ICES WKMEDS3 REPORT 2015

ICESS ACOM COMMITTEE

ICES CM 2015\ACOM:39

REF. WGFTFB, SCICOM

Report of the Workshop on Methods for Estimating Discard Survival 3 (WKMEDS 3)

20–24 April 2015

London, UK

ICES International Council for the Exploration of the Sea

CIEM Conseil International pour l'Exploration de la Mer

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H. C. Andersens Boulevard 44–46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

Recommended format for purposes of citation:

ICES. 2015. Report of the Workshop on Methods for Estimating Discard Survival 3 (WKMEDS 3), 20-24 April 2015, London, UK. ICES CM 2015\ACOM:39. 47 pp.

Contents

Executive Summary

The primary aim of this meeting was to begin to address terms of reference c and d:

c) Critically review current estimates of discard mortality, with reference to the guidelines detailed in a), and collate existing validated mortality estimates; and

d) Conduct a meta-analysis, using the data detailed in c), to improve the understanding of the explanatory variables associated with discard mortality and identifying potential mitigation measures.

From discussion in previous meetings, it was recognized that these terms of reference could be met as part of a "Systematic Review" (e.g. CEE, 2013; Hughes *et al.*, 2014; Higgins and Green (eds.), 2011). A full-scale systematic review (SR) is a substantial undertaking that will demand considerable resources, particularly researcher's time. One of the first important steps in planning a SR is to establish a review team to draft preliminary protocols, for assessment and development by a wider stakeholder group, including end-user (CEE, 2013). This was the purpose of this meeting and, outside plenary sessions, the meeting was divided into two subgroups to discuss and draft protocols on the relevant components of a SR, in context with ToRs c and d.

Group 1: Critical Review of Survival Assessments and Estimates:

The precision and accuracy of discard survival estimates is likely to vary between different assessments, even on the same species in the same fisheries. This group discussed and developed a protocol for a systematic critical review process, using methods such as those recommended by the Collaboration for Environmental Evidence (http://www.environmentalevidence.org/), to assess different survival studies in terms of essential criteria derived from the WKMEDS Guidelines. These protocols are presented in sections 3.2 and 3.3 of this report.

This review process was applied to a number of case studies to establish a database of validated discard survival estimates (with appropriate measures of uncertainty):

- North Sea Flatfish, in particular Plaice (*Pleuronectes platessa*) and Sole (*Solea solea*)
- Norway Lobster (*Nephrops norvegicus*)
- Skates and Rays (regulated commercial species)

The species addressed in these case studies have attracted attention as potential candidates for "High survival" exemptions from EU Landing Obligation (EU Common Fisheries Policy, Art. 15, para. 2b). At the meeting the work focused on the first two of these three case studies, the work on the skates and rays case study was initiated only, and is not reported further in this report.

Group 2: Meta-analysis of Survival Estimates:

The group provides a quantitative synthesis of the effect size of key explanatory variables from different but related studies. If performed correctly, and using reliable data (see above), this synthesis could substantially increase the power on an analysis to interpret the modifying effects of different variables on discard survival. This group reviewed the different approaches available for conducting a meta-analysis, as part of the systematic review discussed by group 1, including weighted random/fixed effects

models and fuzzy logic, and considered their applicability to the case-study data from task group 1. Based on these discussions, a set of protocols were drafted and are presented in sections 3.2 and 3.5.

In addition, some participants in group 2 reviewed text relating to data analysis in the draft guidelines, in support of:

e) Develop guidelines and where possible identify best practice for undertaking discard survival studies (using the framework detailed in the report of STECF Expert Working Group EWG 13-16) (2014 Workshop).

This report presents a summary of the discussion of these groups and the protocols they proposed for undertaking a systematic review, and associated meta-analysis, of Discard Survival assessments and estimates. These protocols are organized in a format that follows the steps taken in a typical Systematic Review (SR):

A. Problem identification;
B. Identify relevant studies and original data;
C. Data extraction and evaluation;
D. Meta-analysis; and
E. Presentation of review and conclusions.

Note - Proposed protocols for steps A – D are presented in this report. The presentation of the findings of a SR (E) will be addressed at the next meeting of WKMEDS.

1 Introduction

ICES established a Workshop on Methods for Estimating Discard Survival (WKMEDS), in January 2014, in response to a request from the European Commission to address the urgent need for guidance on methods, as identified by STECF EWG 13-16 (STECF, 2014).

EU Member States and Advisory Councils are interested in commissioning survival studies to investigate the feasibility of exemptions to the Landings Obligation, under Art. 15, para. 2b of the new EU Common Fisheries Policy. There are practical and scientific limitations to the methods currently available for estimating discard survival (ICES, 1995, 1997, 2000, 2004 and 2005; Revill, 2012; Gilman *et al.*, 2013). Therefore, there is an urgent requirement for the provision of guidelines, or identification of best practice, for undertaking discard-survival studies.

1.1 Terms of Reference

This workshop was chaired by Mike Breen (Norway) and Thomas Catchpole (UK), and will work by correspondence as well as a series of meetings during 2014–2016 to:

1) Develop guidelines and where possible identify best practice for undertaking discard survival studies (using the framework detailed in the report of STECF Expert Working Group EWG 13-16) (2014 Workshop);

2) Identify approaches for measuring and reducing, or accounting for, the uncertainty associated with mortality estimates;

3) Critically review current estimates of discard mortality, with reference to the guidelines detailed in a), and collate existing validated mortality estimates;

4) Conduct a meta-analysis, using the data detailed in c), to improve the understanding of the explanatory variables associated with discard mortality and identifying potential mitigation measures; and

5) Based on ToR a) to d) a CRR should be developed for SCICOM consideration.

The first and second meetings were held on 17-21 February and 24-28 November, 2014, at ICES HQ in Copenhagen, to address ToR a).

The third meeting was held on 20-24 April 2015, at the Department of Environment, Food and Rural Affairs, London, to address ToR c) and d).

2 Meeting Overview

2.1 Meeting Objective

The primary aim of this meeting was to begin to address terms of reference c and d:

c) Critically review current estimates of discard mortality, with reference to the guidelines detailed in a), and collate existing validated mortality estimates; and

d) Conduct a meta-analysis, using the data detailed in c), to improve the understanding of the explanatory variables associated with discard mortality and identifying potential mitigation measures.

2.2 Meeting Structure

From discussion in previous meetings, it was recognized that these terms of reference could be met as part of a "Systematic Review" (e.g. CEE, 2013; Hughes *et al.*, 2014; Higgins and Green (eds.), 2011). A full-scale systematic review (SR) is a substantial undertaking that demands considerable resources, particularly researcher's time. One of the first important steps in planning a SR is to establish a review team to draft preliminary protocols, for assessment and development by a wider stakeholder group, including end-user (CEE, 2013). This was the purpose of this meeting and, outside plenary sessions, the meeting was divided into two subgroups to discuss and draft protocols on the relevant components of a SR, in context with ToRs c and d.

Critical Review of Survival Assessments and Estimates:

The precision and accuracy of discard survival estimates is likely to vary between different assessments, even on the same species in the same fisheries. This group discussed and develop a protocol for a systematic critical review process, using methods such as those recommended by the Collaboration for Environmental Evidence (http://www.environmentalevidence.org/), to assess different survival studies for essential criteria derived from the WKMEDS Guidelines. These protocols are presented in sections 3.2 and 3.3 of this report.

This review process was applied to a number of case studies to establish a database of validated discard survival estimates (with appropriate measures of uncertainty):

- North Sea Flatfish, in particular Plaice (*Pleuronectes platessa*) and Sole (*Solea solea*)
- Norway Lobster (*Nephrops norvegicus*)
- Skates and Rays (regulated (quota) commercial species)

The species addressed in these case studies have attracted attention as potential candidates for "high survival" exemptions from EU Landing Obligation (EU Common Fisheries Policy, Art. 15, para. 2b). The focus of the work at the meeting was on the first two of these three case studies, the work on the skates and rays case study was initiated only, and is not reported further in this report.

