

Prospective severity classification of scientific procedures in cephalopods: Report of a COST FA1301 Working Group survey

Cooke Gavan M^{1,*}, Anderson David B², Begout Marie-Laure³, Dennison Ngaire⁴, Osorio Daniel⁵, Tonkins Belinda⁶, Kristiansen Tore⁷, Fiorito Graziano^{8,9}, Galligioni Viola^{9,10}, Ponte Giovanna^{8,9}, Andrews Paul Lr^{8,9}

¹ Anglia Ruskin University, Faculty of Life Sciences, UK

² LASA, UK

³ Ifremer, Laboratoire Resources Halieutiques, France

⁴ Biological Services, University of Dundee, UK

⁵ School of Life Sciences, University of Sussex, UK

⁶ The College of Animal Welfare, Middlesex University, UK

⁷ Institute of Marine Research, Norway

⁸ Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Italy

⁹ Association for Cephalopod Research 'CephRes', Italy

¹⁰ Comparative Medicine Unit, Trinity College, Ireland

* Corresponding author : Gavan M. Cooke, email address : Gavan.Cooke@anglia.ac.uk

Abstract :

Cephalopods are the first invertebrate class regulated by the European Union (EU) under Directive 2010/63/EU on the protection of animals used for scientific purposes, which requires prospective assessment of severity of procedures. To assist the scientific community in establishing severity classification for cephalopods, we undertook a web-based survey of the EU cephalopod research community as represented by the participants in the European COoperation on Science and Technology (COST) Action FA1301, CephSInAction'. The survey consisted of 50 scenarios covering a range of procedures involving several cephalopod species at different life stages. Respondents (59 people from 15 countries) either allocated a severity classification to each scenario or indicated that they were unable to decide (UTD). Analyses evaluated score distributions and clustering. Overall, the UTD scores were low ($7.0 \pm 0.6\%$) and did not affect the severity classification. Procedures involving paralarvae and killing methods (not specified in Annex IV) had the highest UTD scores. Consensus on non-recovery procedures was reached consistently, although occasionally non-recovery appeared to be confused with killing methods. Scenarios describing procedures above the lower threshold for regulation, including those describing behavioural studies, were also identified and allocated throughout the full range of severity classifications. Severity classification for scenarios based on different species (e.g. cuttlefish vs. octopus) was consistent, comparable and dependent on potentially more harmful interventions. We found no marked or statistically significant differences in the overall scoring of scenarios between the demographic subgroups (age, sex, PhD and cephalopod experience). The COST Action FA1301 survey data provide

a basis for a prospective severity classification for cephalopods to serve as guide for researchers, project assessors and regulators.

Keywords : Cephalopods, Directive 2010, 63, EU, invertebrates, 3Rs, severity classification

Introduction

Prediction of the likely effects of experimental scientific procedures on live animals is a requisite of the European Union's Directive 2010/63/EU for the protection of animals used for scientific purposes.¹ When planning experiments, all procedures considered to be above the lower threshold for regulation should be classified into the categories "non-recovery", "mild", "moderate" or "severe" (as required by Article 15 of Directive 2010/63/EU). Prospective assessment and assignment of 'severity' based on the worst expected outcome for any animal undergoing the procedure must be included in applications for project authorization^{1, 2}, and the actual severity experienced by each individual animal which has undergone the procedure must be documented and reported accordingly (as required by Articles 38, 39 and 54).

Severity assessment is also an essential aspect of the 3Rs (reduce, refine, replace) principle, which is at the core of the Directive (Article 1) and is considered fundamental for experimental biology.³ Adequate assessment of severity classification is an important element in promoting good animal welfare, but is also important for scientific validity of the study since physiological and behavioural responses to any putative suffering, if not adequately assessed, may affect data quality and outcomes. Prospective severity classification is pivotal for good scientific study, and fundamental to inform the harm-benefit analysis undertaken by regulatory bodies, ethical and institutional review committees⁴.

The inclusion of "live cephalopods" (nautiloids, cuttlefish, squid and octopus) in Directive 2010/63/EU¹ brought, for the first time, a class of invertebrate animals (Phylum Mollusca: Class Cephalopoda) within the scope of legislation regulating the use

of animals for scientific and educational purposes. The term “live cephalopod” operationally includes all species (~800 known)⁵⁻⁷ from hatching.^{2, 8-10} In Europe a limited number of cephalopod species are currently utilized, including the common cuttlefish *Sepia officinalis*, the bobtail squid *Euprymna scolopes*, the European squid *Loligo vulgaris*, and the octopuses *Octopus vulgaris*, *Eledone cirrhosa* and *E. moschata*.¹¹ Analysis of cephalopod research assessing the potential impact of Directive 2010/63 identified multiple studies including procedures which would now fall within the scope of the Directive¹²

Directive 2010/63/EU (Article 3) defines the threshold for regulation as a procedure “which may cause pain, suffering, distress or lasting harm equivalent to, or higher than that caused by the insertion of a hypodermic needle in accordance with good veterinary practice”. There is clear evidence that cephalopods may be utilized for scientific purposes in a large number of studies in which this regulatory threshold is likely to be exceeded.^{2, 9, 12}

Although the European Commission Working-Group guidance on severity classification^{13, 14} is of general applicability to cephalopods the illustrative examples are exclusively based on laboratory mammals¹⁴, and this is also the case for the FELASA/ECLAM/ESLAV Working Group report published in 2018.⁴ For fish, the paucity of guidance on severity classification was recognised in 2009 by the Norwegian Consensus-Platform for the 3Rs (Norecopa) that established a Working Group to produce guidance on severity classification for scientific procedures involving fish.¹⁵ Even in the UK, where *Octopus vulgaris* has been regulated since 1993¹¹, neither in the current transposition of the Directive nor in the “Code of Practice for the Housing and

Care of Animals Bred, Supplied or Used for Scientific Purposes" is there information on severity classification applicable to any cephalopod*.

Similarly to what was achieved by Norecopa¹⁵, the COST Action FA1301 "A network for improvement of cephalopod welfare and husbandry in research, aquaculture and fisheries (*CephsInAction*)" initiated an activity specifically designed to derive a preliminary guidance document on prospective severity classification of procedures (within the meaning of Directive 2010/63/EU) for cephalopods to facilitate and act as a reference for project applications to the National Competent Authorities. This initiative was carried out as part of those planned within the aims of COST Action FA1301-Working Group 4 (Cephalopod Welfare) dedicated to facilitate and promote a collaborative framework for increasing knowledge on biological, physiological and behavioural aspects of cephalopods' welfare. Different from the approach used by Norecopa (a meeting with a published consensus statement),¹⁵ COST Action FA1301 *CephsInAction* launched a 'Prospective Assessment of Severity Survey' (PAS-C Survey) applying a Delphi methodology in the initial stages, to develop a consensus on prospective severity classification of regulated procedures using cephalopod molluscs[†]. The PAS-C Survey comprised 50 scenarios, each giving a brief description of a procedure based on the literature. Scenarios contained sufficient information to make an assessment, and supporting notes and definitions were provided to help respondents identify the main issues to consider in making the assessment.

* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/388535/CoPanimalsWeb.pdf

† <http://www.cephsinaction.org/activities/projects/pas-c/>

Here we report and analyze the outcomes of the PAS-C Survey in order to: *i.* identify examples of procedures in cephalopods in each of the prospective severity categories; *ii.* identify procedures where severity classification may be difficult; *iii.* provide researchers, project reviewers and regulators with examples of procedures in each severity category to inform allocation of prospective severity classifications.

Methods

The approach followed here differs from the approach adopted by Norecopa,¹⁵ by using a ‘Prospective Assessment of Severity Survey’ (PAS-C Survey) with the initial survey design inspired by the Delphi method^{16, 17} i.e., which aims to obtain a reliable opinion consensus on draft scenarios from a group of experts by subjecting them to a series of questionnaires interspersed with controlled opinion feedback. As described below, we carried out rounds of a modified Delphi process to develop the survey scenarios.¹⁷

Scenarios are based on data derived from a non-systematic review of the use of live cephalopods in studies published between 1940 and 2014, including biochemistry, physiology, immunology, neuroscience, eco-toxicology, and behavioural sciences. Procedures were selected for inclusion in the scenarios by the senior authors of this paper with the intention of encompassing the full range of possible severities, and were also based on a preliminary list in Fiorito et al.¹²

A critique of the methodology is included in the Discussion.

Survey development

- i) ***Proof of concept (first round of Delphi approach).*** The principle of using hypothetical scenarios based upon the literature was tested on four national regulators (United Kingdom, Irish Republic) using a pilot of 10 scenarios. Responses and feedback were assessed and informed the drafting of additional scenarios.
- ii) ***Drafting and revision of scenarios for the survey (second round of Delphi approach).*** Sixty-three scenarios and guidance notes were piloted on a group including a graduate student studying animal welfare (United Kingdom), a post-doctoral cephalopod researcher (Italy), a regulator (United Kingdom), two veterinarians (United Kingdom, Italy), an aquarium technician (United Kingdom) and established researchers (France, Italy, Norway, Portugal, United Kingdom). Scenarios with >60% consensus in any severity category were included in the final survey with modifications as required. Questions with <60% consensus were redrafted based on feedback and sent out for further review. Fifty scenarios were selected for the final PAS-C survey with the intention of covering the full range of severity classifications.

Scenarios included examples from twelve cephalopod species including *Nautilus*, *Sepia officinalis*, *Loligo vulgaris*, *Doryteuthis pealeii*, *Octopus vulgaris* and a number of other octopuses and other species also encompassing different life stages (**Table 1**).

Killing is not a regulated procedure under Directive 2010/63/EU provided that *Article 6* and *Annexe IV* are followed. However, as *Annexe IV* does not specify any method for

cephalopods we included four examples of killing methods in the scenarios to obtain a view from the community about their likely severity and to inform discussions of amendments to *Annexe IV*.

All the scenarios were considered to contain sufficient information to make a judgement about the procedure described. Variations around the same scenario were used to obtain insights into the factors influencing severity classification, and to check internal consistency.

Unable to decide category

This category was used to help identify scenarios where there was unintended ambiguity in the question, inadvertent omission of sufficient information to make a decision despite careful screening or where respondents felt that they were really unable to make a clear judgement.

***Guidance notes* (see **Supplementary Material**).**

To help and inform recipients of the survey, guidance notes were provided. These included:

- a) **Instructions.** These indicated that only one answer should be given and that if it was not possible to reach a clear decision about a category then the “*unable to decide*” (UTD) category should be used.

Emphasis was placed on making the prospective classification based upon the “*highest severity anticipated for any animal*”.

- b) **Definitions of the severity classifications.** Definitions of sub-threshold, non-recovery, mild, moderate and severe classifications were provided based on Annex VIII of the

Directive 2010/63/EU.^{4, 13, 14, 18} The list of definitions provided as guidance is included in **Table 2**. An “upper threshold” category was included to provide an insight into procedures which the respondents considered to exceed the “severe” classification and hence may not be acceptable or justifiable (see **Table 2**). It should be noted that procedures with this classification would not normally be authorised (see Directive Article 55,3). The inclusion of “upper threshold” aligns with the range of classifications used in the Norecopa report on severity assessment in fish.¹⁵

- c) **Factors to take into account.** This was based on Directive 2010/63/EU Annex *VII* and the EC Severity Assessment Framework.¹⁴
- d) **A list of general assumptions.** This broadly covers “contingent suffering” and includes handling of the animal, methods of anaesthesia, and post-operative recovery. The assumed competence of all personnel caring for the animals or performing procedures was also noted.

