sea test protocol was followed using 15 tanks for each test i.e. 5 tanks control, 5 tanks control plus dispersant added as type 2, and 5 tanks plus the same dispersant added as type 3 (the Type 2 exposures were included to allow some comparison to these results).

Type 2 dispersants were added by first diluting to 10% with seawater then adding 18ml to the surface of the oil (equivalent to a 1:10 dispersant to oil ratio). Type 3 dispersants were added in the same manner as for the heavy oil tests i.e. 1.8ml of neat dispersant added to 18ml of oil in a containment ring (also equivalent to a 1:10 dispersant to oil ratio).

3.4.3 Type 3 Tests - Results

Again it was necessary to generate a sufficiently extensive dataset that would enable, through statistical analysis, the assessment of new pass/fail criteria for products tested as a Type 3 product. New data generated as part of this project was combined with historically available test data to create the test dataset. The results are shown in table 4.

Table 4. Difference in mortality for a range of dispersants when applied to Kuwait crude oil as type 2 or type 3.

| Dispersant | Test ref | Control | Type 2 | Type 2 mortality increase (% over control) | type 3 | Type 3 mortality increase (% over control) |
|------------------------------------|------------|---------|--------|--|--------|--|
| Slickgone | C685 | 23 | 10 | -13 | 35 | 12 |
| AGMA | C685 | 23 | 23 | 0 | 50 | 27 |
| Superdispersant 25 | C666 | 10 | 37 | 27 | 78 | 68 |
| Superdispersant 25 | C665 | 25 | 42 | 16 | 67 | 41 |
| Superdispersant 25 | C662 | 29 | 45 | 16 | 57 | 28 |
| Slickgone EW | C661 | 21 | 40 | 19 | 75 | 54 |
| Slickgone EW | C680 | 17 | 45 | 28 | 72 | 55 |
| Superdispersant 25 | 2005 1 | 22 | 46 | 24 | 69 | 47 |
| Slickgone EW | 2005 2 | 30 | 69 | 39 | 74 | 44 |
| Finasol OSR-51 | 2005 3 | 68 | 73 | 5 | 81 | 13 |
| Caflon OSD | 2005 4 | 43 | 60 | 17 | 75 | 32 |
| Agma OSD 369 | 2005 5 | 29 | 52 | 23 | 60 | 31 |
| Corexit 9527 (rep 1) | Pre 2000 A | 13 | 76 | 63 | 85 | 72 |
| Corexit 9527 (rep 2) | Pre 2000 B | 52 | 75 | 23 | 79 | 27 |
| Corexit 9527 (rep 3) | Pre 2000 C | 43 | 52 | 9 | 69 | 26 |
| Corexit 9527 (rep 4) | Pre 2000 D | 33 | 42 | 9 | 56 | 23 |
| Corexit 9527 (rep 5) | Pre 2000 E | 32 | 66 | 34 | 78 | 46 |
| Dasic Slickgone SC100 | Pre 2000 F | 35 | 48 | 13 | 56 | 21 |
| Petrolite Tretolite W29-22 (rep 1) | Pre 2000 G | 8 | 45 | 37 | 58 | 50 |
| Petrolite Tretolite W29-22 (rep 2) | Pre 2000 H | 16 | 32 | 16 | 46 | 30 |
| Petrolite Tretolite W29-22 (rep 3) | Pre 2000 I | 12 | 55 | 43 | 61 | 49 |
| Dispolene 34S (rep 1) | Pre 2000 J | 49 | 50 | 1 | 59 | 10 |
| Dispolene 34S (rep 2) | Pre 2000 K | 45 | 26 | -19 | 39 | -6 |
| Dispolene 34S (rep 3) | Pre 2000 L | 37 | 53 | 16 | 75 | 38 |
| Dispolene 34S (rep 4) | Pre 2000 M | 25 | 42 | 17 | 45 | 20 |
| Finasol OSR 5 (rep 1) | Pre 2000 N | 52 | 67 | 15 | 70 | 18 |
| Finasol OSR 5 (rep 2) | Pre 2000 O | 21 | 43 | 22 | 46 | 25 |

4 Statistical Analysis (Test validity and setting of pass/fail criteria)

4.1 Statistical Test of the Difference in Mortality Rates of the Control and Test Treatments

The theory and practice for this approach has been developed previously for type 2 dispersant tests. We summarise this here and then explain why it needs to be modified for the type 3 approach.

4.2 Experimental Design

The experimental design consists of five replicate tanks for each of the *Control* and *Test Treatment*, each tank containing n_i animals, $i = 1..5$. The target for n_i is 20, although in practice n_i may vary slightly. At the end of the experimental period, the number of dead animals in each tank, d_i , is noted, from which the percentage mortality in each tank

$$m_i = \frac{d_i}{n_i} \times 100\%$$

is computed.

For each treatment, the total number of animals and total number of dead animals are

$$n = \sum_{i=1}^5 n_i$$

and

$$d = \sum_{i=1}^5 d_i$$

respectively, from which the total percentage mortality is

$$m = \frac{d}{n} \times 100\% .$$

In an obvious notation, n_C , d_C and m_C correspond to the totals for the *Control Treatment*, and n_T , d_T and m_T to the totals for the *Test Treatment*.

4.3 Data Screening and Test of Between-Tank Homogeneity

Kirby *et al* (1996) described several criteria for assessing whether a test is valid. These include that m_T must fall in the range 10-80% and that for both the *Control* and *Test Treatments* the underlying mortality rates in each tank should be the same (i.e. homogeneous). This second criterion can be judged statistically by assuming that the mortality of one animal is independent of the mortality of another. Between-tank homogeneity can then be tested using a simple chi-squared test comparing

$$\chi^2 = \sum_{i=1}^5 \frac{(d_i - n_i d/n)^2}{n_i d/n} + \sum_{i=1}^5 \frac{(d_i - n_i d/n)^2}{n_i (n-d)/n}$$

and a 5% critical value given by the 95% quartile of a chi-square distribution with 4 degrees of freedom (i.e. one less than the number of tanks).

This test is conducted for both the *Control* and *Test* Treatments. If, in both cases, the computed values of χ^2 do not exceed the 95% quantile, then there is no evidence against

between-tank homogeneity. It is then appropriate to proceed to the next stage of the analysis, a comparison of the *Control* and *Test Treatment* mortality rates.