Meta-analysis of Survival Estimates:

It provides a quantitative synthesis of the effect size of key explanatory variables from different but related studies. If performed correctly and using reliable data (see above), this synthesis could substantially increase the power on an analysis to interpret the modifying effects of different variables on discard survival. This group reviewed the

different approaches available for conducting a meta-analysis, as part of the systematic review discussed by group 1, including weighted random/fixed effects models and fuzzy logic, and considered their applicability to the case-study data from task group 1. Based on these discussions, a set of protocols were drafted and are presented in sections 3.2 and 3.5.

In addition, some participants in group 2 reviewed text relating to data analysis in the draft guidelines, in support of:

e) Develop guidelines and where possible identify best practice for undertaking discard survival studies (using the framework detailed in the report of STECF Expert Working Group EWG 13-16) (2014 Workshop).

2.3 Participants

The third meeting was attended by 18 people (Table 2.1).

Table 2.1: Participant in WKMEDS 3 (§=lead).

Name	Organization	Country	Group
Hugues Benoît	Gulf Fisheries Centre, Fisheries and Oceans (DFO)	Canada	2.Meta-analysis
Mike Breen §	Institute for Marine Research (IMR)	Norway	2.Meta-analysis
Tom Catchpole §	Centre for Environment, Fisheries and Aquaculture Science (Cefas)	UK	1.Critical Review
Chun Chen	Wageningen University and Research Centre (WUR)	Netherlands	2.Meta-analysis
Jim Ellis	Centre for Environment, Fisheries and Aquaculture Science (Cefas)	UK	1.Critical Review
Dorothée Kopp	French Research Institute for Exploitation of the Sea (Ifremer)	France	1.Critical Review
Alessandro Ligas	Centro Interuniversitario di Biologica Marina (CIBM)	Italy	2.Meta-analysis
Niels Madsen	Danish Technical University (DTU-Aqua), Hirtsals	Denmark	2.Meta-analysis
Sonia Mehault	French Research Institute for Exploitation of the Sea (Ifremer)	France	1.Critical Review
Caroline Methling	Danish Technical University (DTU-Aqua), Hirtsals	Denmark	1.Critical Review
Pieke Molenaar	Wageningen University and Research Centre (WUR)	Netherlands	1.Critical Review
Hans Nilsson	Swedish University of Agricultural Sciences (SLU)	Sweden	1.Critical Review
Peter Randall	Centre for Environment, Fisheries and Aquaculture Science (Cefas)	UK	1.Critical Review
Karine van der Reijden	Wageningen University and Research Centre (WUR)	Netherlands	1.Critical Review
Sebastian Uhlman	Institute for Agricultural & Fisheries Research (ILVO)	Belgium	2.Meta-analysis
Bob van Marlen	Wageningen University and Research Centre (WUR)	Netherlands	2. Meta-analysis
Inger Wilms	CVO-Visserij	Netherlands	1.Critical Review

2.4 Agenda

The agenda for the third meeting of WKMEDS is detailed in Appendix 1.

3 Protocols for a Systematic Review

3.1 Introduction – What is a Systematic Review?

A systematic review (SR) is a review of the scientific literature that is focused on a specific research question that tries to identify, appraise, select and synthesize all high quality research evidence relevant to that question (Higgins and Green, 2011; CEE, 2013). This process has evolved to encompass rigorous synthesis and analysis, both quantitative and qualitative, of available research (see section 3.5) (Herman *et al.*, 2008; CEE 2013). Analyses may be narrative, such as a structured summary and discussion of the studies' characteristics and findings, or quantitative, that is involving statistical analysis (Higgins and Green, 2011).

Synthesis of existing research is a crucially important way of advancing science, especially where science can inform decision-making. Critical evaluation through narrative reviews can provide significant conceptual advances, as they can examine patterns in the flaws of studies and identify priorities for additional research. However, they can often be hampered by subjectivity in study selection, low repeatability, and limited ability to quantify and explain variation in study outcomes. In contrast, a systematic review, in combination with appropriate meta-analysis, provides a quantitative approach to undertake synthesis, statistical analysis, and summary of a collection of independent studies by systematically combining results of previous research.

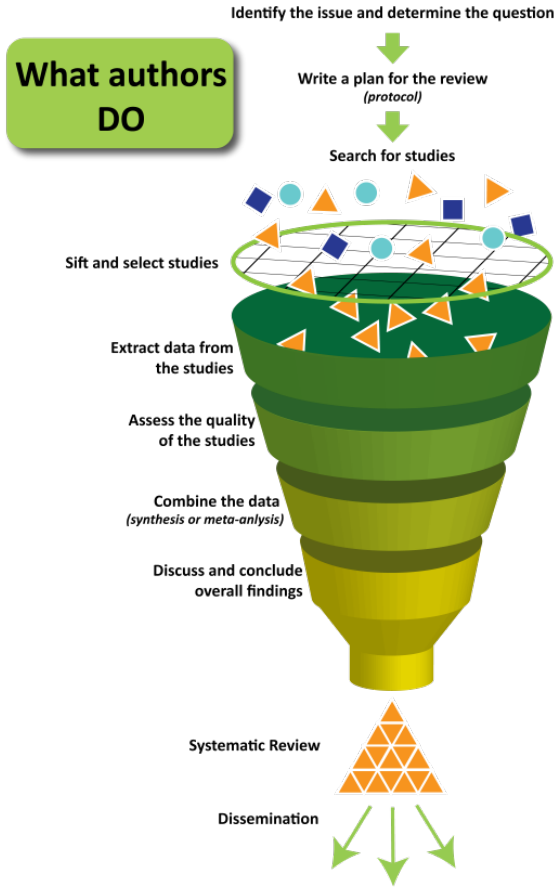

Figure 3.1: An overview of the Systematic Review process. Source: Centre for Health Communication and Participation La Trobe University, Australasian Cochrane Centre http://navigatingeffectivetreatments.org.au/exploring_systematic_reviews_what_authors_do.html

This report presents a summary of the discussion and protocols proposed by WKMEDS for undertaking a systematic review, and associated meta-analysis, of Discard Survival assessments and estimates. These protocols are organized in a format that follows the steps taken in a typical Systematic Review (SR):

A) Problem identification;

B) Identify relevant studies and original data;

C) Data extraction and evaluation;

D) Meta-analysis; and

E) Presentation of review and conclusions.

[Note - Proposed protocols for steps A – D will be presented in this report. The presentation of the findings of a SR (E) will be addressed at the next meeting of WKMEDS.]

In drafting the protocols, WKMEDS has taken guidance from two sources, in particular:

- Cochrane Collaboration - Handbook for Systematic Reviews of Interventions (ed. Higgins and Green, 2011). (http://handbook.cochrane.org/)
- Collaboration for Environmental Evidence (CEE) 2013. Guidelines for Systematic Review and Evidence Synthesis in Environmental Management. Version 4.2. Environmental Evidence: (www.environmentalevidence.org/Documents/Guidelines/Guidelines4.2.pdf)

3.2 A) Problem Identification

The first essential step in conducting a systematic review (SR) of scientific data is to correctly formulate the question/s to be directed at that data. These questions should be scientifically relevant, and answerable, and should be perceived as unbiased by stakeholders and end-users of the data (CEE, 2013).

3.2.1 Stakeholder Involvement

In particular, the CEE stress the importance of involving the stakeholder and user groups at an early stage in the review process, to ensure both the question and the review protocol have utility and validity for the issue they intend to address. To this end, stakeholders and fisheries managers were invited to attend the WKMEDS 3 meeting and contribute to defining the review questions and protocols. Further input from stakeholders, managers and the wider scientific community on these proposed protocols will also be invited with the publication of this report.