Personal profile

Personal information was requested from respondents for age, gender, academic and professional qualifications, number of publications (overall and cephalopod specific), current position or role, research experience with cephalopods. Survey respondents were not identifiable in any other way. Analysis was carried out on fully anonymised personal profile information (see also Results).

The on-line PAS-C survey (see Supplementary File 2)

Fifty scenarios, ranging in length from 18-292 words (median 108), were delivered online via the COST Action FA1301 web site.[‡] Scenarios were presented in 10 tabs, five scenarios per screen page. The guidance and definitions pages were always accessible. Scenarios could be completed in any order and answers saved and modified at any time until submission.

The survey was accessible to all COST Action FA1301 Cephs/nAction participants, who were emailed directly to inform them of the survey launch. This activity and the general principles included in the PAS-C survey was approved by the COST FA1301 Management Committee.

Consent by the COST Action FA1301 Cephs/nAction participants was implied by responding to the survey.

The on-line platform including the survey was open for three months (July to September 2015) with several reminders sent. At the end of the time, the on-line survey was closed and all responses and personal profile information were downloaded from the website and processed into a form suitable for numerical analysis. Any individual identifiers were not included in the download. Initial processing was by personnel not involved in the analysis of the data. The sorted anonymised raw data was then analysed.

Analysis of data

Demographics of respondents

The limited set of personal data (i.e., Personal Profile information) was used to investigate whether there was any relationship between the overall severity

[‡] <http://www.cephsinaction.org/activities/projects/pas-c/>

classification (see below) for scenarios and age, gender, possession of a PhD, number of publications or experience with cephalopods (as defined in the survey).

The sub-groups for the dichotomous variables of gender, PhD and cephalopod experience were fortuitously reasonably balanced with almost equal numbers in the sub-groups. To generate balanced sub-groups for the continuous variables of age and publication number we used a median split. Seven respondents had a veterinary qualification, but this sub-group was too small for a meaningful sub-group analysis (7 respondents with veterinary qualification vs. 52 without; see **Supplementary Data**).

Population responses to individual questions

For each scenario we initially analysed the percentage (%) of respondents answering in each of the seven (unable to decide, sub-threshold, non-recovery, mild, moderate, severe and upper threshold) answer categories. The scores were used to classify a scenario using dendrogram analysis (see **Allocating procedures to severity categories** below) and the scenario score profile (by bar and radar plots, see below).

Although “non-recovery” is in the list of prospective severity classifications, it is not part of a graded sequence of severity from sub-threshold (below the lower threshold for regulation) to the upper threshold. Therefore, we recalculated the % responses for all scenarios (except the three scenarios assessed as non-recovery) excluding the “non-recovery” scores for the 47 scenarios that were clearly not in this category. Subtraction of the non-recovery scores had no significant impact on the mean scores in each category over all 47 scenarios not in the non-recovery category (data not shown). The scores for 47 scenarios without the non-recovery scores are used in radar plots, and also to calculate a single score for each scenario to enable comparison between sub-groups

and ranking of scenarios (see *Allocating procedures to severity categories* and *Numerical scoring of scenarios for comparison of sub-groups* below).

Allocating procedures to severity categories

Three approaches were used to allocate scenarios to a particular severity classification.

First, two authors (GC and PA) independently reviewed the radar plots and the score profiles for each scenario. This allowed classification of scenarios where this was clear from the dominant category score and distribution of scores in adjacent categories. In the case of ambiguity, intermediate classifications (e.g., sub-threshold/mild, mild/moderate, moderate/severe) were allocated initially. Any scenario where allocation was not possible was placed initially in an ambiguous category.

Second, the distribution of % scores in each of the main and intermediate severity classes was averaged for each group of scenarios and the distribution of scores plotted.

Third, although the method above reflects the way in which a committee or project evaluator might arrive at a decision, it could be argued to involve some degree of subjective judgement so dendograms were constructed following an average linkage method^{19, 20} by using SPSS Statistics (PASW Statistics for Windows, Version 18.0, Chicago: SPSS Inc.). The average linkage method was applied to % scores in each of the categories for each scenario, and this allowed clustering of the data (i.e. scenarios).

Numerical scoring of scenarios for comparison of respondent sub-groups

Comparison of the responses to each of the scenarios between the demographic subgroups requires a single numerical score for each scenario.

In brief, we recalculated the % responses, excluding the “non-recovery” scores for the 47 scenarios that were clearly not in this category. To derive the single numerical score for each scenario we used the % score in each category and multiplied it by ‘1’ for unable to decide (the impact of the UTD score is minimal for most scenarios), ‘3’ for sub-threshold, ‘5’ for mild, ‘7’ for moderate, ‘9’ for severe and ‘11’ for upper threshold and then summed the scores. This transformation had no significant impact on the overall severity score for the scenario. Thus, if all (100%) respondents were unable to decide on a severity category the score would be 100. Similarly, for complete agreement on a scenario describing a sub-threshold study it would be 300, for mild it would be 500, and the highest possible score for any scenario would be 1100, if all respondents agreed that the scenario was in the upper threshold category.

The summed scores ranged from 332 (scenario 2) to 940 (scenario 43)[§]. These single values for each scenario: **i.** have been utilized exclusively for comparison of demographic sub-groups and to rank scenarios; **ii.** have not been used to determine severity classifications, which are only allocated using the original % responses in the survey (see ***Population responses to individual questions*** and ***Allocating procedures to severity categories*** above).

[§] the maximum possible score being 1100.

Statistics and data analysis

Descriptive statistics were used to describe participants' demographic characteristics and responses to scenarios. Further information on data processing and analysis are provided above and in particular for the analysis of: i. demographic data (see 'personal profile', 'demographics of respondents'), ii. responses and their allocation to severity categories (see 'Population responses to individual questions', Allocating procedures to severity categories', 'Numerical scoring of scenarios for comparison of sub-groups'). Whenever appropriate we reported the number of respondents to a particular scenario, personal profile question or number of scenarios (indicated as n), unless otherwise stated. In addition, the average responses are expressed as mean \pm SEM, unless otherwise indicated. For the non-dichotomous subgroups (i.e., age and number of publications) we utilized a median split to compare matched size groups above and below the median values (age and number of publications respectively). Further subdivision of the population was not possible as the sub-groups became too small for meaningful analysis. The processing of the raw survey data (e.g., grouping , sorting) was carried out using Microsoft Excel (2016). Data are non normally distributed (after Shapiro-Wilk test) and only non-parametric tests were used for any statistical comparisons. Graph plotting, descriptive statistics and correlations were carried out using were carried out using GraphPad Prism 7 & 8 for Windows (GraphPad Software, La Jolla California , USA). In particular, we calculated mean and median, standard error of the mean (SEM) and standard deviation (SD). We also used the Mann-Whitney U test, to evaluate differences in overall severity scores, and Spearman Rank Correlation for the analysis of demographic sub-group correlations. SPSS (PASW Statistics for Windows, Version 18.0, Chicago: SPSS Inc.) was used for cluster analaysis (dendrogram

construction) following an average linkage method^{19, 20}... Statistical significance was set at $p < 0.05$.

Results

Fifty-nine people from 15 EU countries and participants of COST Action FA1301 responded to the PAS-C survey, giving an overall response rate of 33.5% (see **Supplementary Tables 1 and 2** for details and demographics of respondents). Due to the biased distribution among EU countries of people either involved or potentially interested in cephalopod science, it is not surprising that responses primarily originated from five countries: Italy (21.2%), United Kingdom (21.2%), France (13.5%), Portugal (11.5%) and Spain (5.8%).

Scores in the unable to decide category

For all 50 scenarios the scores in the “unable to decide” (UTD) category** ranged from 3.4% to 25.9% of respondents with a mean of $7.0 \pm 0.61\%$ (**Supplementary Figure 1**). Eighty-six percent of scenarios had UTD scores less than 10% (64% of scenarios: < 5% UTD). There was no correlation between the overall severity score for each scenario, and the UTD score.

Seven scenarios (i.e., 10, 22, 27, 30, 31, 41, 48; see Supplementary Information) had UTD scores >1 standard deviation from the mean (i.e. > 11%). Of these, three involved studies of *O. vulgaris* paralarvae (i.e., scenarios 22, 27 and 31), one studied an analgesic

** Guidance notes provided for the PAS-C Survey indicated that this answer should be used when a respondent was unable to allocate a scenario to a severity classification

(scenario 10), one a general anaesthetic (scenario 48), and one a killing method (scenario 41). Scenario 22 had the highest UTD score (25.9%), and the most diverse range of responses of all 50 scenarios (see below).

Identification of scenarios describing non-recovery procedures

Three scenarios (5, 29 and 46) had a score in the non-recovery classification of $65.5 \pm 3.1\%$ with a UTD score of $5.7 \pm 0.5\%$. **Figure 1** illustrates the individual scores for these three scenarios and in each case shows the highest score in the non-recovery category as compared to all other categories. As the sub-threshold score for these three scenarios was <2%, respondents clearly recognised that these scenarios all came within the scope of Directive 2010/63/EU and allocated them “correctly” to the non-recovery classification.

As an independent measure of the accuracy with which the non-recovery scenarios were assessed and also to identify factors that may lead to an erroneous classification, we analysed the non-recovery scores for the remaining 47 scenarios that clearly did not meet the Directive definition of non-recovery. The score in the non-recovery classification for scenarios that are clearly not in this category was $3.4 \pm 0.6\%$ but for three scenarios (22, 26 and 28) the score was > 10% (mean $16.1 \pm 2.3\%$; range 13-20%). These three scenarios, with “erroneous” high non-recovery scores all described a method for killing a cephalopod with no other procedure involved; one involved an octopus paralarva (scenario 22), one a cuttlefish hatchling (scenario 26) and one an adult squid (scenario 28).

It is interesting to note that the final clustering of the overall % scores (**Figure 2**) indicated that the three non-recovery scenarios appeared in a single cluster (A in **Figure**

2). Additionally, three scenarios (22, 26 and 28) with non-recovery scores >10% (see above) were outside the non-recovery cluster A, and clustered in C2 (**Figure 2**).

Identification of scenarios below and around the lower threshold for regulation

For four scenarios (2, 32, 38, 42) the sub-threshold score was > 68% (range 68%-79%; Mean \pm SEM = $72.7 \pm 2.4\%$), with a clear separation of the sub-threshold from scores in the mild category and higher classifications (**Figure 3A**). The radar plot for these four scenarios also supports the allocation to the sub-threshold classification (**Supplementary Figure 2**). For these four scenarios the UTD score was $4.3 \pm 0.5\%$ so had no impact on the overall classification. These four scenarios cluster together (sub-cluster B1) in the dendrogram (**Figure 2**).

Whilst the survey population scored four scenarios predominantly below the threshold for regulation (sub-threshold), we also assessed how often the remaining forty-six scenarios (i.e. all scenarios except sub-threshold scenarios 2, 32, 38 and 42) were also scored as sub-threshold.

The percentage of the survey population assessing the remaining scenarios as sub-threshold was $11.2 \pm 2.4\%$ and for 16 of the 46 scenarios (35%) this value was < 5% of the survey population. However, for thirteen scenarios (i.e. 3, 8, 11, 12, 13, 16, 20, 22, 23, 27, 31, 34 and 40), over ten percent (range 10.3% - 52.6%) of the survey population considered each of these scenarios to describe procedures that were sub-threshold. Of these thirteen scenarios, five involved behavioural studies (i.e., 3, 12, 13, 16 and 34), three involved growth/feeding in paralarvae/hatchlings (20, 27 and 40), two killing paralarvae (22 and 31), one involved a non-invasive procedure performed under general anaesthesia with recovery (scenario 8), one a digestive tract transit study (scenario 11)

and one a study of post-reproductive senescence (scenario 23). The individual scores for eight of these scenarios are plotted in **Figure 3 (B and C)** together with the scores from the four sub-threshold scenarios (**Figure 3 A**) to illustrate the gradual shift from high sub-threshold scores to reduced sub-threshold scores with increased scores in mild or more severe categories.