4.4 Comparison of Control and Test Treatment Mortality Rates

Writing

$D = m_T - m_C$ (where m_T and m_C are expressed as proportions)

for the difference between the *Test* and *Control Treatment* percentage mortality rates with

standard error of $D = \sqrt{m_T(1-m_T)/n_T + m_C(1-m_C)/n_C}$,

then an appropriate test statistic is

$$t = \frac{D}{\text{standard error of } D} \quad (1)$$

This is compared to the 95% quartile of a standard Normal distribution (which is 1.645). If t does not exceed the 95% quartile, then there is no evidence that the mortality rate in the *Test Treatment* exceeds that in the *Control Treatment*.

4.5 Decision Procedure for Approval Applications for type 2 dispersant tests (current approach)

Following the comparison of the mortality rates, a compound decision making procedure is followed to determine whether a licence is issued. Note that the value 20% defined here is later referred to as DIFF.

- If the outcome of the D test is not statistically significant (i.e. t is less than the 95% quartile), then the licence is issued.
- If the test is significant and $D \geq 20\%$, then a licence is refused.
- If the test is significant but $D < 20\%$, then the test is repeated. If the second test is not significant, a licence is issued. (n.b. if the second or subsequent tests show between-tank heterogeneity, they are repeated as necessary)
- Otherwise, for a significant second test, the licence is refused.
- NOTE: The above process is the basis of the core assessment process. For all assessments scientific expertise and judgement is a central part of the final decision making and may, for example, require extra testing to increase confidence in the final assessment and licensing decision.

4.6 Choosing decision procedures for type 3 tests

It was established that using similar decision procedures for type 3 tests may not be appropriate because the higher mortalities associated with type 3 procedures mean that substances will almost never pass. In the following section we explain how this conclusion was drawn and a modified decision procedure for type 3 tests is described.

4.7 Power of the D statistical significance test

The statistical significance test comparing control and treatment mortalities using the t-statistic defined in equation (1) is important in the procedure for type 2 tests. If the test yields a result which is statistically significant on both the first and second occasions then

the treatment will fail. This was probably done because the power of the test fitted in well with the DIFF value of 20% used in the Nicholson and Kirby procedure (see Appendix 1). However, it is important that the power of the test is understood, particularly if higher values of DIFF than 20% are considered.

The power of the test can be defined in terms of the difference in mortalities, $P_t - P_c$, between the control and treatment tests, as well as the value of P_c . This second component is needed because the power is not the same for different values of P_c . The power of the test is defined as the probability that the test is statistically significant at the 95% level.

Ten thousand simulations were carried out to determine the power. Four values of P_c were selected: 10, 20, 30 and 40%. The value of P_t went from $(P_c + 1)$ to 100%. The value of the difference in mortality ($P_t - P_c$) for which the power was 99% was also calculated. That is to say that, for greater differences than this, it is virtually certain that the test will yield a statistically significant result. The results of the power studies are shown in Figures 1 and 2. Figure 1 is for 5 tanks (100 animals) and Figure 2 is for 3 tanks (60 animals). The 99% power differences are shown in Table 5 below.

Table 5: Differences $P_t - P_c$ for which the power is 99%, for four different values of P_c .

| P_c | 5 tanks | 3 tanks |
|-------|---------|---------|
| 10 | 0.22 | 0.30 |
| 20 | 0.26 | 0.33 |
| 30 | 0.27 | 0.34 |
| 40 | 0.27 | 0.34 |

From Table 5 it is seen that if P_t is more than 27% points greater than P_c then, using the Nicholson and Kirby procedure, the substance will never pass – because the statistical significance test will always be significant. For 3 tanks, the equivalent figure is about 34%. Thus, if a scheme was designed to require treatments that have mortalities such as these to pass, then there is a need to modify the test procedure. This is illustrated further in the

next section where the proportion of tests that pass under various selected pass/pass criteria is considered.

4.8 A decision procedure for heavy fuel oil and type 3 dispersant tests

From the power studies described above it was shown that it is difficult to pass the Nicholson and Kirby test if the treatment and control mortalities are more than about 20% (5 tanks) or 30% (3 tanks) different. This is confirmed by the plots in Figures 3 and 4 where the whole procedure as outlined above has been simulated. The plots show the percentage passing the tests as a function of $P_t - P_c$. The value of DIFF used is 20% in both plots.

A value judgement needs to be made in order to decide whether it is acceptable to have treatments passing if they have, say, mortalities of 30 percentage points higher than the control. This would be needed in order to test the type 3 and heavy oils. If this is acceptable, then we need to develop a new test procedure.

The following new procedure is proposed (which assumes that the tests for between-tank homogeneity are continued).

1. Choose your value of DIFF at an appropriate level such as 30%.
2. The test fails if the observed difference in mortality between control and treatment tests is greater than DIFF and if the D statistical significance test is significant. If either of these criteria are not true then the treatment passes.
3. As with previous procedures, scientific judgement will play an important part in the assessment process and the eventually agreed levels of DIFF will need to be selected by the scheme regulators (Marine Management Organisation).

Below, how this new procedure works for a number of test experiments using heavy fuel oil (Falmouth oil) and type 3 procedures using the standard Kuwait crude oil is evaluated.

4.9 Results for Falmouth oil (the selected heavy fuel oil)

Table 6 summarises the results for the Falmouth oil. In practice, some of these experiments might have been repeated because of differences in mortalities between tanks. However, the results presented in table 6 still give a good idea of how different values of DIFF affect pass/fail for the Falmouth oil. The last row of Table 6 gives the percentage of tests that would pass at each level of DIFF. It can be seen that only 50% pass when DIFF is set to 20, but that 69% pass when DIFF is at 30.

4.10 Results for Type 3 tests using Kuwait Crude Oil as control

Table 7 summarises the results for the type 3 dispersant tests using Kuwait crude oil during this study. The mortalities here are generally higher than for the Falmouth oil procedure. This makes it easier for the chi-squared test to detect differences between tanks. This is reflected in the fact that on only two of the seven experimental dates do all the controls and treatment pass the between-tank homogeneity test. However, all results have been included in the dataset for illustration and for improved overall assessment.