Inger Wilms from the Netherlands fishing organization (CVO-Visserij) contributed to subgroup 1 and the plenary discussions on the selection of suitable review questions. Furthermore, she took part in the review and compilation of available data on flatfish survival. Sarah Adcock, from the UK Department of the Environment, Fisheries and Rural Affairs (DEFRA), also attended part of the meeting and presented information about the implementation of the Landing Obligation (LO) by the UK Ministry. Based on discussion, including scientists, fishery managers and fishery representatives, WKMEDS prioritized *Nephrops*, Plaice and Sole as candidate species for case study reviews within the meeting.

3.2.2 Defining the Review Question

In their guidelines on SR, the CEE identify that the most common questions for SR have four definable elements, often referred to as the PICO or PECO (Population, Intervention/Exposure, Comparator, Outcome) elements (Table 3.2.2.1). Furthermore, the CEE advise that a SR should focus on a primary question. Although given sufficient reliable data, with suitable scope, it may be practical to also consider secondary questions within a SR.

Table 3.2.2.1: The key elements of a PICO/PECO question for a Systematic Review (from CEE, 2013).

Question element	Definition
Population(of subjects)	Unit of study (e.g. ecosystem, species) that should be defined in terms of the statistical populations of subject(s) to which the intervention will be applied.
Intervention/ exposure	Proposed management regime, policy, action or the environmental variable to which the subject populations are exposed.
Comparator	Either a control with no intervention/exposure or an alternative intervention or a counterfactual scenario.
Outcome	All relevant outcomes from the proposed intervention that can be reliably measured or outcome that might result from exposure to an environmental variable.

One of the main objectives of WKMEDS is to provide suitable information on discard survival for fisheries management in relation to the Landing Obligation, based on available scientific data. This information on discard survival may be used to allow for exemptions to the Landing Obligation, based on survival being sufficiently high.

To this end, following discussions in group 2 and in plenary, WKMEDS identified the following questions as candidates to be addressed by the application of a systematic review and meta-analysis (see also Table 3.2.2.2):

Primary Question:

- What is the discard survival (and variability) of a particular species/taxa in a particular scenario (i.e. fishery/métier, region, discard practice...)?

Secondary Questions:

- Is the survival 'high' enough for a particular species in a particular scenario?
- What is the effect of covariate X (e.g. species/taxa, gear type, season, handling processes...) on survival?
- What is the effect size (and variability) of experimental factors (e.g. cage study vs. tagging study, monitoring period) on survival?
- How consistent is an effect on survival across the studies applicable to a particular scenario?

The primary questions for the candidate case studies are therefore:

- What is the discard survival (and variability) of *Nephrops norvegicus* in European trawl fisheries?
- What is the discard survival (and variability) of plaice (*Pleuronectes platessa*) and sole (*Solea solea*) in European fisheries?

Table 3.2.2.2: Case studies primary questions presented in PECO format.

POPULATION	EXPOSURE	COMPARATOR	OUTCOME
Nephrops norvegicus	Discarded in European trawl fisheries	Suitable control groups	Survival (and variance)
Pleuronectes platessa	Discarded in European fisheries	Suitable control groups	Survival (and variance)
Solea solea	Discarded in European fisheries	Suitable control groups	Survival (and variance)

3.3 B) Identify Relevant Studies and Original Data

To identify all relevant studies that have generated discard survival estimates in the selected case studies, a two stage literature search was completed.

3.3.1 Stage 1 – literature search

The first stage was a literature search using the scientific citation search engine 'Web of Science'. Web of Science (WoS, previously known as Web of Knowledge) is an indexing service that provides a comprehensive scientific citation search. It gives access to multiple databases that reference cross-disciplinary research, which allows for in-depth exploration of specialized sub-fields within an academic or scientific discipline.

The following search terms were used to identify the literature. These terms were based on the papers with which the participants were familiar and the expertise of the group. Search criteria 1-12 below were each combined with terms used to describe the taxon for each case study. Therefore, in each case study, 12 searches were completed and all of the results were entered onto an Excel spreadsheet. Those references that contained original discard survival estimates were selected and acquired.

Search engine: Web of science

Search criteria ('Topics'):

1) Discard* AND surviv*
2) Discard* AND mortality
3) Discard* AND vitality
4) Bycatch* AND surviv*
5) Bycatch* AND mortality
6) Bycatch* AND vitality
7) By-catch AND surviv*
8) By-catch AND mortality
9) By-catch AND vitality
10) Post-release AND mortality
11) Post-release AND surviv*
12) Post-release AND vitality

Each of the above were used in combination with the taxon terms relevant to each case study:

Case study 1 – Nephrops norvegicus

AND *Nephrops norvegicus* OR Norway lobster OR Dublin Bay Prawn OR Langoustine OR Norwegian lobster

Case study 2 – Pleuronectes platessa

AND *Pleuronectes platessa* OR European plaice OR plaice NOT American plaice

Case study 3 – *Solea solea*

AND *Solea solea* OR sole

3.3.2 Stage 2 – extended literature search

The second stage was to examine the selected articles and identify other sources of original discard survival data from the reference lists given in the selected articles. These articles and data sources were then acquired, where possible, and a final list of literature sources for each case study were was compiled. (Tables 3.3.2.1-3.3.3.3). During the first stage it was evident that in all but one study plaice and sole were investigated together, so these articles were combined. At the meeting, the extended literature search focused on sole, with a view to complete searches for plaice outside the meeting.

Table 3.3.2.1: Results from the literature search. The results for *Solea solea* and *Pleuronectes platessa* are presented together because in all but one case the selected articles presented survival data on both species.

Case study species	Number of WoS results	Number WoS hits with original discard survival estimates	Number of referenced sources with original discard survival estimates	Total no. of references with original discard survival estimates
Nephrops norvegicus	30	2	10	10+1*
Solea solea/ Pleuronectes platessa	52	3	4**	7**

***one study was made available for the first time at the meeting and was included in the review process.**

****sole only**

Table 3.3.2.2 Final list of identified sources of original data on discard survival rates for case study 1 *Nephrops norvegicus*.

Title	Authors	Journal	Year
Discarded Nephrops survival after trawling.*	Méhault, Morandau, Fifas.	Working Document. Ifremer	2011
Impact of capture method and trawl duration on the health status of the Norway lobster, *Nephrops norvegicus*.	Ridgway, I.D., Taylor, A.C., Atkinson, R.J.A., Chang, E.S., Neil, D.M.	Journal of Experimental Marine Biology and Ecology 339, 135–147.	2006
Physiological changes in the Norway lobster *Nephrops norvegicus* (L.) escaping and discarded from commercial trawls on the West Coast of Scotland - I. Body fluid volumes and haemolymph composition after capture and during recovery	Harris, RR; Andrews, MB	Journal of Experimental Marine Biology and Ecology, 320: 195-210	2005
The efficacy of releasing caught *Nephrops* as a management measure.	Castro, M., Arau´ jo, A., Monteiro, P., Madeira, A. M., and Silvert, W.	Fisheries Research, 65: 475-484	2003
Roundfish and *Nephrops* survival after escape from commercial fishing gear.	Wileman, D.A., Sangster, G.I., Breen, M., Ulmestrand, M., Soldal, A.V., Harris, R.R.,	Final report to the EC (FAIR-CT95-0753), Brussels.	1999
Survival of discarded *Nephrops* norvegicus in the Bay of Biscay and in the Celtic Sea.	Charuau, A, Morizur,Y, Rivoalen, JJ.	ICES CM 1982/B:13	1982
Survival of discarded *Nephrops*.	Edwards, E., Bennett, D.B.	ICES CM 1980/K:10	1980
Essai de détermination de taux de survie des langoustines hors taille rejetées lors des opérations de pêche commerciales	Gueguen et Charuau	ICES CM 1975/K:12	1975
The survival of small *Nephrops* returned to the sea during commercial fishing	Symonds, D.J., Simpson A.C.	Journal Du Conseil Conseil International Pour L'Exploration De La Mer 34 (1): 89–98	1971
Survival of *Nephrops* returned to the sea	Symonds D.J., Simpson A.C.	ICES CM 1968/K:14	1968
Gear effects on mortality of discarded Norway lobster *(Nephrops norvegicus*) in Swedish fisheries*	Nilsson H.C., Ulmestrand M., Thorvaldsson B., Hilvarsson A. and Valentinsson D.	Unpubl.	2015