Five scenarios (3, 12, 23, 34 and 40; **Figure 3 B and C**) have a distribution of scores indicating that they should be considered as describing procedures which are on the borderline for regulation (i.e., on the sub-threshold-“lower” mild borderline), whereas for the other eight scenarios the balance shifts to scores in severity classes predominantly above the lower threshold for regulation (**Figure 3 C** for examples).

In **Figure 2**, scenario 23 (with a sub-threshold score of 43%) describing a non-invasive observational study of senescent octopuses is allocated to the same cluster (B2) as scenarios 3, 12, 34, and 40, therefore just within the mild category. Cluster B includes closely related sub-clusters (B1) with sub-threshold scenarios and B2 with scenarios assessed as around the regulatory threshold (sub-threshold/mild borderline). However, as the scores in the sub-threshold category and the sum of scores for all categories above the lower regulatory threshold (sub-threshold $48.1 \pm 1.5\%$ vs. above lower threshold $48.4 \pm 1.8\%$, $n = 5$; see **Figure 3 B** and **Supplementary Figure 2** for radar plot profiles) in a prospective classification it is appropriate to allocate these scenarios to the mild category applying the precautionary principle.

The above score profile of the scenarios considered to be around the lower regulatory threshold differs markedly from all other supra-threshold scenarios (sub-threshold score $5.9 \pm 1.1\%$ vs. sum of supra-threshold scores $87.7 \pm 1.2\%$, n=40; p < 0.0001).

The remaining scenarios identified above (8, 20 and 27) cluster with scenarios in C1_A** (**Figure 2**) which are clearly above the lower threshold for regulation, and are therefore considered to describe procedures which should also be regulated under the Directive.

Identification of scenarios in mild, moderate and severe classifications and around borderlines

As depicted in **Figure 2** some scenarios have been allocated to the mild and higher severity classes (see also score profiles in **Supplementary Figures 2 and 3**). Below we also identify scenarios attributed by respondents to borderline categories. These scenarios are allocated to the higher severity classification class according to the precautionary principle which should be applied to prospective assessment of severity.

Mild - Scenarios in this category are identified in the groupings B2 and C1_A** in the clustering (**Figure 2**). The scenarios (3, 12, 23, 34 and 40) on the sub-threshold/mild borderline in B2 (**Figure 2; Supplementary Figure 3**) should ultimately be included with scenarios identified above in the mild category (C1_A**); sub-cluster including scenarios: 8, 11, 13, 20, 27) using the precautionary principle (see for summary **Figure 4**).

Sub-cluster C1_A* adjacent to C1_A**, both included in the main cluster C1_A, identifies four scenarios (1, 7, 17 and 48) with a predominant score in the mild category but also with scores in higher, but not lower, severity categories. The

overall scenario scores for the five scenarios in the C1_A** sub-cluster (8, 11, 13, 20 and 27) are lower than those in the C1_A* cluster (1, 7, 17 and 48; see also **Appendix 3** for individual scores).

Scenarios 1, 7, 17 and 48 have relatively high mild scores and scores in higher severity categories placing them on the mild/moderate borderline (**Supplementary Figure 4**). For these latter scenarios the individual survey scores in the mild category are similar to those in the moderate and higher categories (mild $43.3 \pm 2.2\%$ vs. higher $44.7 \pm 1.8\%$; see **Supplementary Figure 3**), so using the principle outlined above they should be allocated to the moderate category (**Figure 4**).

Moderate. Cluster C1_B (**Figure 2**) identifies scenarios with high scores in the moderate category (C1_B*) and scenarios with moderate but also higher severity scores (C1_B**). The overall scores (see Methods) for the fifteen scenarios in the C1_B* sub-cluster (4, 10, 15, 16, 18, 19, 21, 24, 31, 33, 35, 36, 37, 45 and 49) are lower than the six scenarios (6, 25, 39, 44, 47 and 50) in the C1_B** sub-cluster (662 ± 8.7 vs. 757 ± 12.8 ; $p < 0.0001$).

The %scores in each severity category for scenarios in the moderate category and for scenarios with high moderate scores, but also with scores in more severe categories, are shown in **Supplementary Figure 3**. Analysis of the individual %scores in the C1_B** sub-cluster (6, 25, 39, 44, 47, 50) shows that these six scenarios are at the upper-end of the moderate category, hence they are in cluster C1 as opposed to cluster C2 (severe and upper threshold). However, based upon

the distribution of scores and adopting the precautionary principle these should be classified as severe (see below).

Applying the precautionary principle, the scenarios identified above on the mild/moderate borderline (sub-cluster C_{1A}*), i.e. scenarios 1, 7, 17, 48) should be included with the scenarios identified above in the moderate category (C_{2B} sub-cluster: 4, 10, 15, 16, 18, 19, 21, 24, 31, 33, 35, 36, 37, 45 and 49; see summary **Figure 4**).

Severe and upper threshold. All remaining scenarios were allocated by cluster analysis to grouping C2 (**Figure 2**) with sub-clusters C_{2A} (scenarios 9, 14, 28, 30, 41) and C_{2B} (scenarios 26, 43). Scenarios 26 and 43 (C_{2B}) have high and comparable scores in the severe and upper threshold categories so these are classified as upper threshold while scenarios in sub-clusters C_{2A} (9 and 14) and C_{1B} (6, 25, 39, 44, 47, 50; see **Moderate** above) are classified as severe (**Figure 4**).

Overview of scenario allocation to severity categories

The cluster analysis (**Figure 2**) identified groupings and sub-groupings of scenarios which were used to allocate them to sub-threshold (B1), mild (B2 and C_{1A}**), moderate (C_{1A}* and C_{1B}**), severe (C_{2A} and C_{1B}**) and above upper threshold (C_{2B}) categories.

The original % severity scores for all scenarios allocated to the severity categories, as defined in Directive 2010/63/EU, are plotted in **Figure 4** in which scenarios on borderlines between categories (see also **Supplementary Figure 2**) have been allocated to the higher severity category as described above and adopting the precautionary principle.

An unclassifiable scenario

Scenario 22 (“*Newly hatched paralarvae of Octopus vulgaris (<1mm, DML) will be killed by immersion in liquid nitrogen for subsequent biochemical analysis*”) falls into a category of its own based on two criteria. Firstly, it had the highest score of any scenario in the unable to decide category (25.9%), with a sub-threshold score of 25.9%, a non-recovery score of 20.7% and an aggregate score of 26.6% in the supra-threshold severity categories. These values give it a unique profile (see **Supplementary Figure 2**). Secondly, it stands alone in the dendrogram (**Figure 2**).

In contrast to all other scenarios no predominant category could be identified.

Examples of procedures in cephalopods derived from the survey analysis

Using the classification of the scenarios from the dendrogram (**Figure 2**), the severity allocation based on % category scores and their distribution (**Figure 4, Supplementary Figure 2**), the scenarios used in the survey are listed in rank order, with the assigned severity classification indicated, to provide examples of the prospectively assessed severity of procedures in cephalopods (**Appendix 3**).

The main procedures have been also extracted from the full scenario descriptions to provide concise examples in each severity category (**Table 3**).

Discussion

This paper reports the first attempt to allocate regulated procedures in cephalopods prospectively to severity categories. An integrated cephalopod community, with people

working either directly or indirectly in cephalopod science, as represented by the participants in COST Action FA1301 ‘CephsInAction’, allocated consistently scenarios *i*, describing procedures below and above the lower threshold for regulation defined by Directive 2010/63/EU, and *ii*, across the entire range of severity classifications. Furthermore, respondents recognised that severity classification applies to all cephalopod life-stages post-hatching and species.

Critique of methodology

A discussion of the limitations of the Delphi method is outside the scope of this paper but here we overview some of the issues that may have influenced the outcomes of the survey.

Potentially, a survey can gather views from a larger and more diverse group than may be possible in a round-table discussion, and can include those either unable to attend or not invited to a “self-selected” expert group meeting. The survey views are independent (assuming no collusion), equivalent and anonymous, minimising the impact of group dynamics and responses can readily be linked to personal profiles to assess variations between sub-groups. In the survey we included an “unable to decide” option whereas in a round-table meeting of “experts” peer pressure may make it be difficult for an individual to give this as their opinion. With an on-line survey the time pressure to reach a decision is reduced compared to a meeting and modification of response is possible until submission.

In our case, the PAS-C survey facilitates engagement of the cephalopod community in responding to the Directive 2010/63/EU as the opinions are representative of the community rather than views imposed by those more familiar with this type of

assessment in vertebrates. This is of particular relevance since it has been suggested that the Directive attempts to apply a “mammalocentric” approach to cephalopods.²¹ However, there are also negative factors to consider.

The PAS-C Survey was based on a selected series of short scenarios to focus attention on the procedures used, and avoided including any justification so the judgement was based on the scenario procedures and not the scientific or other justification. Scenarios included in the survey were based on data derived from a non-systematic review of published works spanning from 1940 to 2014 that utilized cephalopods in studies including biochemistry, physiology, immunology, neuroscience, eco-toxicology, and behavioural sciences. The use of methods sections from published works publications was not considered practical for our aims because of: *i.* the need for extensive copyright clearance, *ii.* respondents recognising the original paper, and hence being influenced by the authorship, *iii.* the length of many methods sections would make the time required to complete the survey impractical. We are aware that the amount of information given in the scenarios is limited; however, the outcomes indicate that this was sufficient to reach a decision as overall scores in the “unable to decide” category were low.

The survey outcomes are obviously dependent on the respondents. However, the respondents are a self-selecting sample of a population (the COST Action FA1301 participants) whom it is assumed are a representative sample of the EU cephalopod research community. We recognise that the respondents comprise a number of sub-groups (**Supplementary Tables 1 and 2**) which it was not possible to analyse separately because of the sub-group sizes. However, we used a median split for the main sub-groups (age, gender, PhD, cephalopod experience and publications) and this did not reveal any major differences in overall assessment of the scenarios but this analysis may not have sufficient resolution to identify more

subtle differences. The answers given by the respondents assume diligence (e.g., reading the Guidance notes) and honesty (e.g., giving an independent view) with participants needing to be sufficiently altruistic to spend about a couple of hours to complete the survey. By not selecting only “experts” to participate, the responses will depend on the respondents’ knowledge and understanding of the Directive, knowledge of some basic biology relevant to each cephalopod species, comprehending the experimental methods in scenarios, their level of relevant expertise and possibly their “emotional” response to the scenario (e.g., descriptions of surgical procedures may be challenging for researchers who may only be involved in behavioural studies). In deriving the overall severity assignments, objective criteria have been used based on the survey responses, but the process by which individual respondents reached their assessments is likely to include both objective and subjective elements.

Whatever the approach used by the respondents there is a high degree of consensus for many scenarios, and in addition we were unable to identify any major systematic differences in the overall assessments between the demographic sub-groups.

Finally, we should highlight that the survey was only delivered in English, but COST Action FA1301 has participants from many EU states and English is not the first language for many, although meetings are conducted in English. We are unable to assess the impact of language on the results but it may have reduced the number of respondents.