Historical Type 3 dispersant testing results (from 2005 and before) are also included in Table 7 and the overall assessment. For some, the original raw test data are not fully available and so the chi-squared tests to test for homogeneity between tanks or the ion of

the D-test for these data cannot be completed. However, the value of DIFF is known. Therefore the results of these earlier tests have been included so as to provide a much larger pool of tests to evaluate how the various values of DIFF would perform if used for these data (ignoring the chi-squared and D-test components of the testing procedure). It can be seen that, from the last row of Table 7, 59% of tests would pass if DIFF was set at 40, and 78% would pass if DIFF was set at 50.

Table 6: Summary of results for Falmouth oil (using type 3 procedure)

| Substance | N | Pass or fail for various DIFF | | | | D-test stat significant? | Observed D |
|--|-----|-------------------------------|-----------|-----------|------------|--------------------------|------------|
| | | 20 | 30 | 40 | 50 | | |
| Enesperse: 22/10/09 | 60 | P | P | P | P | N | 8.3 |
| Slickgone: 22/10/09 ¹ | 60 | P | P | P | P | N | 0 |
| AGMA: 22/10/09 | 60 | P | P | P | P | N | 8.3 |
| Superdispersant: 22/10/09 | 60 | P | P | P | P | N | 10 |
| Enesperse: 3/2/10 ² | 100 | P | P | P | P | Y | 15 |
| Slickgone: 3/2/10 ² | 100 | F | F | F | P | Y | 45 |
| AGMA: 8/2/10 | 100 | F | F | F | P | Y | 47 |
| Finasol: 8/2/10 | 100 | F | P | P | P | Y | 20.3 |
| Corexit: 10/2/10 ³ | 100 | F | P | P | P | Y | 20.3 |
| Slickgone: 10/2/10 ³ | 100 | F | F | F | P | Y | 45 |
| Superdispersant: 22/2/10 | 80 | F | P | P | P | Y | 28 |
| Radiagreen: 22/2/10 | 120 | F | F | P | P | Y | 34 |
| Gamlen OD 400: 25/2/10 | 100 | P | P | P | P | N | 4 |
| Total fluides: 25/2/10 | 80 | P | P | P | P | N | -1 |
| Dispolene: 1/3/10 | 100 | F | F | P | P | Y | 34 |
| Caflon: 1/3/10 | 100 | P | P | P | P | Y | 19 |
| Percentage passing at each DIFF level | | 50 | 69 | 81 | 100 | | |

1 failed between tank homogeneity test

2 used a control tank with 0 mortality

3 removed results for control tank 5 as high outlier

Table 7: Summary of type 3 experiments using Kuwait crude oil as the control.

| Substance | N | DIFF | | | | D-test stat | Observed D |
|---------------------------------------|-----|-----------|-----------|-----------|-----------|-------------|------------|
| | | 20 | 30 | 40 | 50 | | |
| Slickgone: 26/11/09 | 60 | P | P | P | P | N | 12 |
| AGMA: 26/11/09 | 60 | F | P | P | P | Y | 27 |
| Superdispersant: 9/3/09 ¹ | 100 | F | F | F | F | Y | 66 |
| Superdispersant: 4/3/09 ² | 100 | F | F | F | P | Y | 43 |
| Superdispersant: 16/2/09 ² | 100 | F | P | P | P | Y | 28 |
| Slickgone EW: 12/2/09 ¹ | 100 | F | F | F | F | Y | 54 |
| Slickgone EW: 10/2/09 ¹ | 100 | F | F | F | F | Y | 54 |
| Previous data | | | | | | | |
| Slickgone | | P | P | P | P | | 12 |
| AGMA | | F | P | P | P | | 27 |
| Superdispersant 25 | | F | F | F | F | | 68 |
| Superdispersant 25 | | F | F | F | P | | 41 |
| Superdispersant 25 | | F | P | P | P | | 28 |
| Superdispersant 25 | | F | F | F | P | | 47 |
| Slickgone EW | | F | F | F | F | | 54 |
| Slickgone EW | | F | F | F | F | | 55 |
| Finasol OSR-51 | | P | P | P | P | | 13 |
| Caflon OSD | | F | F | P | P | | 32 |
| AGMA OSD 369 | | F | F | P | P | | 31 |
| Corexit 9527 | | F | F | F | F | | 72 |
| Corexit 9527 | | F | P | P | P | | 27 |
| Corexit 9527 | | F | P | P | P | | 26 |
| Corexit 9527 | | F | P | P | P | | 23 |
| Corexit 9527 | | F | F | F | P | | 46 |
| Dasic Slickgone SC100 | | F | P | P | P | | 21 |
| Petrolite Tretolite W29-22 | | F | F | F | P | | 50 |
| Petrolite Tretolite W29-22 | | F | P | P | P | | 30 |
| Petrolite Tretolite W29-22 | | F | F | F | P | | 49 |
| Dispolene 34S | | P | P | P | P | | 10 |
| Dispolene 34S | | P | P | P | P | | -6 |
| Dispolene 34S | | F | F | P | P | | 38 |
| Finasol OSR 5 | | P | P | P | P | | 18 |
| Finasol OSR 5 | | F | P | P | P | | 25 |
| Percentage passing at each | | 19 | 50 | 59 | 78 | | |