***a study recently completed and available for the first time at the meeting**

Table 3.3.2.3 Final list of identified sources of original data on discard survival rates for case studies 2 and 3 *Solea solea* and *Pleuronectes platessa*

Title	Authors	Journal	Year
Short-term survival of discarded target fish and non-target invertebrate species in the "eurocutter" beam trawl fishery of the southern North Sea.	Depestele, J., Desender, M.; Benoît, H.P., Polet, H., Vincx, M	Fisheries Research 154: 82-92.	2014
Mortality of adult plaice, *Pleuronectes platessa* and sole, *Solea solea* discarded from English Channel beam trawlers	Revill, AS, Broadhurst, MK, Millar, RB	Fisheries Research 147: 320–326	2013
Importance of discards of a beam trawl fishery as input of organic matter into nursery areas within the Tagus estuary	Cabral, HN, Teixeira, CM, Gamito, R, Costa MJ	Nutrients and Eutrophication in Estuaries and Coastal Waters Developments in Hydrobiology 164, 449-455	2002
Survival of bycatch from a beam trawl.	Kaiser, M. J., and B. E. Spencer.	Marine Ecololgy Progress Series 126:31–38	1995
Mortality of fish from the bycatch of shrimp vessels in the North Sea	R. Berghahn, M. Waltemath, A.D. Rijnsdorp	Journal of Applied Ichthyology, 8, pp. 293–306	1992
On the survival of plaice and sole discards in the otter trawl and beam trawl fisheries in the North Sea.	Van Beek, F.A., van Leeuwen, P.I., Rijnsdorp, A.D.	Netherlands Journal of Sea Research 26(1): 151-160	1990
Sterblichkeit untermaβiger Plattfische im Beifang der Garnelenfischerei.	Kelle, W.	Ber. dt. wiss. Kommn. Meeresforsch, 25 (1/2): 77-89.	1976/77

3.4 C) Data Extraction and Evaluation

3.4.1 Critical Review Framework

A critical review framework was developed specifically for discard survival data research. The framework was developed to reflect the guidance document on conducting discard survival assessments generated by WKMEDS (ICES, 2014). The framework also benefited by the ongoing survival assessment being conducted by members of the group. Progress and experiences on projects being conducted in Sweden, Denmark, The Netherlands, England, France and Belgium were presented at the meeting to assist in the development of a critical review and continue the method harmonization process that was initiated at the previous meeting.

The critical review framework is separated into the following sections:

- Information on the data source
- Vitality assessments
- Captive observation
- Controls
- Analysis

In each section there are a series of 'yes/no' style questions, designed to allow an assessment on the method, relative quality and utility of the survival estimates. There are also opportunities to make comments about specific observations and add details. The structure of the critical review is given in Table 3.4.2.1. At present the critical review does not include information on tagging methods, this will be added at a later time. None of the references identified in the three case studies included discard estimates based on tagging methods.

The critical review framework was tested by six reviewers each with a separate paper. Feedback from reviewers led to improvements and changes to the framework, including the clarification of questions and a final structure was agreed. The process of critically reviewing all of the identified papers began but was not completed during the meeting. The agreement was for all the references identified for each case study to be reviewed by at least two people familiar with the relevant species, fisheries and/or survival assessment methods.

3.4.2 Meta-analysis data extraction

In parallel with the critical review framework, a database structure was developed into which essential information could be collated from each of the selected articles in a systematic and structured way. The data to be collected were selected based on the main elements of discard survival assessments as identified in the WKMEDs guidance document. A summarized version of the Meta-analysis database is given in Appendix 2. The database includes information on the details of the fishery, the scale of the work, the design of the experiments, and the data from which the survival estimates are derived.

It was recognized that the database had to be sufficiently flexible to capture the most detailed information from the articles. Data were presented in different styles and at different levels of aggregation and when data in the articles were tabulated, it was beneficial to transpose all of the data into the database. The structure developed, allowed each different treatment to be entered separately and it could convert easily into a relational database. Therefore, if trip level discard survival estimates are provided these

are captured as separate treatments, the same is for temperature treatments, seasonal data or data presented at time intervals during captive observation experiments. Each treatment is entered as a separate column of data. The database was tested by four reviewers and was improved following feedback. The data format entered into the database will be determined by the presentational format in the article, but the database is designed to capture all of the data however it is presented.

The database structure was developed at the meeting and the process of entering data on to the database from articles from each of the case studies was commenced.

Table 3.4.2.1: The framework of the critical review used to assess literature on discard survival estimates for the three case studies.

	Critical Review Framework
0	Reviewer
0.1	Reference Number (from References tab)
0.2	Does this data source provide an original estimate of discard (post release) survival?
	If 'No' do not review, if 'Yes' complete critical review process
0.3	Main species of interest
	Vitality assessments
1	Does the method include assessments of the health or vitality of the fish, including immediate mortality estimate?
	If 'No' go to 2, if 'Yes' complete 1.01 to 1.14
1.01	Is there a description how the assessed fish were selected from the catch?
1.02	Is the fish selection method best described as random, stratified, bias towards a part of the catch (e.g. live fish), or unknown?
1.03	Is there a description/protocol provided for each health/vitality category?
1.04	Is there a description provided for category 'Dead'?
1.05	Were assessment responses (reflexes) derived from fish that had not been exposed to capture treatment and were they consistently observed?
1.06	Is there a description of time limits for responses/reflexes (e.g. operculum movement in 5 seconds)
1.07	Is the container in which the health assessments conducted appropriate to the species and adequate to observe responses? (e.g. orientation)
1.08	Is observer bias in health assessments discussed/minimized/account for?
1.09	Are protocols effective in assessing health/vitality/injury?
1.1	Are assessments consistent across all parts of the study? (e.g. captive observationvs.assessing the catch)
	Captive Observation
2	Does the method include captive observation experiments?
	If 'No' go to 3, if 'Yes' complete 2.01 to 2.19
2.01	Are the holding / transfer facilities described?
2.02	Are holding / transfer facilities sympathetic to the biological/behavioural needs of the subjects?
2.03	Remarks on holding tanks (construction material, volume/surface area, water movement, lighting, shelter, nutrition, predators excluded, monitoring with disturbance)
2.04	Are the holding / transfer conditions consistent across treatments / replicates?
2.05	Remarks on consistency in conditions
2.06	Is there potential for additional stress / injury / mortality (or escapes)?
2.07	Remarks on potential additional stresses

	Critical Review Framework
2.08	Are the holding / transfer conditions representative of "ambient" conditions?
2.09	Remarks on environmental conditions in holding/transfer
2.1	Is there a suitable definition of "dead"?
2.11	Are there appropriate protocols for handling/removal/measurement of dead specimens? (e.g. dead subjects removed regularly)
2.12	Are there appropriate protocols for monitoring live specimens?
2.13	Is there enough resolution in the monitoring/observation schedule over time? (e.g. frequency on observation)
2.14	Remarks on monitoring schedule
2.15	Was there potential for inducing stress/injury in subjects during observation?
2.16	Remarks on potential for inducing stress through observation protocols
2.17	Was mortality observed to asymptote?
2.18	Remarks on monitoring to asymptote
	Controls
3	Were controls used to account for experimental biases?
	If 'No' go to 4, if 'Yes' complete 3.01 to 3.09
3.01	Were they representative of the subject population / treatment groups? E.g. With respect to: biological characteristics (length, age, sex, condition); number / density; spatial and temporal origin
	If not, why?
3.02	Were they representative of experimental conditions? i.e. did the treatment and control subjects experience identical experimental conditions, with exception of the treatment effect?
	If not, why?
3.03	Were treatment and control subjects randomly selected to account for selection bias?
3.04	Were "Blind controls" used to account for Performance and Measurement biases?
3.05	Has the potential for confounding effects from acquisition methods been addressed/discussed? (e.g. had the control fish been recently caught?)
3.06	Remarks on use of controls
	Analysis
4	Is the analysis that derived the survival estimates described?
4.01	Is the sample representative of the catch?
4.02	Does the sample adequately describe the population in the wider fishery?
4.03	What form is the data? Cross sectional / continuous / multinomial
4.04	Was mortality modelled to asymptote?
4.05	Remarks on modelling to asymptote
4.06	What form of analysis was conducted? (e.g. GLM)
4.07	Are the conclusions supported by the data / analysis?
4.08	Are the conclusions based on a summary of the data or statistical inference?