Overall classification and comparison of scenarios: Internal validity

To identify the major factors influencing the responses we have focused on the scenarios involving either *S. officinalis* (12 scenarios) or *O. vulgaris* (16 scenarios) as they provide the largest number of scenarios for direct comparison.

i. General anaesthesia and/or surgery

All scenarios involving anaesthesia with recovery, irrespective of any other intervention were assessed as above the lower threshold for regulation in both species. This was also the case for comparable scenarios involving the less studied species such as *Eledone cirrhosa* and *Hapalochlaena lunata*. For example, scenario 40 described a feeding/growth study over the first month of life in cuttlefish which was assessed at the lower end of the mild category and considered by a high percentage of respondents to be sub-threshold (see **Figure 3 B**). However, addition of weekly, brief, general anaesthesia to this scenario (scenario 20) moved the assessment to the upper-end of the mild category (see **Figure 3 C**) and increased the overall scenario score (378 vs. 489). Similarly, a scenario (8) involving brief general anaesthesia combined with non-invasive ultrasound is in the middle of the mild category, but a similar procedure involving surgery with recovery and multiple periods of anaesthesia for ultrasound (scenario 1) increased the overall scenario score (470 vs. 581) and placed the scenario in the lower half of the moderate category.

When scenarios included more major surgical procedures they were allocated to moderate (e.g., data logger implantation, scenario 15) and higher categories (e.g., partial arm amputation, scenario 24; supra-oesophageal brain removal, scenario 30).

We deliberately included four scenarios in *O. vulgaris* describing arm amputation because of the resurgence of interest in regeneration studies.²²⁻²⁴ Although there was little difference in the overall severity classification between the scenarios describing amputation carried out under general anaesthesia, with or without local anaesthesia, the scenario describing 10% amputation without anaesthesia (scenario 9) was clearly in

the severe category with an overall severity score in the upper decile. Interestingly, a comparable scenario (scenario 45) involving removal of 2mm of the arm from the Dumbo octopus for molecular profiling (analogous to rodent tail tip sampling) was clearly assigned to the moderate category whereas induction of arm autotomy in the octopus *Abdopus aculeatus* by crushing without anaesthesia was assigned to the severe category (scenario 39).

ii. Substance administration or haemolymph sampling

Repeated sampling of haemolymph under general anaesthesia over 4 weeks (scenario 17) in octopus was scored in the mild/moderate borderline so was finally allocated to the moderate category. Multiple drug injections (subcutaneous or intramuscular) to investigate analgesic (scenario 37) and anxiolytic (scenario 21) effects in cuttlefish were placed directly in the moderate category. The latter scenario (21) is identical to scenario 7, with the exception that scenario 21 involves multiple drug injections; scenario 7 is in the mild/moderate sub-cluster ($C1_A^*$ in Fig. 2) whereas scenario 21 is in the moderate sub-cluster ($C1_B^*$ in Fig. 2) and the overall scenario scores differ by 85 points (564 vs. 649).

Gavage for drug administration in squid (scenario 4) was judged to be moderate probably because of the risk of brain damage during the passage of the tube.^{25, 26}

iii. Behavioural interventions

Behavioural studies that may potentially cause distress above the threshold for regulation were recognised as falling within the Directive. Scenarios 13 and 49 are identical in describing exposure of juvenile cuttlefish to a predator in a divided tank.

However, the scenario in which the cuttlefish are able to hide and where there is no water exchange was categorised in the sub-threshold/mild group whereas, if there was no refuge and the partition perforated to allow water exchange, the scenario was assessed as moderate. Interestingly, a scenario (3) describing exposure of newly hatched *S. officinalis* to images of predators was also considered to be on the sub-threshold/mild borderline.

We propose that if the cephalopod is able to hide from the predator and there is no physical or chemical contact with the predator then the procedure should be considered to be sub-threshold. However, if such studies are undertaken the actual severity should be monitored to provide objective data to support this prospective assessment.

In cuttlefish, experimental induction of escape reactions (scenario 7) and investigation of agonistic social interactions in a small tank (scenario 18) were assessed as mild and moderate respectively. The latter is consistent with a classification of moderate severity for an agonistic social interaction study in *O. vulgaris* (scenario 16).

Scenarios (19 and 37) describing a training paradigm of *O. vulgaris* involving brief electric shocks as negative reinforcements were assessed as moderate, but the scenario which additionally involved daily injections (scenario 37) was also assessed as moderate, although at the upper end of the score range. A recently published study of serial visual reversal learning in octopus using only positive reinforcement reported that the study was classified as “mild”²⁷ although it is unclear whether this was prospective or actual severity.

iv. Food deprivation

Deprivation of food in a juvenile/adult animal within the duration likely to be normally encountered in the wild was not considered to fall within regulation (e.g., scenario 32) but extended food deprivation was. For example, 7 days food deprivation in adult *S. officinalis* was considered to fall within the moderate category (scenario 36) and this aligns with recent proposals for the severity classification of different periods of food deprivation in cephalopods.²⁵

v. A comment about scenarios allocated to the Upper Threshold classification

Two scenarios (26 and 43) were allocated to the upper threshold category (i.e. higher than severe, and hence should not be authorised except under “exceptional and scientifically justifiable reasons”, *Article 55*, see also **Table 2**). These scenarios were clustered (C2_B in Figure 2) with scenarios allocated directly to the severe category (C2_A in Figure 2) suggesting that it may be more accurate to view these scenarios at the upper end of the severe category. It is notable that one of the scenarios (26) describes a method for killing cephalopods which does not comply with the general principles outlined in *Annexe IV* of Directive 2010/63/EU and this is discussed in more detail in a specific section below.

An LD₅₀ study in adult cuttlefish (scenario 43) was also considered to be upper threshold (i.e., above the upper limit defined in *Article 15*) and had the highest overall scenario score. As described, such a study should not be authorised but with appropriate refinement and well defined humane end-points the prospective assessment could change to severe; the classification allocated by 35.6% of respondents.

Prospective severity classification of procedures in cephalopods

Scenarios around the regulatory threshold. The decision regarding whether a procedure is likely to exceed the threshold for regulation is a critical one, but the regulatory boundary may not be obvious. The progressive shift in the score profile from sub-threshold to mild categories is illustrated in **Figure 3** (see also **Supplementary Figure 3**). Applying the precautionary principle we allocated scenarios on the borderline to the mild category. For example, scenario 40 - describing the effect of artificial diets on growth of hatchling cuttlefish - was considered to be mild. However, it is likely that assessment of actual severity would show that the study described was sub-threshold. This precautionary approach is further supported by the Annual Statistics of Scientific Procedures on Living Animals for the Great Britain²⁸ which reported in 2016 that 12% of procedures performed in the UK were sub-threshold although the prospective classification and application to the National Competent Authority must have indicated that they were considered above the regulatory threshold.

Scenarios most likely to be considered below the lower threshold for regulation (sub-threshold) by a significant minority of the respondents involved either behavioural or dietary manipulation/growth studies. As behaviour is among the most studied aspects of cephalopod biology²⁹⁻³² and metabolic studies are also important for research informing cephalopod aquaculture^{25, 26, 33, 34} we recommend that specific guidance is developed to assist researchers in assessing severity in such studies. However, it should be emphasised that if there is any doubt about whether a proposed procedure is below or above the lower threshold for regulation, the default position is that the procedure

should be regulated and evidence gathered on its' actual severity to inform future applications.

Non-recovery scenarios. Three scenarios fitting the Directive definition of non-recovery were readily identified. However, three other scenarios had relatively high scores (13-20%) in the non-recovery category, although not fitting the definition. These scenarios described a method for killing a cephalopod with no other intervention. This observation suggests that although the Directive 2010/63/EU definition of non-recovery⁴ is precise (see **Table 2**) some additional guidance notes to avoid confusion with killing methods may be required.

Scenarios describing methods for killing cephalopods.

Killing an animal is not a regulated procedure under Directive 2010/63/EU provided that the general principles outlined in *Article 6* and *Annexe IV §1* and *2* are followed and a method described in the Table in *Annexe IV §3* is employed. Alternatively, under *Article 6 §4a,b* the National Competent Authority (NCA) may grant an exemption based on an assessment that the proposed method is at least as humane as those in *Annexe IV §3* or that based on a scientific justification the purpose of the procedure cannot be achieved by methods given in *Annexe IV*. However, the Table in *Annexe IV §3* does not give any method for killing cephalopods and in the recent European Commission review of the objectives, fitness for purpose and currency of Directive 2010/63/EU^{††} the lack of standardised appropriate methods for killing cephalopods was recognised as an omission. The only information relevant to killing cephalopods is in *Annexe IV §1a* and

^{††} <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52017DC0631&from=EN>

2 which specifies that methods other than those in the Table in *Annexe IV* may be used only if the animal is unconscious and killing is completed by one of five methods; proposals for methods compliant with these principles have been published^{2,35}.

However, unconsciousness, particularly if induced by a general anaesthetic or cranial trauma may not be compatible with the scientific objectives of a project and an inspection of the cephalopod literature identified several methods which would not be compliant with the principles in *Annexe IV* §1a if used now. Four scenarios describing a killing method were included in the survey to obtain the views of the COST FA1301 community on the prospective severity of these methods should a researcher apply for authorisation (or if an exemption under *Article 6* §4a,b is sought) to use them. Scenarios describing decapitation of an adult squid (28) and mechanical destruction of the brain using a miniature captive-bolt in a 1.5Kg octopus (41) were classified as “severe”. It is interesting to note that the Table in *Annexe IV* §3 identifies decapitation as acceptable for birds (<250g) and rodents (if other methods not possible) and captive bolt as acceptable for larger reptiles (size not specified), rabbits and large mammals.

The scenario describing killing a hatchling cuttlefish by immersion in formalin without prior anaesthesia was considered to be upper-threshold and hence should not be authorised (see above and **Table 2** for definition); a view consistent with previous recommendations.^{2,35}.

A final scenario describing killing newly hatched octopus paralarvae (<1mm, DML) by immersion in liquid nitrogen was the only one of the 50 scenarios to be impossible to classify. Of particular note is that ~25% of respondents considered this to be sub-threshold but this scenario also had the highest unable to decide score.

The results of the survey show that amending *Annexe IV* to give appropriate methods, not involving prior loss of consciousness^{2, 35}, for killing cephalopods represents an urgent challenge for both legislators and the EU based cephalopod community.