1 failed between tank homogeneity test

2 both control and treatment fail between tank homogeneity test

3 control failed between tank homogeneity test

Power of the significance test with 5 tanks

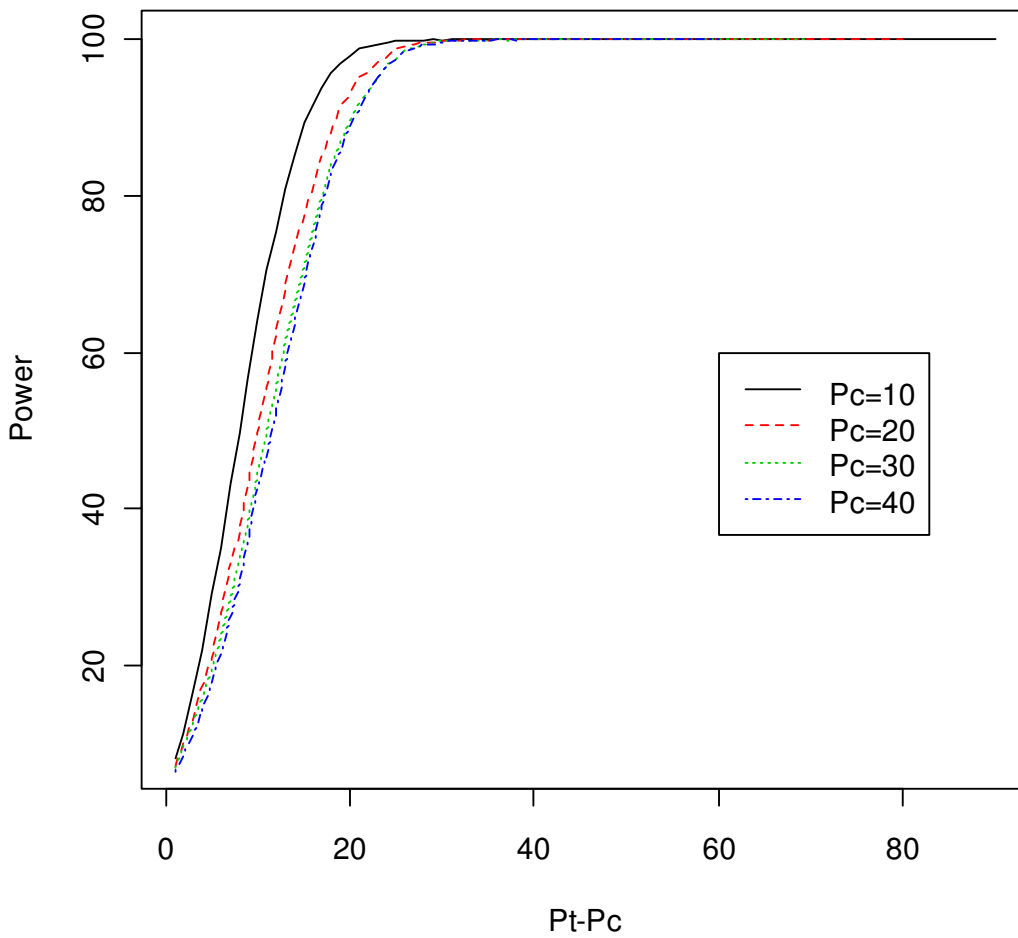


Figure 1: Power of the D statistical test as a function of Pt-Pc and four values of Pc. Five tanks are used.

Power of the significance test with 3 tanks

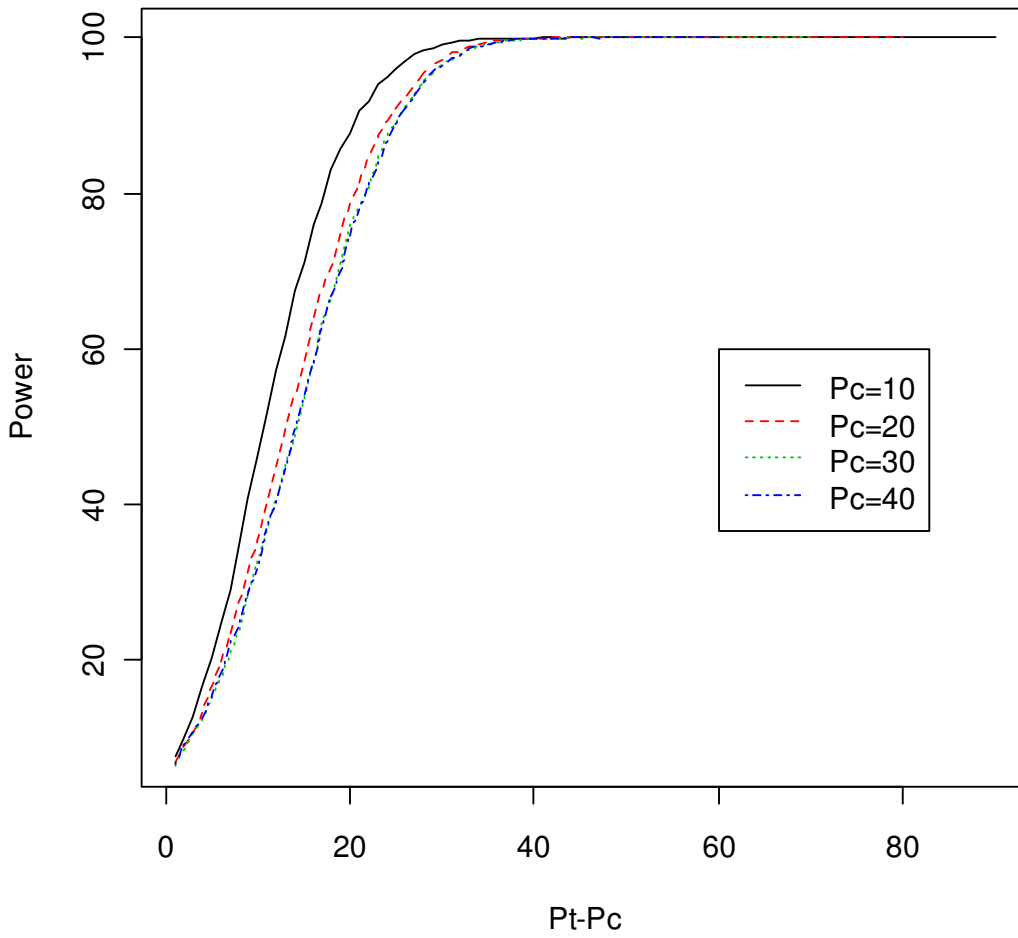


Figure 2: Power of the D statistical test as a function of Pt-Pc and four values of Pc. Three tanks are used.