3.5 D) Meta – analysis

The available methods for conducting a meta-analysis with discard survival data (ToR d) were reviewed and discussed by participants in group 2. The output from these reviews and discussions are presented in this section as an introductory over-view of meta-analysis. In addition, more detailed protocols for conducting a meta-analysis of survival (binary / dichotomous) data are presented in Appendix 3, along with links to relative “R” packages and functions.

This report will limit discussion to the meta-analysis of dichotomous (binary) data, as this is the most likely data format to be generated by survival assessments. Some survival assessments may generate longitudinal (time-to-event) data, however it is anticipated that such data will incorporated into a binary based meta-analysis.

In drafting these proposed protocols this group has relied substantially on the guidelines for meta-analysis provided by the Cochrane Collaboration (http://www.cochrane.org/) and published in their Handbook for Systematic Reviews of Interventions (eds. Higgins and Green, 2011). The primary reason for this is that because the Cochrane Collaboration focuses on the systematic review of medical studies it has very well established protocols for working with survival data.

3.5.1 Introduction – What is meta-analysis?

Meta-analysis is a statistical technique to summarize the numerical results of a range of different studies and produce a summary statistic (together with its confidence interval), which gives the user a means of comparing the effect of an intervention (in this case discarding) compared with a baseline (or control).

The meta-analysis approach first appeared in the 1970s, with its application in medical and social sciences (see Smith and Glass, 1977; Glass and Smith, 1978). The first applications of meta-analysis to ecological data were conducted in the 1990s (see Järvinen, 1991; Arnqvist and Wooster, 1995; Gurevitch *et al.*, 2001), with a sharp increase in use in the last decade.

3.5.2 Advantages and limitations of meta-analysis

A true meta-analysis approach combines the measures of the effects from individual studies into an estimate of the overall strength of the effect (this is then used to determine significance) considering sample size and precision of individual studies and estimate magnitude of effects. There are several other approaches to undertake a quantitative research synthesis, including "vote-counted" and combined probability methods (e.g. Fisher's sum of log methods), but these are limited in their statistical power and they do not provide any information on the magnitude of an effect.

Meta-analysis can be used in a systematic review to: increase statistical power; improve precision, answer questions not addressed by the original studies; address ambiguities due to conflicting results and/or generate new hypotheses. For further discussion on these uses see Deeks *et al.*, (2011). The statistical procedures used in a meta-analysis can be applied to any set of data, however the synthesis will only be meaningful if the studies have been collected systematically. This could be in the context of a systematic review, the process of systematically locating, appraising, and then synthesizing data from a large number of sources (i.e. a selected group of studies) (see sections 3.1-3.4).

The advantages of using a meta-analysis approach are the minimization of subjectivity in evaluating and summarizing existing information, the quantitative measure of overall effects, the statistical significance of conclusions, repeatability, and the fact that more weight can be given to studies with larger sample size and lower variability.

WKMEDS recognized the suitability of applying a meta-analysis approach to the discards survival topic in order to provide stakeholders with reliable estimates of discards survival and the effects of explanatory factors (e.g. gear, season, light/air exposure, etc.) by means of a systematic synthesis and statistical analysis of a collection of previous studies.

In addition, in the event that WKMEDS identifies a lack of suitable information and data for some species and/or areas, and the resultant necessity of undertaking new studies, meta-analysis can be useful in helping to design those studies. For example, meta-analysis may show that in prior studies one outcome index had proven to be more sensitive than others, therefore new studies should be designed in order to be able to provide the same type of outcome index.

3.5.3 The Meta-Analysis Process

Survival assessments use the outcomes of individuals within each study to compare the effects of different interventions (e.g. discarding practices). Meta-analyses focus on pairwise comparisons of interventions, such as an experimental intervention vs. a control intervention, or the comparison of two experimental interventions.

Performing a meta-analysis is a multi-step process, which is highly dependent on the systematic and critical review of the data sources before any analysis takes place. Before conducting a meta-analysis, careful consideration should be given to whether it is appropriate to combine the numerical results of all, or perhaps some, of the studies of interest. Therefore it is recommended that meta-analysis should only be used on data where: studies are related enough to allow meaningful comparison; there is low risk of bias in the primary studies; and where there is no evidence of publication and/or reporting bias (see section 9.1.4 in Deeks *et al.,* 2011 for further discussion).

The first step in this critical review consists of the problem identification with consideration of how appropriate analytical questions should be formulated (see section 3.2). Then, it is necessary to elaborate a strategy for searching and selecting studies (find and read all relevant studies, disregarding those with obvious methodological flaws or that do not report needed information, such as sample size) (section 3.3). The next phase is the extraction of data and information from the selected studies, and the creation of an appropriate database to be populated with the data (section 3.4). Finally, it is necessary to transform data into a 'common currency', the so called effect size (ES), which represent the input data for the meta-analysis.

Deeks *et al*., (2011) define the following principle elements on which most commonly used meta-analyses are based:

1. Calculating a measure of effect;
2. Calculating a summary (pooled) estimate of effect across studies;
3. Assumptions about the variance structure in the meta-analysis model;
4. Assessment of the precision of the summary estimate and the significance of the treatment effect; and
5. Assessment of the likelihood of non-random variation across studies.

These principles are now considered in context with the data WKMEDS is likely to analyse as part of a systematic review.

3.5.3.1 Calculating a measure of effect

Data or test statistics to be combined in a meta-analysis should be transformed into a common format – i.e. effect size (ES). For each study a suitable summary statistic (measure of effect) is calculated to describe the observed intervention effect (compared with some baseline/control). Discard survival data are dichotomous (binary) data, as the outcome can be one of two possibilities, for example, dead or alive. The most commonly encountered effect measures used with dichotomous data are:

- risk difference (RD) (also called the absolute risk reduction);
- risk ratio (RR) (also called the relative risk); and
- odds ratio (OR).

The calculation and interpretation of these measures of effect are discussed in detail in Appendix 3 (section 2), but for a more detailed description and discussion of these measures of effects for binary data see Hamilton, 1979; Scott, 2008, and Newcombe, 2013.

These measures of effect all attempt to account for variations in baseline (i.e. control) survival levels between studies by adjusting the observed treatment survival relative to a corresponding control. Discussions in previous WKMEDS meetings have highlighted concerns over this approach, because it assumes a simple linear relationship between the treatment response and the underlying baseline survival – which may not necessarily be true (ICES, 2014). Although, it is recognized that within a meta-analysis there is a need to normalize the treatment data (i.e. discard survival) with respect to baseline mortality (i.e. controls) across the range of studies and, indeed, this approach may prove informative in identifying the true relationship between the treatment and control mortality. However, it is important to remember that, without an informed insight into the relationship between treatment and control mortality, high levels of baseline mortality simply add to the level of uncertainty associated with the corresponding treatment effects. As such, survival estimates based on such studies, and any derived measures of effect, should be interpreted correctly.