Special considerations for the classification of severity of procedures when using cephalopods

There are over 800 living cephalopod species^{5-7, 36} all marine, living in a wide range of habitats from coastal waters to deep-sea. Despite being limited in number, when compared with the very populous phylum Mollusca³⁷ to which they belong, cephalopods are an astonishing example of diversity of forms and functions. Cephalopods have a molluscan *Bauplan*, but illustrate how evolution can generate great complexity, including sophisticated behavioural repertoire and neural plasticity and organization.^{29, 38, 39} Thus, as Hawkins and co-authors noted for fish,⁴⁰ considering cephalopods as a single ‘unit’ has little meaning as would also be the case if one attempted to produce severity guidance for all mammals. The impact of a given procedure on an animal may differ depending on the species to which it belongs, so prospective assessment of severity should take full account of species-specific characteristics. For example, the impact of restraint may vary widely between the shelled *Nautilus*, cuttlefish, squid and octopus because of their markedly different levels of locomotor activity. Removal of a cephalopod from water may induce stress responses,^{41, 42} and therefore may increase the severity of a procedure but the response to hypoxia differs considerably between species (e.g. *Nautilus*,⁴³ cuttlefish,⁴⁴ squid,⁴⁵ and octopus⁴⁶) so the impact of a period in air (for example during weighing, injection or gavage) needs to be assessed for each species. An additional example is provided by the effects of food deprivation on welfare

(reviewed by Sykes et al.²⁵) where assessment of the impact requires an understanding of the diverse nature of the natural diets and feeding habits of cephalopods.⁴⁷

For cephalopods, as with other regulated species, when assessing severity classification the duration of a procedure or its repetition must also be considered in the context of the biology of the species under study; this may move a given procedure from one severity classification to another.⁴

In addition to the large interspecific diversity, many cephalopods are known for significant changes in their physiology as part of their life-history.⁴⁸ Currently, species-specific guidelines on care and use of cephalopods for experimental purposes are missing. However, for example Fiorito and co-authors² under the aegis of FELASA, the Boyd Group and the COST Action FA1301, provided the first general guidelines on the care and welfare of these animals when used for scientific research. In the guidelines, a series of indicators are recommended to regularly assess animals' health and welfare and identify factors that may cause deviations from optimal status.² Key parameters are proposed to evaluate animal behavior and appearance, and these are recommended to be supplemented by measurement of a number of physiological 'biomarkers'. Fiorito and coworkers² also identified possible signs, proposing a gradation as an indication of an escalation from mild to moderate and severe responses/conditions. These include: *i.* appearance (Physical state, 7 indicators) including skin colour, texture and its integrity, abnormal body appearance (e.g., arms unaligned or dangling); *ii.* behaviour (Psychological state, 12 indicators) including unprovoked behaviors (e.g., withdrawal, abnormal motor or locomotor coordination, grooming, wound directed behaviour), responses to food, and provoked behaviours (defaecation, inking, etc.); *iii.* clinical (Physiological/biochemical state, 8 indicators) such as reduction in body weight and

changes in the rates of ventilation or heart beat frequency. As recommended these indicators should be observed and recorded daily.²

Evidence is emerging for inter-specific differences in the response to noxious stimuli which, if confirmed and extended, will impact severity assessment across cephalopod species.⁴⁹⁻⁵⁴ Techniques to manage pain in cephalopods, apart from local and general anaesthesia^{2, 41}, are not established, and therefore there is an urgent need to develop methods to both assess and alleviate pain for the commonly used species (about 30 according to Smith et al. ¹¹). Finally, there is evidence that cephalopod species may differ in their reaction to the same anaesthetic agent^{2, 55, 56} further emphasising the need to assess the impact of a procedures on a species by species basis.

The further development and validation of ‘welfare indicators’ will ensure accurate assessment of actual severity of procedures and objective measurement of the impact of interventions to alleviate pain, suffering and lasting harm. Data on actual assessment of procedures will in turn enable more accurate assessment of the prospective severity of procedures.

A final special consideration for cephalopods is that animals for research may be taken from the wild (e.g., *Nautilus pompilius*, *Loligo vulgaris*, *Octopus vulgaris*) as described in Article 10 of Directive 2010/63/EU. In all the survey scenarios, capture and transport were not considered as part of the project so their impact does not contribute to the overall severity assessment. However, if a study was undertaken in which capture and transport were part of the scientific objectives of the project then their impact would need to be included in the overall assessment of any procedure performed on arrival in the laboratory.⁴ For example, if a study compared different methods of capture from

the wild on the response to a battery of behavioural tests in the laboratory then the overall severity assessment would be based on the methods of capture, transport and the behavioural tests.

Concluding comments

The results of this COST Action FA1301 PAS-C Survey provide a framework to assist cephalopod researchers in assessing the prospective severity of procedures used in research and in guiding their application to the National Competent Authorities.

The *prospective* classifications of the procedures proposed here (**Tables 3 and 4**) should be regarded as a starting point requiring regular review and modification as information on *actual* severity assessment becomes available for procedures performed on cephalopods under Directive 2010/63/EU. For vertebrates in general, and mammals in particular, there is a body of experience (e.g., regulators, ethical review committees, veterinarians, animal technologists) that can be accessed by project applicants to guide severity assessment and this is facilitated by a number of publications on the topic (e.g. for fish¹⁵). However, for cephalopods there is no history of regulation which can be consulted although this will change with time as Directive 2010/63/EU takes effect and regulated studies are published. The principles of severity classification applicable to cephalopods were reviewed in preparation for regulation,^{11, 12, 35} but these now need to be developed further into guidelines using the evidence gathered here.

The experience of the cephalopod community in responding over the last five years to the multiple challenges resulting from the EU-wide regulation of an entire invertebrate class provides a model should regulation be extended to other invertebrates (e.g.,

decapod crustacea)^{‡‡}.^{2, 12, 57} The EU-wide research community approach using web-based assessment of scenarios that we have taken here has wider applicability to obtaining views about severity assessment in other research areas.

Acknowledgements

We thank all respondents and colleagues who provided comments in the pilot stage of the survey, particularly Professor A. Olsson (Portugal) and Dr. J. Lenehan (Irish Republic). This study was performed under the auspices of COST Action FA1301, benefited from its networking activities, and is considered a contribution to the COST (European COoperation on Science and Technology) Action FA1301 “A network for improvement of cephalopod welfare and husbandry in research, aquaculture and fisheries” (<http://www.cephsinaction.org/>). The authors wish to thank Stefania Grella for assistance in web-enabling the survey and preliminary extraction of data from the survey responses. We also wish to thank all those who responded to the survey and sent comments as well as colleagues who asked insightful questions at various preliminary presentations of the data and which were considered in drafting the manuscript. Dr T. Shaw is thanked for assistance with the initial dendrogram analysis.

Disclaimer. The opinions expressed in this paper should not be taken to reflect either the opinions or policies of any of the institutions or organisations to which the authors or those acknowledged belong. Decisions on whether a procedure falls within Directive

^{‡‡} see Norwegian Regulation on the use of animals in research that is currently protecting living vertebrates, decapod crustaceans and cephalopods (<https://lovdata.no/dokument/SF/forskrift/2015-06-18-761>)

2010/63/EU can only be made by the National Competent Authority and national legislature.

Conflict of Interest. The authors declare no conflicts of interest.

References

1. European Parliament and Council of the European Union. *Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes*. Strasbourg: Concil of Europe, 2010.
2. Fiorito G, Affuso A, Basil J, Cole A, de Girolamo P, D'Angelo L, Dickel L, Gestal C, Grasso F, Kuba M, Mark F, Melillo D, Osorio D, Perkins K, Ponte G, Shashar N, Smith D, Smith J and Andrews PL. Guidelines for the Care and Welfare of Cephalopods in Research - A consensus based on an initiative by CephRes, FELASA and the Boyd Group. *Lab Anim* 2015; 49: 1-90.
3. Sneddon LU, Halsey LG and Bury NR. Considering aspects of the 3Rs principles within experimental animal biology. *The Journal of Experimental Biology* 2017; 220: 3007-3016. DOI: 10.1242/jeb.147058.
4. Smith D, Anderson D, Degryse A-D, Bol C, Criado A, Ferrara A, Franco NH, Gyertyan I, Orellana JM and Ostergaard G. Classification and reporting of severity experienced by animals used in scientific procedures: FELASA/ECLAM/ESLAV Working Group report. *Laboratory Animals* 2018; 52: 5-57.
5. Jereb P and Roper C. *Chambered Nautiluses and Sepioids (Nautilidae, Sepiidae, Sepiolidae, Sepiadariidae, Idiosepiidae and Spirulidae)*. Rome, Italy: FAO, 2005, p.262.
6. Jereb P and Roper C. *Cephalopods of the world. An annotated and illustrated catalogue of species known to date. Volume 2. Myopsid and Oegopsid Squids*. Rome, Italy: FAO, 2010, p.605.

7. Jereb P, Roper C, Norman M and Finn J. *Cephalopods of the World. An Annotated and Illustrated Catalogue of Species Known to Date. Volume 3. Octopods and Vampire Squids.* 2016, p.370.
8. Boletzky Sv. Biology of early life stages in cephalopod molluscs. *Advances in Marine Biology* 2003; 44: 143-203.
9. Sykes AV, Baptista FD, Gonçalves RA and Andrade JP. Directive 2010/63/EU on animal welfare: a review on the existing scientific knowledge and implications in cephalopod aquaculture research. *Reviews in Aquaculture* 2012; 4: 142-162.
10. Young RE and Harman RF. "Larva," "Paralarva" and "Subadult" in Cephalopod Terminology. *Malacologia* 1988; 29: 201-207.
11. Smith JA, Andrews PL, Hawkins P, Louhimies S, Ponte G and Dickel L. Cephalopod research and EU Directive 2010/63/EU: Requirements, impacts and ethical review. *Journal of Experimental Marine Biology and Ecology* 2013; 447: 31-45.
12. Fiorito G, Affuso A, Anderson DB, Basil J, Bonnaud L, Botta G, Cole A, D'Angelo L, De Girolamo P, Dennison N, Dickel L, Di Cosmo A, Di Cristo C, Gestal C, Fonseca R, Grasso F, Kristiansen T, Kuba M, Maffucci F, Manciocco A, Mark FK, Melillo D, Osorio D, Palumbo A, Perkins K, Ponte G, Raspa M, Shashar N, Smith J, Smith D, Sykes A, Villanueva R, Tublitz N, Zullo L and Andrews PLR. Cephalopods in neuroscience: Regulations, Research and the 3Rs. *Invert Neurosci* 2014; 14: 13-36.

13. Commission E. *Examples to illustrate the process of severity classification, day-to-day assessment and actual severity assessment*. Brussels: European Commission, 2013.
14. Commission E and 2010/63/EU NCAftIoD. *Working document on a severity assessment framework*. Brussels: European Commission, 2012.
15. Hawkins P, Dennison N, Goodman G, Hetherington S, Llywelyn-Jones S, Ryder K and Smith AJ. Guidance on the severity classification of scientific procedures involving fish: report of a Working Group appointed by the Norwegian Consensus-Platform for the Replacement, Reduction and Refinement of animal experiments (Norecopa). *Lab Anim-Uk* 2011; 45: 219-224.
16. Dalkey N and Helmer O. An experimental application of the Delphi method to the use of experts. *Management Science* 1963; 9: 458-467.
17. Okoli C and Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Information & Management* 2004; 42: 15-29.
DOI: <https://doi.org/10.1016/j.im.2003.11.002>.
18. Commission E. *Implementation, interpretation and terminology of Directive 2010/63/EU. Questions & Answers on the legal understanding - Q&A document*. Brussels: European Commission, 2014.
19. Sarstedt M and Mooi E. Cluster Analysis. In: Sarstedt M and Mooi E (eds) *A Concise Guide to Market Research*. Berlin, Heidelberg: Springer, 2014, pp.273-324.

20. Yim O and Ramdeen KT. Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. *The Quantitative Methods for Psychology* 2015; 11: 8-21. DOI: 10.20982/tqmp.11.1.p008.
21. Nosengo N. European directive gets its tentacles into octopus research. *Nature News* 2011; 12 April: doi:10.1038/news.2011.1229.
22. Imperadore P, Shah SB, Makarenkova HP and Fiorito G. Nerve degeneration and regeneration in the cephalopod mollusc *Octopus vulgaris*: the case of the pallial nerve. *Scientific Reports* 2017; 7: 46564.
23. Shaw TJ, Osborne M, Ponte G, Fiorito G and Andrews PL. Mechanisms of wound closure following acute arm injury in *Octopus vulgaris*. *Zoological Letters* 2016; 2: 8.
24. Zullo L, Fossati SM, Imperadore P and Nödl MT. Molecular determinants of Cephalopod muscles and their implication in muscle regeneration. *Frontiers in cell and developmental biology* 2017; 5.
25. Sykes AV, Almansa E, Cooke GM, Ponte G and Andrews PL. The Digestive Tract of Cephalopods: a Neglected Topic of Relevance to Animal Welfare in the Laboratory and Aquaculture. *Frontiers in Physiology* 2017; 8: 492.
26. Ponte G, Sykes AV, Cooke GM, Almansa E and Andrews PL. The digestive tract of cephalopods: toward non-invasive in vivo monitoring of its physiology. *Frontiers in Physiology* 2017; 8: 403.