Cardiff (Pc=16%): % passing overall

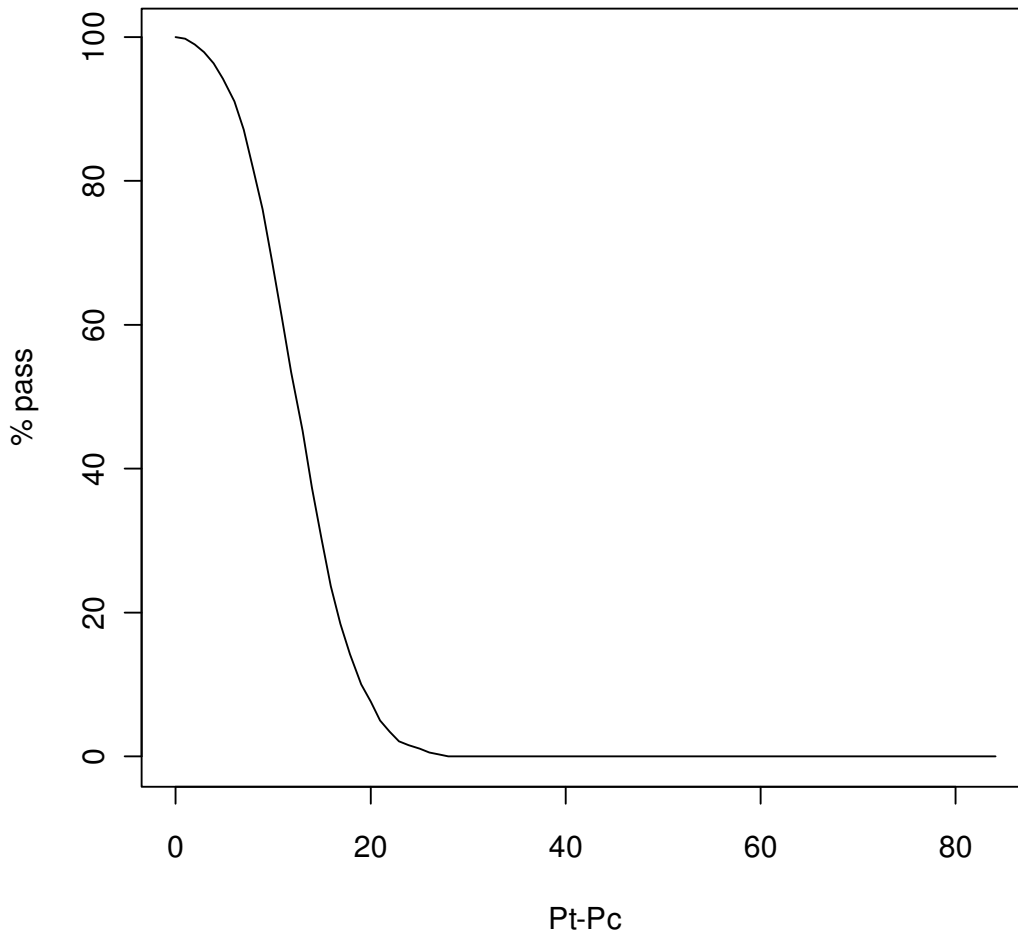


Figure 3: Percentage passing the Nicholson and Kirby test as a function of Pt-Pc. This is for the Cardiff oil mortality which has Pc=16%.

Falmouth ($P_c=0.49\%$): % passing overall

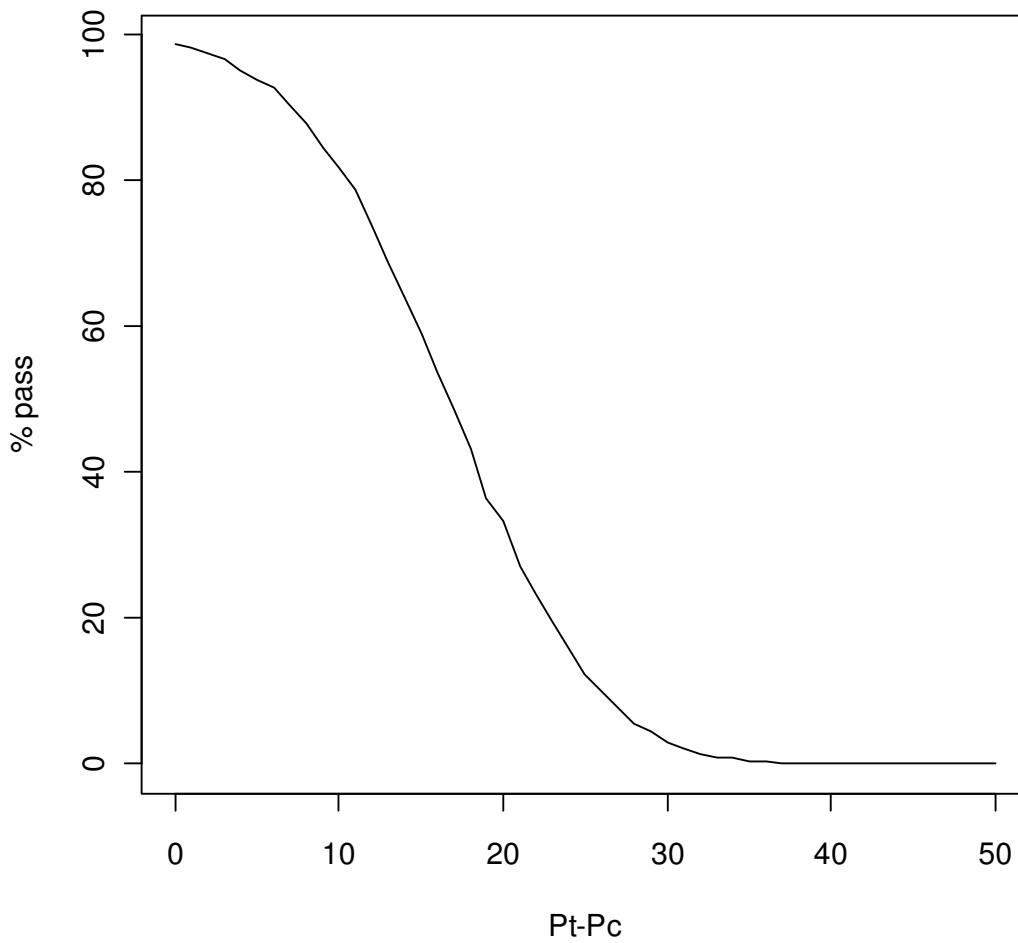


Figure 4: Percentage passing the Nicholson and Kirby test as a function of P_t-P_c . This is for the Falmouth oil mortality which has $P_c=49\%$.