Which measure of effect?

When selecting which measure of effect to use, from a statistical perspective, the Odds Ratio is seen to have a number of advantages over both RD and RR (Newcombe, 2013):

1. Irrespective of the baseline level of risk it is always possible to calculate meaningful OR, unlike RD and RR which may encounter ceiling effects. (For example, starting at a baseline risk of 0.4, it is not possible to achieve an absolute difference of greater than 0.6, or a relative increase of more than 2.5).
2. The OR has symmetrical distribution (because it incorporates contributions from both the observed deaths and survivors), while RR tends towards infinity at either high or low baseline survival levels (depending on the underlying outcome, i.e. death or survival) (Appendix 3, Figure 8.3.2.3.1). This implies that OR is better suited to providing more balanced comparisons between studies with a wide range of baseline survival rates.
3. The odds ratio is a direct output from logistic regression, as its natural logarithm is identical with the regression coefficient.

However, OR does have some disadvantages and inherent paradoxes, not least of all the difficulty that users may have in interpreting the output (Newcombe, 2013). Despite this, WKMEDS III proposes that Odd Ratio be initially considered as the primary measure of effect for meta-analysis of survival assessment data. Although, following the guidance of Schechman (2002), it is recognized that it may be informative to compare the output of both an absolute (e.g. RD) and a relative (e.g. OR) measure. Furthermore, Deeks and Higgins (2010) advise using sensitivity analysis to investigate whether choice of summary statistic is critical to the conclusions of the meta-analysis. The final selection will be made after the preliminary stages of the analysis are completed.

3.5.3.2 Calculating a summary (pooled) estimate of effect across studies

The final output of a meta-analysis is a summary statistic of the estimate of effect (together with its confidence interval) based on the data of all of the individual studies included in the analysis, This summary statistic is derived using a weighted average of the measures of effect (as defined in 3.5.3.1) across the studies of interest.

The weighting in this calculation reflects the amount of information (and hence degree of confidence) in the individual estimates, and the underlying studies. It is applied using a variety of techniques depending on the data format, the measure of effect (see 3.5.3.1) and the variance structure of the data (see 3.5.3.3). These techniques are summarized in Appendix 3 (section 3), along with information on the available software for conducting them.

There was no discussion in this meeting about which would be the most appropriate methods for using with discard survival data, because it will be necessary to examine the available data before selecting the method. However, it was recognized that there would likely be considerable heterogeneity in the available discard survival data, and that this would probably necessitate the use of a rand-effects meta-analysis (see below).

3.5.3.3 Assumptions about the variance structure in the meta-analysis model

Meta-analysis brings together a variety different studies, which will inevitability introduce variability (or heterogeneity) into the observed measures of effect; and the wider the scope of a meta-analysis the greater the potential for heterogeneity. The sources of this variability can, in the case of discard survival, come from the fishing activity being observed (in the population studies, the fishing gear used, the handling procedures, as well as environmental conditions) and from the scientific methods used to make those observations (for examples, the method of survival estimation, the occurrence of captive effects, the use of blind controls, etc.) (see ICES 2014 for further discussion). To obtain a reliable pooled estimate of the measure of effect this heterogeneity needs to be addressed in some way.

At the simplest level, the pooling of the measures of effect may assume that each study is estimating the same intervention effect (i.e. with directly comparable baseline effects and modifiers), in which case a **fixed-effect meta-analysis** is used. If there is heterogeneity in the data, clearly this assumption is undermined. That is, there may be other sources of variability of the survival data, other than intervention effect (e.g. discarding practices), which may be unknown, unique to individual studies or even just random. Therefore, it should be assumed that the different studies are not all estimating the same intervention effect, but that the estimated intervention effects follow a similar distribution across studies; in which case a **random-effects meta-analysis** is performed. These random-effects are typically assumed to be normally distributed around the underlying "fixed" effect. (NB - This is likely to be the most suitable meta-analysis method for many discard survival assessments).

For further discussion on heterogeneity and its role in meta-analysis see Section 9.5 in Deeks *et al*., 2011.

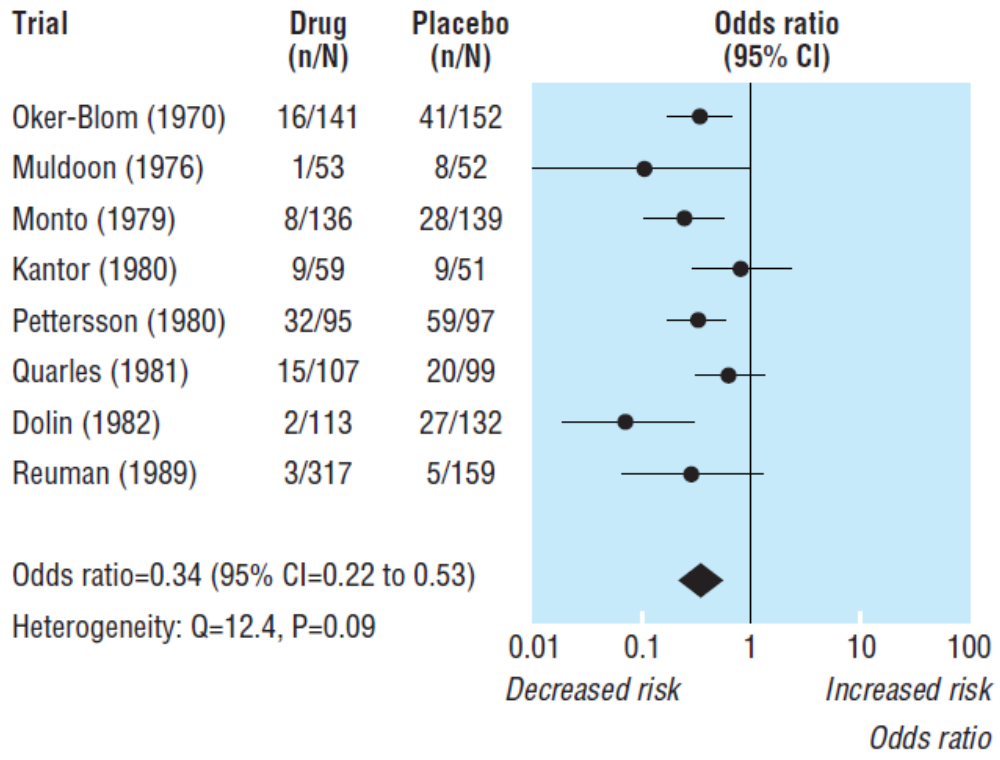

Figure 3.5.3.3.1: An example of the output from a meta-analysis of binary data (Jefferson *et al.*, 2002; reproduced from Higgins *et al.*, 2003)

3.5.3.4 Assessment of the precision of the summary estimate and the significance of the treatment effect

Ultimately, the desired output of a meta-analysis is a single estimate for the chosen measure of effect (with appropriate confidence interval), based on the weighted analysis of the pooled dataset. For example see Figure 3.5.3.3.1, which although a summary of a meta-analysis of medical drug trials, could be seen as analogous to a meta-analysis of discard survival studies, where the drug is the discard treatment (e.g. fishing métier; discarding process) and the placebo is the control. The width of the associated confidence interval of the measure of effect (in this case Odds Ratio; OR) gives a clear indication of the uncertainty associated with the estimate, with wider intervals showing a higher degree of uncertainty, while its location relative to a baseline value (in the case of OR: 1) provides information on the strength of evidence of the treatment effect. Formal hypothesis testing using the standard error of the pooled estimate can be used to define a P value for the null hypothesis (for example the Chi-squared test for the OR).