27. Bublitz A, Weinhold SR, Strobel S, Dehnhardt G and Hanke FD. Reconsideration of Serial Visual Reversal Learning in Octopus (*Octopus vulgaris*) from a Methodological Perspective. *Frontiers in Physiology* 2017; 8: 54.
28. Home Office U. Annual Statistics of Scientific Procedures on Living Animals, Great Britain 2016, <https://www.gov.uk/government/statistics/statistics-of-scientific-procedures-on-living-animals-great-britain-2016> (2017).
29. Marini G, De Sio F, Ponte G and Fiorito G. Behavioral Analysis of Learning and Memory in Cephalopods. In: Byrne JH (ed) *Learning and Memory: A Comprehensive Reference (Second Edition)*. Amsterdam, The Netherlands: Academic Press, Elsevier, 2017, pp.441-462.
30. Huffard CL. Cephalopod neurobiology: an introduction for biologists working in other model systems. *Invertebrate Neuroscience* 2013; 13: 11-18.
31. O'Brien CE, Roumbelakis K and Winkelmann IE. The Current State of Cephalopod Science and Perspectives on the Most Critical Challenges Ahead From Three Early-Career Researchers. *Frontiers in Physiology* 2018; 9: 700. Review. DOI: 10.3389/fphys.2018.00700.
32. Nakajima R, Shigeno S, Zullo L, De Sio F and Schmidt MR. Cephalopods between science, art, and engineering: a contemporary synthesis. *Frontiers in Communication* 2018; 3: 20.
33. Iglesias J, Fuentes L and Villanueva R. *Cephalopod Culture*. Dordrecht: Springer Netherlands, 2014.

34. Garrett SC and Rosenthal JJC. A Role for A-to-I RNA Editing in Temperature Adaptation. *Physiology (Bethesda)* 2012; 27: 362-369.
35. Andrews PL, Darmillacq AS, Dennison N, Gleadall IG, Hawkins P, Messenger JB, Osorio D, Smith VJ and Smith JA. The identification and management of pain, suffering and distress in cephalopods, including anaesthesia, analgesia and humane killing. *Journal of Experimental Marine Biology and Ecology* 2013; 447: 46-64.
36. Sweeney MJ and Roper CF. Classification, type localities, and type repositories of recent Cephalopoda. *Smithsonian Contributions to Zoology* 1998: 561-599.
37. Ponder W and Lindberg DR. *Phylogeny and Evolution of the Mollusca*. Univ of California Press, 2008, p.469.
38. Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, Brenner S, Ragsdale CW and Rokhsar DS. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 2015; 524: 220-224.
39. Shigeno S, Andrews PLR, Ponte G and Fiorito G. Cephalopod Brains: An Overview of Current Knowledge to Facilitate Comparison With Vertebrates. *Frontiers in Physiology* 2018; 9. Review. DOI: 10.3389/fphys.2018.00952.
40. Hawkins P, Dennison N, Goodman G, Hetherington S, Llywelyn-Jones S, Ryder K and Smith AJ. Guidance on the severity classification of scientific procedures involving fish: report of a Working Group appointed by the Norwegian Consensus-Platform for the Replacement, Reduction and Refinement of animal experiments (Norecopa). *Laboratory Animals* 2011; 45: 219-224.

41. Malham SK, Coulson CL and Runham NW. Effects of repeated sampling on the haemocytes and haemolymph of *Eledone cirrhosa* (Lam.). *Comparative Biochemistry and Physiology A-Molecular & Integrative Physiology* 1998; 121: 431-440.
42. Malham SK, Lacoste A, Gelebart F, Cueff A and Poulet SA. A first insight into stress-induced neuroendocrine and immune changes in the octopus *Eledone cirrhosa*. *Aquatic Living Resources* 2002; 15: 187-192.
43. Boutilier R, West T, Webber D, Pogson G, Mesa K, Wells J and Wells M. The protective effects of hypoxia-induced hypometabolism in the *Nautilus*. *Journal of Comparative Physiology B* 2000; 170: 261-268.
44. Capaz JC, Tunnah L, MacCormack TJ, Lamarre SG, Sykes AV and Driedzic WR. Hypoxic induced decrease in oxygen consumption in cuttlefish (*Sepia officinalis*) is associated with minor increases in mantle octopine but no changes in markers of protein turnover. *Frontiers in Physiology* 2017; 8: 344.
45. O'Dor RK, Pörtner HO and Shadwick RE. Squid as elite athletes: Locomotory, respiratory and circulatory integration. In: Gilbert DL, Adelman WJ and Arnold JM (eds) *Squid as experimental animals*. New York, 1990, pp.481-503.
46. Smith PJS, Duthie GG, Wells MJ and Houlihan DF. Continuous recording of arterial blood pO₂ in *Octopus vulgaris* during hypoxia and movement. *Journal of Experimental Biology* 1985; 117: 475-479.

47. Villanueva R, Perricone V and Fiorito G. Cephalopods as predators: a short journey among behavioral flexibilities, adaptions, and feeding habits. *Frontiers in Physiology* 2017; 8: 598.
48. Nixon M and Mangold K. The early life of *Sepia officinalis*, and the contrast with that of *Octopus vulgaris* (Cephalopoda). *Journal of Zoology* 1998; 245: 407-421.
49. Crook RJ, Lewis T, Hanlon RT and Walters ET. Peripheral injury induces long-term sensitization of defensive responses to visual and tactile stimuli in the squid *Loligo pealeii*, Lesueur 1821. *The Journal of Experimental Biology* 2011; 214: 3173-3185.
50. Crook RJ, Hanlon RT and Walters ET. Squid have nociceptors that display widespread Long-Term Sensitization and spontaneous activity after bodily injury. *Journal of Neuroscience* 2013; 33: 10021-10026.
51. Alupay JS, Hadjisolomou SP and Crook RJ. Arm injury produces long-term behavioral and neural hypersensitivity in octopus. *Neuroscience Letters* 2013; 558: 137-142.
52. Cartron L, Darmillacq A-S and Dickel L. The "prawn-in-the-tube" procedure: What do cuttlefish learn and memorize? *Behavioural Brain Research* 2013; 240: 29-32.
53. Darmillacq AS, Dickel L, Chichery MP, Agin V and Chichery R. Rapid taste aversion learning in adult cuttlefish, *Sepia officinalis*. *Animal Behaviour* 2004; 68: 1291-1298.

54. Borrelli L. *Testing the contribution of relative brain size and learning capabilities on the evolution of Octopus vulgaris and other cephalopods*. PhD Thesis, Stazione Zoologica Anton Dohrn, Italy & Open University, UK, 2007.
55. Gleadall IG. The effects of prospective anaesthetic substances on cephalopods: summary of original data and a brief review of studies over the last two decades. *J Exp Mar Biol Ecol* 2013; 447: 23-30.
56. Butler-Struben HM, Brophy SM, Johnson NA and Crook RJ. In vivo recording of neural and behavioral correlates of anesthesia induction, reversal, and euthanasia in cephalopod molluscs. *Frontiers in Physiology* 2018; 9: 109.
57. Berry A, Vitale A, Carere C and Alleva E. EU guidelines for the care and welfare of an "exceptional invertebrate class" in scientific research. *Annali dell'Istituto Superiore di Sanità* 2015; 51: 267-269.
58. Sweeney MJ and Roper CFE. Classification, type localities, and type repositories of recent Cephalopoda. In: Voss GL and al. e (eds) *Systematics and Biogeography of Cephalopods*. Washington, DC: Smithsonian Institution, 1998, pp.561-599.
59. Sole M, Lenoir M, Durfort M, Lopez-Bejar M, Lombarte A, van der Schaal M and Andre M. Does exposure to noise from human activities compromise sensory information from cephalopod statocysts? *Deep-Sea Research Part II-Topical Studies in Oceanography* 2013; 95: 160-181.

60. Sole M, Lenoir M, Durfort M, Lopez-Bejar M, Lombarte A and Andre M.

Ultrastructural Damage of *Loligo vulgaris* and *Illex coindetii* statocysts after Low Frequency Sound Exposure. *PLoS ONE* 2013; 8.

Table 1. List of cephalopod species and relative life-stages included in the 50 scenarios of the COST Action FA1301 PAS-C Survey. For each species, the total number of scenarios (and their survey scenario number, in italics) where the species has been included and the relative life-stage (egg, paralarva, hatchling, juvenile and adult) is reported. The species are listed in taxonomic order following Sweeney and Roper.⁵⁸

Species	Life Stages					Total number of scenarios
	Eggs	Paralarvae	Hatchlings	Juveniles	Adults	
<i>Nautilus pompilius</i>					1 <i>11</i>	1
<i>Sepia officinalis</i>	1 <i>3</i>	3 <i>20*, 26, 40*</i>	8 <i>1, 5, 7, 13, 20*, 21, 40*, 49</i>	12, 15, 18, 34, <i>36, 4*, 46</i>	7	16+1*
<i>Loligo vulgaris</i>	1 <i>14</i>				2 <i>28, 33</i>	3
<i>Doryteuthis pealeii</i>					2 <i>4, 6</i>	2
<i>Watasenia scintillans</i>	1 <i>27</i>					1
<i>Grimpoteuthis</i> spp.					1 <i>45</i>	1
<i>Argonauta argo</i>					1 <i>42‡</i>	1
<i>Octopus vulgaris</i>	3 <i>22, 31, 50</i>				15 <i>8, 9, 10, 16, 19, 24, 25, 30, 32, 35, 37, 38, 41, 47, 48</i>	18
<i>Octopus hummeli</i>					1 <i>23</i>	1
<i>Abdopus aculeatus</i>					1 <i>39</i>	1
<i>Haplochlaena lunulata</i>					1 <i>44</i>	1
<i>Eledone cirrhosa</i>					3 <i>2, 17, 29</i>	3

* cuttlefish species non specified

† Same scenario, animals were analysed from hatchlings to juveniles

‡ Life stage non specified, presumably adult

Table 2. Definitions of severity classifications adopted in the COST Action FA1301 PAS-C survey (see also Supplementary Information).^{4, 13, 14, 18}

SEVERITY CLASSIFICATION	DEFINITION
SUB-THRESHOLD	The use of an animal which does not meet the Directive threshold for regulation of a procedure which may cause the animal pain, suffering, distress or lasting harm equivalent to or higher than that caused by the introduction of a needle in accordance with good veterinary practice.
NON-RECOVERY	Procedures which are performed entirely under general anaesthesia from which the animal shall not recover consciousness.
MILD	Procedures on animals as a result of which the animals are likely to experience short-term mild pain, suffering or distress, as well as procedures with no significant impairment of the well-being or general condition of the animals shall be classified as ‘mild’.
MODERATE	Procedures on animals as a result of which the animals are likely to experience short term moderate pain, suffering or distress, or long-lasting mild pain, suffering or distress as well as procedures that are likely to cause moderate impairment of the well-being or general condition of the animals shall be classified as ‘moderate’.
SEVERE	Procedures on animals as a result of which the animals are likely to experience severe pain, suffering or distress, or long-lasting moderate pain, suffering or distress as well as procedures, that are likely to cause severe impairment of the well-being or general condition of the animals shall be classified as ‘severe’. Note that the Competent Authority will require retrospective assessment of projects involving “severe” procedures.
UPPER THRESHOLD	Although not specifically a severity level it is described as a level of severity greater than that described by the definition of “severe”. It is defined as a “ <i>procedure involving severe pain, suffering or distress that is likely to be long-lasting and cannot be ameliorated</i> .” Article 55 indicates that a Member State may allow such procedures under “exceptional and scientifically justifiable reasons”.