5 Discussion & Conclusions

5.1 Heavy fuel oil test development

This project has successfully adapted the current Sea Test toxicity assessment (Kirby et al 1996) approach to allow the assessment of products for use in treating heavier oils (e.g. fuel oil spills) at sea. The process essentially comprised the testing of a range of candidate oils to select one of suitable toxicity and dispersability and then the subsequent running of an adapted Sea Test using the oil with a range of dispersants to generate a dataset for statistical assessment.

In the initial assessment of test oil suitability a range of test parameters that could affect the test results (e.g. temperature, mixing energy, oil: dispersant ratios and concentrations etc.) were considered with the aim of establishing appropriate test conditions as necessary for the development of a standard testing protocol. These preliminary investigations suggested that appropriate dispersability and toxicity could be gained with candidate heavy fuel oils under similar test conditions already established for the standard Sea test which uses Kuwait crude as the standard oil.

The biggest challenge which the project addressed was in the selection of an acceptable standard oil, especially with respect to it being of an appropriate 'heaviness' or viscosity to represent the types of oil that the newly assessed products will be used to treat in response scenarios. The finally selected oil was an IFO 180 sourced from WFS in Falmouth based on its performance under the test conditions. The Falmouth IFO 180 oil had a measured kinematic viscosity (measured at 50°C) of around 90 cSt (the current standard Sea test oil, Kuwait crude, has a comparatively low kinematic viscosity of <10 cSt).

It has been shown that certain dispersants, under certain conditions, can successfully disperse oils of much higher viscosity than the Falmouth IFO 180 (e.g. IFO380 during the MSC Napoli incident) and the issue, therefore, was whether this test oil could be regarded as of too low a viscosity compared to that of the upper range of oils that could potentially be treated with dispersants. After wide consultation it was agreed that very heavy or residual fuel oils (VHFO's) are generally considered as non-dispersible and therefore oils categorised as IFO 380 are likely to represent the very upper limit of those that could be readily dispersed – and even then only under certain, favourable, conditions. Therefore an IFO180 was considered to be the highest viscosity category that would be *routinely* amenable to dispersion and it is on that basis that the Falmouth IFO180 was considered as an appropriate standard oil for the development of the test protocol. Furthermore, the use of a more dispersible oil allowed for a greater level of reproducibility within the proposed tests and was considered likely to rank products in a more similar manner than if a heavier option were selected.

Importantly, the development and implementation of the new test procedure will allow the approval of products with the specific aim of treating heavier oils. This enhancement of the current response toolbox is essential when one considers that the majority of oil spills are now from fuel spills rather than cargo and that for European coasts bordering the Baltic, Barents and North Seas and the English Channel, there is an heightened risk from such

oils due to the increased transport of heavy and residual fuel oils from the former Soviet Union (Kirby and Law 2010).

5.2 Type 2 and Type 3 testing

This project also addressed successfully the issue pertaining to the need for separate testing of products as type 2 (addition after water dilution) or type 3 (neat addition) dispersants. The comparative test data generated within this project has expanded and confirmed preliminary indications that dispersants added as a Type 3 product routinely cause higher levels of toxicity (under Sea Test conditions) than when added as a Type 2 dispersant – even though exactly the same amount of dispersant has been added (as type 2 dispersants are diluted 1:10 with water but 10x as much dispersant is added).

The aim of the project, on this issue, was to generate sufficient test data such that the toxicity difference between the two dispersant application methods could be properly quantified, in order that appropriate statistical analysis could be conducted to establish new pass/fail criteria if products need to be retested for Type 3 use.

Current test procedures, whereby products can be tested as a Type 2 product and given approval as a Type 2/3 were established in the early period of the UK scheme (late 1970's/early 1980's) when evidence of the difference in Sea Test performance was not manifest. However, more effective products developed after this time have subsequently proved to exhibit this difference in performance/toxicity. Furthermore, products are more operationally deployed as a Type 3 product, especially from aircraft, and therefore it is important to approve products on the basis of their mode of application.

5.3 Implementation into Statutory Scheme

The research provided in this project provides the underpinning science that will allow for the introduction of new test protocols and assessment criteria to be implemented into the statutory oil treatment product testing and approval scheme. However, the actual use of the information and the final shape of a new scheme will be determined by the Marine Management Organisation (MMO) – the administrators of the scheme. These detailed discussions will take place during 2010.

The issues to be discussed as part of the implementation process include:

5.3.1 Setting of Pass/Fail criteria

The current UK toxicity tests have pass/fail criteria based around whether a treatment significantly ($p < 0.05$) increases toxicity compared to an oil only control. In practice this equates to a percentage mortality difference of ~20%. If this same criterion were applied to the newly developed heavy fuel oil test or, indeed, assessment as a Type 3 dispersant, it is likely that very few products would pass. It would, therefore, not seem appropriate to apply the same pass/fail criteria for the new assessments and consideration should be given to how the new procedures can be applied to screen out the worst and only approve the best performing products. Reference to tables 6 and 7, for heavy oil and type 3 tests respectively, suggest that DIFF levels of 30-40% would be required to differentiate, and approve, the top 70-80% of products for treating heavy oils and the top 50-60% of products to be used as Type 3 products. The exact levels of DIFF chosen will have implications for the breadth of products available to responders and may need to be

informed by specific use scenarios. The shape of the final implementation will be developed by the Marine Management Organisation after appropriate consultation.

5.3.2 Need for a Rocky Shore Test

In the current approval scheme it is compulsory that all products have to pass both the Sea Test and Rocky Shore test and that to gain approval a product had to pass both assessments. During the 2007 review it was argued by some stakeholders that a Rocky Shore test assessment (for which a number of products are not approved in the UK because they failed this test) was inappropriate for products whose primary use would be offshore and not in the vicinity of rocky shores. It is likely that the application of dispersant products on heavy fuel oils and/or in the Type 3 application mode will normally only occur in offshore environments. With that in mind, this project did not aim to develop a Rocky Shore test equivalent for use with heavy oils, however, products applied as a Type 3 product can already be assessed using the current Rocky Shore test if necessary (though comparative toxicity data for products added as a Type 2 or Type 3 product are not available to enable assessment of appropriate pass/fail criteria). Therefore the MMO, in consultation with its scientific advisors, will need to decide the status of the new product assessments and potential use on or near rocky shores.