The estimated effect (and CI) should not be viewed in isolation, however, and it is useful to compare that pooled estimate with the original data (and their confidence intervals). The plot in Figure 3.5.3.3.1 displays the pooled estimate (and CI) of the Odds Ratio in the form of a diamond at the bottom of the plot. Above it, the OR (and CIs) for the individual studies are displayed, alongside the raw data for the treatment and control groups. It is clear that the pooled estimate (from its CI) is more precise than any of the individual studies, moreover there appears to be a significantly decreased risk associated with treatment (although no P value is presented to support this inference).

3.5.3.5 Assessment of the likelihood of non-random variation across studies

There are various tools available that can be used in most meta-analysis methods to determine whether the variation among the results of the separate studies is compatible with random variation, or whether it is large enough to indicate inconsistency of intervention effects across studies. For further details on these methods and how to address heterogeneity in meta-analysis see Appendix 3 (section 6).

With respect to the data arising from the discard survival assessments, it was recognized that these data are likely affected by both heterogeneity and the issue of limited numbers of studies. When conducting a meta-analysis of these data, every effort should be made to explore the nature of the heterogeneity, and address it.

One important aspect of addressing heterogeneity is how the risk of bias, among the included studies and with respect to reporting and publication bias, can be assessed and addressed in the meta-analysis (For more details see Higgins and Green, 2011 – chapters 8, 10, and 12). This was not discussed in detail at this meeting, but will be an important component of the next meeting.

4 A "Fuzzy" Logic Approach to Understanding and Predicting Potential Discard Survival

Quantifying the survival probabilities of 'discarded' organisms during commercial or recreational fishing has received attention for improving predictions of stock status and fisheries management. In Europe, with an exemption rule from the landing obligation under the premise of "high" survival, estimates of discard survival are of considerable interest for both the fishing industry and fisheries managers. However, traditional methods to assess discard fate via captive monitoring or tagging experiments may be too expensive and too difficult to standardize to provide empirical estimates of discard survival at population level for every species of interest. Further, the plethora of (interacting) factors that may induce acute stress, injury and possibly mortality during the catch-and-discarding process in combination with individual susceptibility to cope with stress and injury may complicate attribution of causality. In such cases, where "*causality plays a role but where our understanding of what is actually going on is incomplete*", fuzzy logic expert inference (FLEI) or Bayesian networks (BN) can be used to model risks and conditional probabilities of an event occurring (Charniak, 1991).

These conceptual frameworks allow quantitative predictions (i.e. risk estimates or discard survival probability) when combining scant empirical data with expert judgement. In both FLEI and BN approaches, relevant parameters are selected and their relationships in influencing the likelihood of outcome states (i.e. high, moderate, poor survival) are described either as IF-THEN rules in the fuzzy-logic approach or mapped out as part of influence diagrams with probability nodes which illustrate expected causality in BN. Both FLEI (Cheung *et al.*, 2005) and BN models (Lee and Rieman, 2007) have been used in applied fisheries science. The development of BN models within an ecological setting have been well described by Marcot *et al.*, (2001; 2006) and may provide a "roadmap" to follow when converting influence diagrams of causal relationships to initial BN models which will be the focus of the following sections.

The likelihood of an individual fish to survive the capture-and-discarding process will depend on the severity of stressors and their interactions in causing stress and injury; and species-specific coping mechanisms to alleviate stress and avoid injury. A fish from a given population may perfectly handle the capture-and-discarding process in a certain fishery one day, but on another day under different environmental conditions, the same species may suffer severe death tolls, although the fishing and handling process itself was not altered and the fish came from the same population as the day before. The next day, the fisher may tow the net through a school of fish that were in poor body condition (e.g. just after spawning) and again these fish died 24 hours after being discarded despite mild environmental and unchanged technical conditions.

Survival potential is then the combination of intrinsic sensitivity towards hypoxia or mechanical damage, among others, and the range of measurable stressful factors during fishing. Thus, a BN model to predict survival potential for a given fish species needs to account for the influence of (i) the environment such as salinity and temperature changes, (ii) the fishing and handling processes such as gear deployment duration and air exposure and handling times; and (iii) generic life history and susceptibility/sensitivity traits such as length, sex, hypoxia tolerance, body condition, skin type (scale morphology and presence of a mucus layer) or parasite tolerance (see section 7 "Explanatory variables" in ICES WKMEDS 2014; Benoît *et al.*, 2013). Among technical

and environmental factors, frequently associated with discard mortality were gear configuration, handling, gear deployment duration, water and air temperature and air exposure (p. 51, ICES, 2014).

To construct a BN model requires (1) a list of key influencing variables or stress factors; (2) the possible states or ranges of these discrete or continuous variables, respectively; (3) the dependent relationships between these variables; and (4) conditional probabilities that quantify this dependence. Picturing the dependence between these variables results in an influence diagram whereby each relationship node (parent or child notes) is associated with a conditional probability table (CPT). "*CPT represent the probability or frequency with which a node takes on each discrete state given the states of any antecedent ("parent") nodes that interact with it. CPTs can be derived and updated from empirical information, and expert judgement or case examples*" (Marcot *et al.*, 2001).

It is proposed that, with the support of WKMEDS, a BN model is developed to predict the survival potential of captured-and-released fish based on known relationships between published records of discard mortality and ranges of environmental, technical or biological stress during fishing, in combination with generic life-history traits.

5 Important Dates and Deadlines

ITEM	DATE
Draft guidelines circulated for review	19/06/2015
Comments on draft guidelines report	03/07/2015
Submission of guidelines to ACOM and SCICOM	17/07/2015

6 Conclusions and Recommendations

1. Based on the outputs of this meeting, it is recommended that a scoping exercise is conducted to determine the resources required to undertake a full-scale systematic review of discard survival data for North Atlantic Species. A full systematic review, conducted with input from stakeholders, would help fishery managers assess the benefits for survival assessments for different fishery-species combinations and inform the most suitable design of future survival assessments.

2. It is recommended that funding sources are identified to support a full-scale systematic review of discard survival data for North Atlantic Species.

3. It is recommended that the critical review framework developed by WKMEDS will be made available to the EU Commission and the Scientific, Technical and Economic Committee for Fisheries (STECF), to assist in the evaluation of any proposed exemptions made on the basis of high survival in regional Discard Plans. Along with this, the outcome of critical reviews of relevant papers that have been completed by WKMEDS so far, will be made available in time for the evaluation of the regional Discard Plans in June 2015.

4. Based on those exemptions from the landing obligation awarded on the grounds of high survival from 2016, and results from European studies on estimated discard survival rates, it is recommended that WKMEDS link with relevant stock assessment groups to discuss possible implications of current assumptions on the survival of discarded organisms that are made in ICES stock assessments.

5. It is recommended that, where vitality data provide an effective indicator for survival rates, the potential for ongoing monitoring of vitality as part of the EU Data Collection Framework is explored. Particularly in fisheries where species exemptions have been awarded on the basis of high survival.

6. It is recommended that a formal database is established for storing and processing reviewed discard survival data, the option for ICES to provide this resource will be explored.

7. It is proposed that the next WKMEDS meeting is held in Gent, Belgium, proposed dates 30 November–4 December 2015;

8. To report progress with critical review and meta-analysis;

9. Define the criteria and weighting methods for critical appraisal of reviewed studies; and

10. Expand current database to include more species based on management priorities.

7 References

Arnqvist G., Wooster D. 1995. Meta-analysis: synthesizing research findings in ecology and evolution. *Trends in Ecology and Evolution*, 10: 236-240.

Benoît HP, Plante S, Kroiz M, Hurlbut T. A comparative analysis of marine fish species susceptibilities to discard mortality: Effects of environmental factors, individual traits, and phylogeny. ICES J Mar Sci. 2013;70(1):99–113.

Borenstein M., Hedges L.V., Higgins J.P.T, Rothstein H.R. 2009. Introduction to Meta-Analysis. Chapter 7. Converting Among Effect Sizes. John Wiley & Sons, Ltd (DOI: 10.1002/9780470743386.ch7)

Charniak E. Bayesian Networks without Tears. AI Mag. 1991; 12–50.