Table 3. Outline examples of procedures utilizing living cephalopods classified by severity (prospective) category: sub-threshold (green), mild (yellow), moderate (orange), severe (red). Allocation to a particular category utilises the analysis of the survey data (see text for details). The original scenario numbers from the survey are also reported; the full scenarios (see ‘Scenarios’ in Supplementary Information) should be consulted for details. Non-recovery scenarios (5, 29 and 46) and scenarios only describing a method for killing a cephalopod (22, 26, 28 and 41) are not included in this table but are discussed in text and are shown in full in the **Supplementary Files**.

Severity	Key procedures from scenario	Scenario no.
Sub-threshold	Testing behaviour in adult octopus using measurement of attack latency, alertness and response to a threat. Each test session lasts ~5min and is repeated 5 times/day over 5 days.	2
Sub-threshold	Food deprivation of an adult octopus for up to 24h in their home tank followed by feeding their normal diet including an inert marker and collection of faeces.	32
Sub-threshold	Comparative study of natural behaviours of singly housed octopus in tanks (compliant with Guidelines) with different environments (rocky vs. sandy bottom) mimicking two different natural environments.	38
Sub-threshold	Investigation of preference in adult octopus for a range of natural foods in comparison to an artificial diet that has been used successfully in a related species.	42
Mild	Comparison of growth rate in cuttlefish from one month of age using groups fed on the natural diet and comparator groups on three artificial diets that fulfil the nutritional needs. 3D video recording of behaviour and for measurement of growth.	40
Mild	Ultrasound imaging of the arm and brain in adult octopus under short duration (<30min) general anaesthesia with recovery on a single occasion.	8
Mild	Food deprivation for 5 days in <i>Nautilus</i> followed by administration of a barium sulphate marker in the food and measurement of oro-anal transit time using hourly X-rays over the next 12h followed by humane killing.	11
Mild	Exposure of juvenile cuttlefish to a predator for 60min but with complete physical separation and a refuge for the cuttlefish to hide.	13
Mild	Exposure of hatchling cuttlefish to images of predators in a tank with escape refuges and substrate.	3
Mild	Consecutive daily testing (5 days) of the predation/attack response of adult cuttlefish to a prawn placed in a tube so that it is inaccessible.	34
Mild	Comparison of growth rate in cuttlefish from one month of age using groups fed on the natural diet and comparator groups on three artificial diets that fulfil the nutritional needs. Weekly anaesthesia over 4 weeks	20

Severity	Key procedures from scenario	Scenario no.
	to permit detailed morphometry and cuttlebone density measurement using ultrasound.	
Mild	Taste aversion paradigm in adult cuttlefish using quinine treated crabs in adult cuttlefish.	12
Mild	Investigation of efficacy of novel synthetic diets in squid hatchlings over one month. Efficacy assessed by daily mortality but study terminated if >20% (pre-set humane end point).	27
Mild	Observation of post-reproductive female octopus until death by natural causes or euthanasia after loss of >20% body weight/cataracts/skin lesions.	23
Moderate	Removal of a small piece of anterior fin from juvenile cuttlefish under general anaesthesia followed by anaesthesia every five days for 30 days to permit microphotography and high resolution ultrasound.	1
Moderate	Weekly haemolymph sampling under general anaesthesia for a period of four weeks in an adult octopus.	17
Moderate	Administration of an experimental drug into the stomach of adult squid using a semi-rigid feeding tube inserted into the stomach via the beak with the animal manually restrained. Groups of animals will be killed humanely at 30min intervals over the next 12h to investigate drug pharmacokinetics.	4
Moderate	Exposure of juvenile cuttlefish to a predator for 60min but with a perforated partition between the tanks that allows water (but not animal) exchange between tanks. The cuttlefish tank does not contain a refuge.	49
Moderate	Implantation of data logging tags (<1% of body weight) into the cuttlebone of adult cuttlefish under general anaesthesia with recovery for one month.	15
Moderate	Training regime involving reward or aversive stimuli (brief noxious electric shock to an arm) in adult with a daily intramuscular injection of vehicle or study drug into an arm on each of 7 days of the study after which the animals are killed humanely.	37
Moderate	Sedation of newly hatched octopus paralarvae with progressively chilled sea water until movement and breathing stop, followed by substitution of the sea water with isotonic chilled formalin.	31
Moderate	Removal of a 2mm portion of arm tip from octopus without general anaesthesia.	45
Moderate	Food deprivation of adult cuttlefish for a period of seven days followed by humane killing.	36
Moderate	Investigation of the potential general anaesthetic effects of a substance not previously investigated in cephalopods using immersion of octopus in escalating concentrations.	48
Moderate	Investigation of opiate analgesia in octopus using subcutaneous administration of vehicle, opiate agonist or antagonist in a Latin square design with a week between test which involve noxious levels of mechanical stimulation to the arm and dorsal mantle.	47
Moderate	Investigation of potentially agonistic interactions between either sexually mature male and female octopus or cuttlefish with or without refuges.	16, 18
Moderate	Surgical removal of 10% of an arm in adult octopus under general anaesthesia with or without application of local anaesthetic.	25, 50

Severity	Key procedures from scenario	Scenario no.
Moderate	Investigation of the potential effects of an anxiolytic drug on escape behaviour in juvenile cuttlefish using a crossover design and intramuscular injections of drug or vehicle.	21 (cf.17)
Moderate	Exposure of sub-adult squid to low frequency ultrasound for a period of two hours at an intensity sufficient to cause permanent damage to the statocysts.	33
Severe	Exercising adult squid in a flume to exhaustion (~45-60min). At the point of exhaustion animals are killed humanely for metabolic analysis of tissues.	6
Severe	Surgical removal of the distal 1% of the tip of an arm in adult octopus without general anaesthesia and followed by recovery for 24h after which the animal is killed humanely.	9
Severe	Removal of the distal 50% of one arm under general anaesthesia in adult octopus with prior nerve block using injection of local anaesthetic. After 24h the animal will be killed humanely.	47
Severe	Induction of arm autotomy by crushing injury in octopus followed by recovery from 1-7days followed by humane killing.	39
Severe	Fertilised eggs from squid are reared at higher temperatures to simulate climate change with the expectation that ~25% of hatchlings will have a malformation that impairs the ability to feed but will survive for the four weeks of the study after which they will be humanely killed.	14
Severe	Exposure of newly hatched octopus paralarvae to a range of concentration of ammonium or nitrite ions to identify the median lethal concentration.	50
Severe	Transection of one mantle connective in adult octopus under general anaesthesia followed by recovery over a period of three months with weekly high resolution video recording of the mantle with the animals previously trained to enter the smaller recording chamber.	44
Severe	Administration of a parasite to adult octopus followed every 10 days by behavioural monitoring and haemolymph sampling under brief general anaesthesia for a period of three months. The parasite is expected to cause skin lesions in 50% of the animals after two months.	25
Severe	Surgical removal of the entire supra-oesophageal brain from octopus under general anaesthesia with recovery and investigation of the sensitivity to noxious stimuli.	30
Upper threshold	Exposure of adult cuttlefish to a potential pollutant to identify LD ₅₀ .	43

Table 4. List of possible procedures using live cephalopods (not an exhaustive list) as deduced from the set of scenarios included in the COST Action FA1301 PAS-Survey. See text, Table 2 and also examples provided by Fiorito and coauthors.^{2, 12} Procedures are allocated to a *prospective severity* category.

Non-recovery and upper-threshold procedures using cephalopods (e.g., exposure of a cephalopod to a potential pollutant to identify LD₅₀) are not included in this list.

Sub-threshold procedures

Behavioural studies that do not cause pain, suffering, distress or lasting harm exceeding the lower threshold for regulation (e.g., choice of shelter or backgrounds and enrichment mimicking the natural environment; measurement of predatory attack latency where there is no risk of the prey attacking the predator)

Feeding studies where food restriction does not cause any harm (e.g., food preferences).

Feeding a diet (we refer here to diets that meet the full nutritional requirements of the animal, or that have been used successfully in a closely-related* species)

Food deprivation (i.e., withdrawal of food) for a short interval relative to normal food intake and to the life-stage

Adding inert markers in the diet to follow the passage of digesta to be estimated through collection of faeces

Manipulations of housing parameters (e.g., water temperature, photoperiod, water gases and ion levels) within the ranges experienced by the species in its natural habitat where the speed of change is such to be known to be tolerated by the species and where such changes may occur gradually

Mild procedures

Behavioural studies involving short-term exposure to an artificial predator (e.g., pictures or movies of predators)

Non-invasive imaging of animals (e.g., ultrasound) with appropriate sedation or general anaesthesia

Induction and maintenance of short duration general anaesthesia (< 30 min) with recovery on a single occasion using an appropriate agent and concentration for the species and life stage (e.g., weigh and measure a cephalopod for a scientific purpose)

Feeding cephalopods on modified diets (fulfilling animals' nutritional needs according to life stage, and assuming no reduction in quantity or quality of the diet compared with normal feed); use of non-invasive or minimally-invasive methods for the evaluation of behaviour and growth

Feeding cephalopods with modified diets that do not meet all of the animal's nutritional needs where daily mortality would be assessed**

Mild procedures, continued

Studies involving short-term food deprivation followed by administration of a marker in the food and measurement of oro-anal transit time

Daily testing (e.g., 5 days) of the predation and/or attack response of adult cephalopods to an inaccessible prey

Studies of taste aversion in adult cephalopods in behavioural/training paradigms

Studies of post-reproductive senescence until death by natural causes (this assumes that humane end points are in place, veterinary advice is sought throughout and treatment given if appropriate and is compatible with the aims of the study)

Haemolymph sampling under general anaesthesia on a single occasion with recovery (volumes and techniques are limited to those recommended in published guidelines² and/or national legislation)

Moderate procedures

Repeated haemolymph sampling (e.g., weekly sampling for a given period) under general anaesthesia within acceptable limits for haemolymph removal for the species and life stage

Surgery under general anaesthesia (e.g., removal of a small piece of fin; removal of 10% of an arm; implantation of data logging-tags; see also Fiorito et al.^{2, 12}) with recovery where the surgical intervention is likely to result in mild to moderate impairment of normal function

Food deprivation of adult cephalopods for prolonged times beyond the normal interval between meals (e.g., 5 days in adult *O. vulgaris*²²) followed by humane killing

Exposure to substances with unknown biological effects in cephalopods (e.g., a potential general anaesthetic substance not previously investigated using immersion of the animal in escalating concentrations)

Exposure of sub-adult animals to a given sound frequency for a period of two hours at an intensity known to cause tissue damage resulting in abnormal behaviour***

Removal of a portion of arm tip (e.g. for genotyping) from an animal without general anaesthesia

Behavioural studies that are expected to cause moderate aversive reactions or distress or moderate interference with normal behaviour

Pharmacokinetic studies where a single dose of an experimental drug is administered to a cephalopod using a semi-rigid feeding tube inserted via the beak with the animal manually restrained; the pharmacokinetic part of the study will require repeated haemolymph sampling under sedation(see above)

Training protocols involving an inescapable aversive stimulus in an adult cephalopod and a daily intramuscular injection of vehicle or drug (not longer than 7 days at the end of which the animals are humanely-killed)

Investigation of an opiate analgesic using subcutaneous administration of vehicle, opiate agonist or antagonist (appropriate experimental design should be applied) and a noxious level of mechanical stimulation to parts of the body of the animal

Severe procedures

Surgery or other procedures carried out under general anaesthesia which are expected to result in severe or persistent moderate postoperative pain, suffering or distress, or severe and persistent impairment of the general condition of the animal followed or not of humane killing (e.g., removal of the distal 50% of one arm in adult cephalopods with prior nerve block using injection of local anaesthetic; induction of arm autotomy by crushing injury; removal of the entire supra-esophageal mass from adult with recovery and investigation of the sensitivity to noxious stimuli)

Surgical removal of the distal the tip of an arm without general anaesthesia and followed by recovery for 24h (after which the animals is killed humanely)

Toxicity testing where death is the endpoint or fatalities are to be expected, or severe pathophysiological states/malformations are induced (e.g. exposition of fertilised cephalopod eggs to higher temperatures to simulate climate change that might affect hatchlings; exposure of newly hatched paralarvae or juvenile cephalopods to a range of concentration of ammonium or nitrite ions to identify the median lethal concentration)

Forced swimming tests with exhaustion as the endpoint

Disease studies where the disease or parasite in question is known to cause death or skin lesions and where the study cannot be controlled to avoid mortality

* caution should be provided when transposing a procedure between different species, considering the wide biological and physiological divergence that characterize cephalopods.