5.3.3 Putting the new scheme to the industry

A further important issue will be how to implement the new assessment processes in a way that is acceptable to the industry. The oil spill treatment product industry is quite unlike major industries (pharmaceutical, agrochemical, oil & gas, etc.) in that it is limited in scale and some of the major manufacturers are small companies. As such, the cost of product assessment has always been seen as high by many in the industry, compared to the size of the market. This has, in the past, been responsible for the withdrawal of products and manufacturers from the sector because it makes it financially non-viable for them. In introducing the new assessments for heavy oil treatment products and for Type 3 use, the regulator is potentially introducing initial reassessment costs some of the manufacturers will not feel able to meet. The risk is that some products/manufacturers are withdrawn from the 'toolbox' that our national responders draw upon. The aim of the amended scheme is to strengthen the response toolbox and therefore consideration needs to be given as to the impact on the industry of the new assessments. Options for consideration might include interim arrangements whereby product assessments for current products (or products submitted within a set time period) are subsidised in some way. Consideration will also need to be given as to how currently approved type 2/3 products are treated. At what point will their Type 3 approval be withdrawn and a reassessment required?

Ultimately this research and the subsequent consultations may lead to the amendment of the current standard testing protocols to reflect the new procedures. This cannot, however, be completed until the above issues have been discussed and agreed with the MMO. Once this has happened MMO and Cefas can jointly agree a protocol drafting and implementation schedule.

5.4 The Future

The research conducted under this project will allow Defra/MMO to address the primary issues identified in the 2007 review. However, there were a range of additional recommendations made as part of the review that have not yet been addressed and, in

amending the scheme, the opportunity should be taken to revisit the need for further work. For example, the new test procedures by necessity have been designed to stay as close to the current testing procedures as possible so as to maintain consistency across the scheme as much as possible. However, this project has not addressed some of the other issues that will need to be considered to keep the scheme 'fit for purpose' moving forward. These issues include:

- The need to ensure that the tests do not bias against highly efficacious products. Products of high efficacy, it is argued, may increase the oil exposure, and potential effects, in the short term, but that they may allow the oil to disperse more rapidly and therefore reduce the overall potential exposure time. This aspect is not currently taken account of in the current (and new) suite of tests, but the addition of a dilution element to the Sea Test protocol during the initial dispersed oil exposure phase could take account of this.
- Potential combined toxicity and efficacy test. In principle an efficacy measurement could be made using the water from the Sea Test. The development of a combined efficacy/toxicity test could be considered.
- Tests for surface cleaners. Currently the Sea and Rocky Shore tests are used for the assessment of a variety of products other than dispersants. For most, e.g. sorbents, the tests are perfectly appropriate. However, for the category of surface cleaners, the Sea and Rocky Shore test are not suitable, because the surface cleaners would never be applied to oil on the surface of the water as other products are. Development of a new protocol for testing of surface cleaners should be considered.
- Climate change etc.: Should the testing scheme be a simple screen for approval, or should the wider assessment take account of other factors? For example, oil and dispersed oil have been shown to exhibit phototoxicity (Kirby et al 2006?) which means that under conditions of UV exposure (as might occur on cloudless bright days or in clear waters) toxicity could be between 10-100 times higher to certain invertebrates and larvae. Conversely, cloudier, stormier weather which could increase suspended solids may reduce sunlight penetration. Climate Change issues are an important potential source of future variations in oil/dispersant efficacy and effects. These and other issues were highlighted as requiring further research in the conclusions of the 2007 review.

6 References

Kirby, M.F., Law, R.J. (2010) Accidental spills at sea – Risk, impact, mitigation and the need for coordinated post-incident monitoring. *Marine Pollution Bulletin, Volume 60, Issue 6, June 2010, Pages 797-803*

Kirby, M.F., Lyons, B.P., Barry, J., Law, R.J. (2007). The toxicological impacts of oil and chemically dispersed oil: UV mediated phototoxicity and implications for environmental effects, statutory testing and response strategies. *Marine Pollution Bulletin, 54: 464-488.*

Kirby, M.F., Matthiessen, P. and Rycroft, R.J. (1996) Procedures for the approval of oil spill treatment products. Fish. Res. Tech. Rep., MAFF Direct. Fish. Res., Lowestoft, (102): 19pp

7 Acknowledgements

This work was funded by the Department for Environment, Food & Rural Affairs under project ME1309. Thanks go to past colleagues in the Marine Fisheries Agency (Michael Meekums and Malcolm Peddar) and to present colleagues in the Marine Management Organisation (Nick Greenwood and Neil Wellum) for their input and support. Thanks also go to colleagues at the Maritime and Coastguard Agency and the suppliers of the test oils listed in the report. Finally, particular thanks are owed to Leah Nelson and Cheryl Moran for their commitment and technical support throughout the project.

8 Appendix 1: Overview of Statistical Toxicity Test

Overview of Statistical Toxicity Test for the Approval of Oil-Spill Treatment Products

Mike Nicholson and Mark Kirby

Kirby *et al* (1996) describe the experimental procedures for testing the toxicity of oil-spill treatment products relative to a control treatment. Here, the experimental design and the statistical analyses of the resulting data are defined. The analyses consist of a preliminary assessment of the homogeneity between replicate samples, followed by a simple comparison of the mortality rates of the control and test treatments if the preliminary test has been passed. The properties of this test are evaluated.

9 Introduction

A detailed description of the experimental procedures for testing the toxicity of an oil-spill treatment product is given by Kirby *et al* (1996). However, this description gives only a brief outline of the statistical method employed. The objective of this note is to provide a detailed description of the method, together with a worked example and the formulae to reproduce it. The statistical terms are defined in many textbooks e.g Campbell (1967), which also provide further statistical background and discussion.

The statistical method is applicable to two experimental procedures: The *Sea* test using the brown shrimp *Crangon crangon* as the test organism and the *Rocky Shore* test using the common limpet *Patella vulgata*. For the Sea Test the method measures and tests the statistical significance of the additional mortality in the test organisms of a candidate product when applied to a control oil. Hence the mortality rate in a *Test Treatment* consisting of a mixture of oil and product is compared with the mortality rate in a *Control Treatment* consisting of oil. For the Rocky Shore test the method measures and tests the significance of the difference between an oil-only *control treatment* and a product-only *test treatment*.

The experimental design and the statistical method are described in Section 2. This section also describes the decision-making procedure that follows the statistical tests, and evaluates the probability that a licence application will be successful. An example is given in Section 3.

Statistical Test of the Difference in Mortality Rates of the Control and Test Treatments

Experimental Design

The experimental design consists of five replicate tanks for each of the *Control* and *Test Treatment*, each tank containing n_i animals, $i = 1..5$. The target for n_i is 20, although in practice n_i may vary slightly. At the end of the experimental period, the number of dead animals in each tank, d_i , is noted, from which the percentage mortality in each tank

$$m_i = \frac{d_i}{n_i} \times 100\%$$

is computed.

For each treatment, the total number of animals and total number of dead animals are

$$n = \sum_{i=1}^5 n_i$$

and

$$d = \sum_{i=1}^5 d_i$$

respectively, from which the total percentage mortality is

$$m = \frac{d}{n} \times 100\% .$$

In an obvious notation, n_C , d_C and m_C correspond to the totals for the *Control Treatment*, and n_T , d_T and m_T to the totals for the *Test Treatment*.

Data Screening and Test of Between-Tank Homogeneity

Kirby *et al* (1996) describe several criteria for a valid test. These include that m_T must fall in the range 10-80% and that for both the *Control* and *Test Treatments* the underlying mortality rates in each tank should be the same (i.e. homogeneous). This second criterion can be judged statistically by assuming that the mortality of one animal is independent of the mortality of another. Between-tank homogeneity can then be tested using a simple chi-squared test comparing

$$\chi^2 = \sum_{i=1}^5 \frac{(d_i - n_i \frac{d}{n})^2}{n_i \frac{d}{n}} + \sum_{i=1}^5 \frac{(d_i - n_i \frac{d}{n})^2}{n_i \frac{(n-d)}{n}}$$

and a 5% critical value given by the 95% quantile of a chi-square distribution with 4 degrees of freedom.

This test is carried out for both the *Control* and *Test Treatment*. If, in both cases, the computed values of χ^2 do not exceed the 95% quantile, then there is no evidence against between-tank homogeneity. It is then appropriate to proceed to the next stage of the analysis, a comparison of the *Control* and *Test Treatment* mortality rates.

Comparison of Control and Test Treatment Mortality Rates

Writing

$$D = m_T - m_C$$

for the difference between the *Test* and *Control Treatment* percentage mortality rates with

$$\text{standard error of } D = \sqrt{m_T(1-m_T)/n_T + m_C(1-m_C)/n_C},$$

then an appropriate test statistic is

$$t = \frac{D}{\text{standard error of } D}.$$

This is compared to the 95% quantile of a standard Normal distribution. If t does not exceed the 95% quantile, then there is no evidence that the mortality rate for the *Test Treatment* exceeds that of the *Control Treatment*.

Decision Procedure for Licence Applications and Corresponding Error Rates

Following the comparison of the mortality rates, a compound decision making procedure is followed to determine whether a licence is issued:

- If the outcome is not statistically significant (i.e. t is less than the 95% quantile), then the licence is issued.
- If the test is significant and $D \geq 20\%$, then a licence is refused.
- If the test is significant but $D < 20\%$, then the test is repeated. If the second test is not significant, a licence is issued. (n.b. if the second or subsequent tests show between-tank heterogeneity, they are repeated as necessary)
- Otherwise, for a significant second test, the licence is refused.

Figure 1 shows the probability that a licence will be issued as a function of the true difference between the mortality rates of the *Test* and *Control Treatments*. This shows that at small differences between the mortality rates ($< 5\%$), the probability of a successful licence application is high ($> 95\%$). Similarly, at large differences between mortality rates ($> 22\%$) the probability of a successful licence application is low ($< 5\%$).

10 Example

Table 1 demonstrates the analysis using an artificial example. Here, there is no evidence against between-tank homogeneity for either the *Control* or *Test Treatment*, nor any evidence of any difference in their mortality rates. No further testing would be required, and in this instance, a licence would be issued.

Campbell, R.C. (1967) *Statistics for Biologists* Cambridge University Press. Cambridge, England.

Kirby, M.F., Matthiessen, P. and Rycroft, R.J. (1996) Procedures for the approval of oil spill treatment products. Fish. Res. Tech. Rep., MAFF Direct. Fish. Res., Lowestoft, (102): 19pp.



About us

Cefas is a multi-disciplinary scientific research and consultancy centre providing a comprehensive range of services in fisheries management, environmental monitoring and assessment, and aquaculture to a large number of clients worldwide.

We have more than 500 staff based in 2 laboratories, our own ocean-going research vessel, and over 100 years of fisheries experience.

We have a long and successful track record in delivering high-quality services to clients in a confidential and impartial manner.

Cefas Technology Limited (CTL) is a wholly owned subsidiary of Cefas specialising in the application of Cefas technology to specific customer needs in a cost effective and focussed manner.

CTL systems and services are developed by teams that are experienced in fisheries, environmental management and aquaculture, and in working closely with clients to ensure that their needs are fully met.

For on-line information see <http://www.cefas.co.uk>.

Customer focus

With our unique facilities and our breadth of expertise in environmental and fisheries management, we can rapidly put together a multi-disciplinary team of experienced specialists, fully supported by our comprehensive in-house resources.

Our existing customers are drawn from a broad spectrum with wide ranging interests. Clients include:

- international and UK government departments
- the European Commission
- the World Bank
- Food & Agriculture Organisation of the United Nations
- oil, water, chemical, pharmaceutical, agro-chemical, aggregate and marine industries
- non-governmental and environmental organisations
- regulators and enforcement agencies
- local authorities and other public bodies

We also work successfully in partnership with other organisations, operate in international consortia and have several joint ventures commercialising our intellectual property.

HEAD OFFICE

Centre for Environment, Fisheries & Aquaculture Science
Pakefield Road,
Lowestoft,
Suffolk NR33 0HT
UK
Tel +44 (0) 1502 562244
Fax +44 (0) 1502 513865

For more information see <http://www.cefas.co.uk>