** if high mortality rates, depending from assessed values for the species and life-stage, are observed (e.g., > 10%) the study will be terminated using a pre-set humane end-point.

*** see for example ^{59, 60}

Figure Legends

Figure 1. Distribution of scores (% responses) for three individual scenarios (5, 29 and 46) with the highest scores in the non-recovery category. UTD = unable to decide category. See text for definitions.

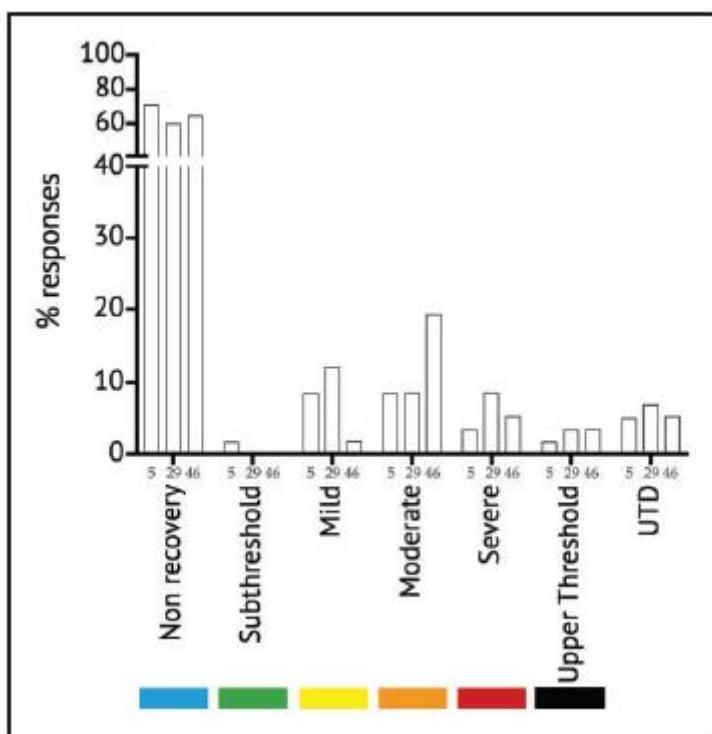


Figure 2. Relationship between scenarios deduced after cluster analysis utilizing the average linkage method. In the dendrogram, the distance between two clusters is defined as the average distance between all pairs of the two clusters' members and shown as rescaled distance (%). The clustering algorithm was applied to scores for all 50 scenarios (numbers refer to the scenario identifier in the original survey). For clarity, the groupings of scenarios clusters have been colour coded to indicate the severity category

using the same colour coding as used in all figures: Sub-threshold () Non-recovery () Mild () Moderate () Severe () Upper threshold (). Scenario 22 is an outlier in the dendrogram, reflecting that this was the only scenario for which it was not possible to assign a severity classification (see text for details). The final assigned severity classification derived from all data (see text) is indicated although in the cluster some scenarios on category borderlines. For example: B2 assigned to mild, but on the sub-threshold/mild border so clusters close to sub-threshold, B1; C1_A^{*} assigned to moderate, but on the mild/moderate border so clusters close to mild C1_A^{**}; C1_B^{*} assigned to severe but on the mild/severe border so clusters close to mild C1_B^{*}. Note that the two scenarios in the upper threshold category cluster together (C2_B and C2_A close to the severe scenario grouping). See also **Figure 4** (and **Supplementary Figures 2 and 3**) for additional data used in making final severity classifications.

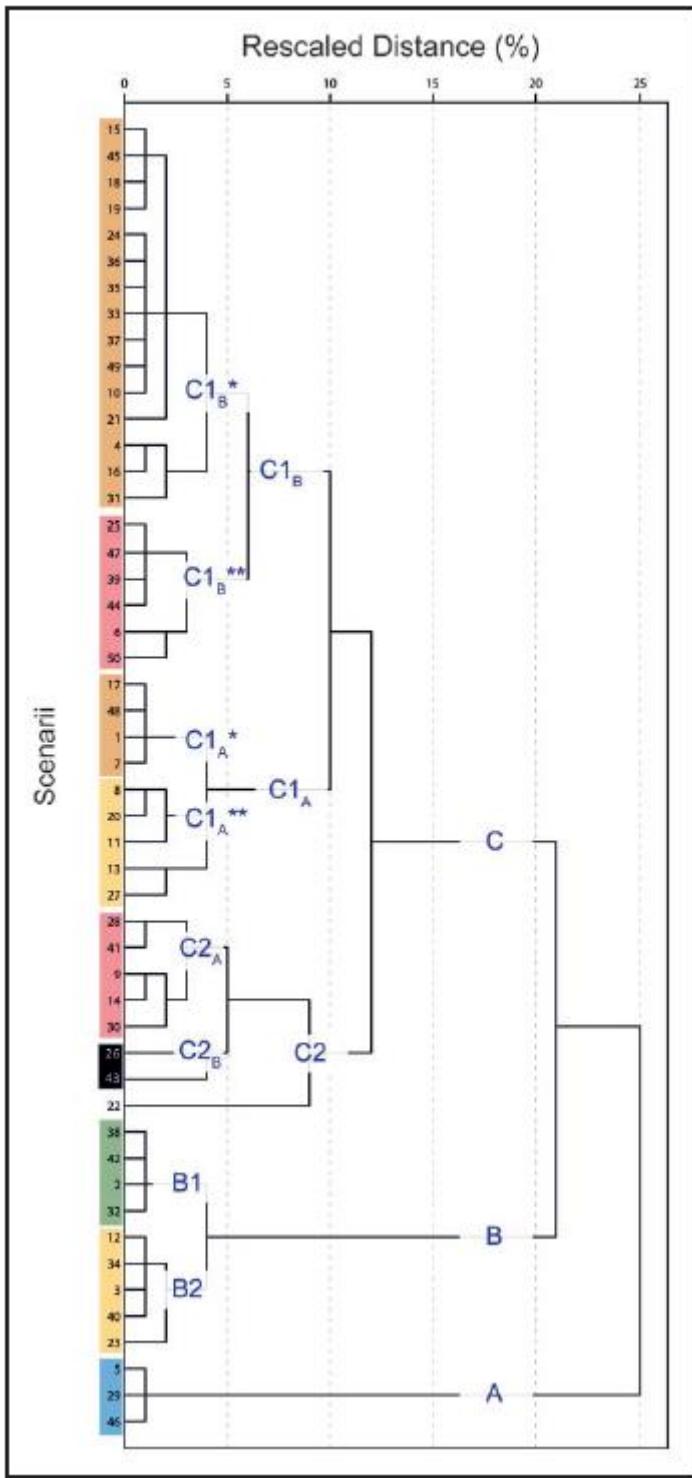


Figure 3. **A:** The distribution of scores (% respondents) in each severity category for four scenarios (32, 42, 38, 2) with dominant sub-threshold scores. **B, C:** the change in the distribution of scores from scenarios with predominantly sub-threshold scores but

increasing scores in severity categories above the regulatory threshold (scenarios 40, 3, 43, 12, 23). The bars illustrate scenarios around the threshold for regulation, to scenarios with dominant “mild” scores (scenarios 8 and 20) and increasing moderate scores (scenario 11) indicating scenarios which should be considered supra-threshold.

Abbreviations: UTD = unable to decide. See text and Supplementary Information for definitions.

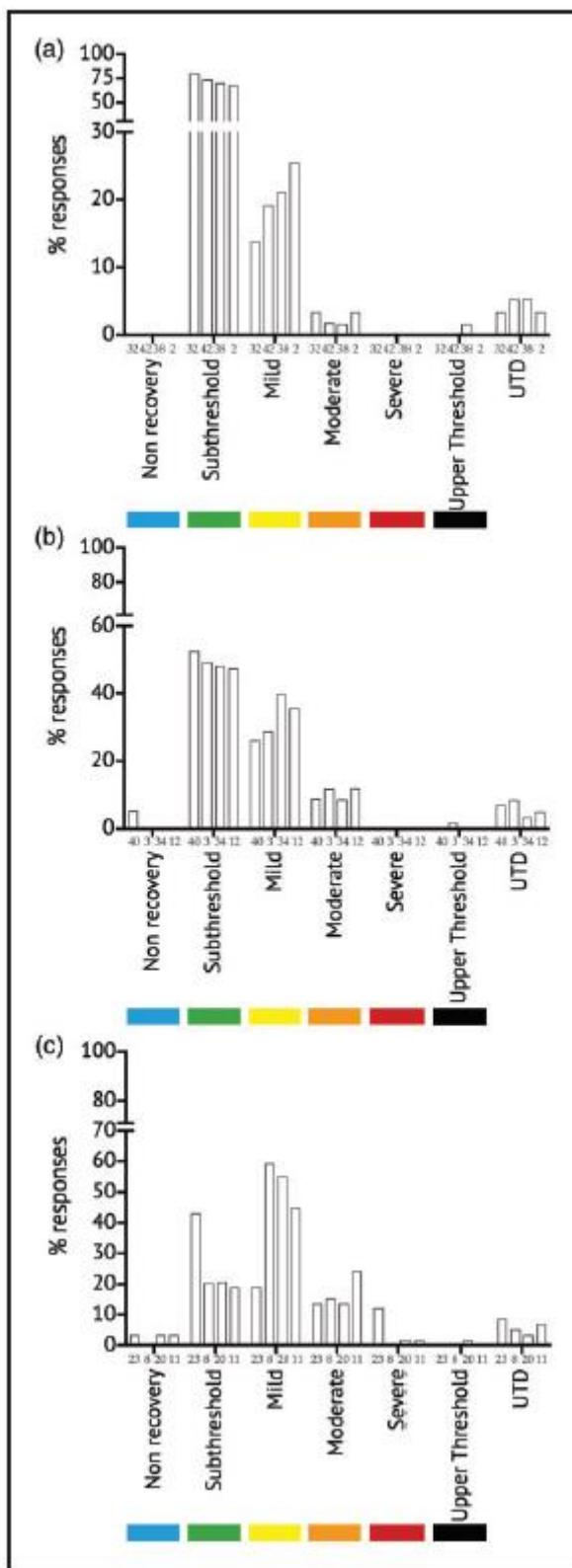


Figure 4. The distribution of scores (% respondents; mean ± SEM) in each severity category for scenarios allocated to the categories defined in Directive 2010/63/EU.

Numbers in parenthesis indicate scenario number in original survey. See text for details
of method of allocation.