Cheung WWL, Pitcher TJ, Pauly D. A fuzzy logic expert system to estimate intrinsic extinction vulnerabilities of marine fishes to fishing. Biol Conserv. 2005; 124:97–111.

Collaboration for Environmental Evidence. 2013. Guidelines for Systematic Review and Evidence Synthesis in Environmental Management. Version 4.2. Environmental Evidence: www.environmentalevidence.org/Documents/Guidelines/Guidelines4.2.pdf

Deeks JJ, Higgins JPT, Altman DG (editors)(2011). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Deeks J.J., Higgins J.P.T. 2010. Statistical algorithms in Review Manager 5. Statistical Methods Group of The Cochrane Collaboration. 11 pp.

Glass G.V., Smith M.L. 1978. Meta-analysis of research on the relationship of class size and achievement. *Educational Evaluation and Policy Analysis*, 1: 2-16.

Gurevitch J., Curtis P.S., Jones M.H. 2001. Meta-analysis in ecology. *Advances in Ecological Research*, 32: 199-247.

Hamilton, M.A. (1979). Choosing the parameter for 2x2 Table or 2x2x2 Table Analysis. Am. J. Epidemiology. 109 (3), 362-275.

Herman J. Ader; Gideon J. Mellenbergh; with contributions by David J. Hand (2008). Methodological quality (chapter 3) in Advising on Research Methods: A consultant's companion. Johannes van Kessel Publishing. ISBN 978-90-79418-02-2.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., Altman, D.G. (2003). Measuring inconsistency in meta-analyses. British Medical Journal. 327, 557-560

Higgins JPT, Green S (eds.) 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

ICES. Report of the Workshop on Methods for Estimating Discard Survival (WKMEDS), 17–21 February 2014, ICES HQ, Copenhagen, Denmark. ICES CM 2014/ACOM:51 2014. 114 pp.

Järvinen A. 1991. A meta-analytic study of the effects of female age on laying-date and clutch-size in the great tit Parus major and the pied flycatcher Ficedula hypoleuca. Ibis, 133: 62-66.

Jefferson TO, Demicheli V, Deeks JJ, Rivetti D. Amantadine and rimantadine for preventing and treating influenza A in adults. Cochrane Database Syst Rev 2002;(4):CD001169.

Lee DC, Rieman BE. Population Viability Assessment of Salmonids by Using Probabilistic Networks. North Am J Fish Manag. Taylor & Francis; 1997 Nov 1;17(4):1144–57.

Marcot BG, Holthausen RS, Raphael MG, Rowland MM, Wisdom MJ. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. For Ecol Manage. 2001;153(1-3):29–42.

Marcot BG, Steventon JD, Sutherland GD, McCann RK. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. Canadian Journal of Forest Research. 2006. p. 3063–74.

Newcombe, R.G. 2013. Confidence intervals for proportions and related measures of effect. CRC Press. Pp442.

Schwarzer, G. 2015. "R" Package "Meta": A General Package for Meta-Anlysis. http://cran.r-project.org/web/packages/meta/meta.pdf

Scott, I (2008). Interpreting risks and ratios in therapy trials. Australian Prescriber, 31(1), 12-16.

Schechtman, E (2002) Odds ratio, relative risk, absolute risk reduction and the number needed to treat – Which of these should we use? Value in ealth 5, 431-436.

Smith M.L., Glass G.V. 1977. Meta-analysis of psychotherapy outcome studies. American Psychologist, 32: 752-760.

Viechtbauer W (2015). R Package "Metafor": Meta-Analysis Package for R http://cran.r-project.org/web/packages/metafor/metafor.pdf).

Appendices

Appendix 1 – Agenda ICES Workshop on Methods for Estimating Discard Survival (WKMEDS II); 20–24 April 2015, Defra, London, UK

Monday 20 April

13.00 Arriving

13.30 Welcome - Thomas Catchpole (TC)

Introductions to each other - all

A brief history of WKMEDS - Mike Breen (MB)

Progress so far, including update Guidance Report - MB

WKMEDS and Terms of Reference for this meeting –MB

1. Critically review current estimates of discard mortality, with reference to the guidelines, and collate existing validated mortality estimates;
2. Conduct a meta-analysis, using the data, to improve the understanding of the explanatory variables associated with discard mortality and identifying potential mitigation measures.

Brief update on EU Policy on the Landing Obligation – Thomas Catchpole (Cefas)

14.30 Agreeing a detailed plan for the week and what we want to achieve – lead TC/MB

15.30 Coffee

16.00 Open session on current survival assessment work – presentations/videos welcome (up to 20 minutes)

Tuesday 21 April (coffee breaks allowed but not scheduled)

09.00 Devising an evaluation framework to critically review current estimates of discard mortality based on the guidelines from WKMEDS

12.00 lunch

13.00 Working in assigned groups to deliver ToRs:

Critically review current estimates of discard mortality, with reference to the guidelines, and collate existing validated mortality estimates – lead TC

Group 1 – flatfish (TAC, European species) – lead Karin van der Reijden

Group 2 – *Nephrops* – lead Sonia Mehault

Group 3 – Skates and rays (TAC, European species) – lead Jim Ellis

Conduct a meta-analysis, using the data, to improve the understanding of the explanatory variables associated with discard mortality and identifying potential mitigation measures – lead MB

15.00 Presentation and discussion with UK Ministry lead on implementation of landing obligation – Dr Sarah Adcock

16.00 Open session on current survival assessment work – presentations/videos welcome (up to 20 minutes)

Wednesday 22 April

09.00 - Plenary session - overview of ToR Groups

1. Critical review –TC

2. Meta-analysis - MB

09.30 Group Sessions

12.30 lunch

13.30 Presentation from Peter Randall (Cefas) on Cefas survival assessments

14.00 Group Sessions – drafting text

16.00 Open session on current survival assessment work – presentations/videos welcome (up to 20 minutes)

Thursday 23 April

09.00 - Plenary session - overview of ToR Groups

1. Critical review –TC

2. Meta-analysis - MB

1000 Group Sessions – drafting text

12.30 lunch

13.30 Group sessions – drafting text

16.00 Open session on current survival assessment work – presentations/videos welcome (up to 20 minutes)

Friday 24 April

09.00 - Plenary session - overview of ToR Groups

1. Critical review –TC

2. Meta-analysis – MB

Summarizing progress, identifying tasks and owners

10.00 Group Sessions – drafting text
13.00 Close

Appendix 2 – Summarized version of meta-analysis data database structure

1	Reference number	15	Maximum tow duration/soak time	29	Max air temperature
2	Information source	16	Mean depth (metres, numbers only)	30	Mean sea surface temperature
3	Species	17	Max depth (metres, numbers only)	31	Min sea surface temperature
4	Area of study (as stated in data source)	18	Min depth (metres, numbers only)	32	Max sea surface temperature
5	Area (regional seas)	19	Time of year (free text)	33	Mean fishing depth temperature
6	Fishing gear type	20	Number of different vessels	34	Min fishing depth temperature
7	Fishing gear as described	21	Number of hauls	35	Max fishing depth temperature
8	Target species	22	Total number of trips	36	Sea state (Douglas sea scale)
9	Treatment description (survival estimates presented by 'trip', 'haul', 'temperature', 'tow duration', 'sorting method' etc.)	23	Seasonal coverage	37	Other environmental variables provided on fishing operation
10	Vessel	24	Haul catch weight mean (kg, numbers only)	38	Catch sorting practice
11	Fishing practice	25	Haul catch weight min (kg, numbers only	39	Description of modified sorting practice
12	Description of modified catching practice	26	Haul catch weight max (kg, numbers only)	40	Health/vigour assessment conducted (vitality score)
13	Mean tow duration/soak time	27	Mean air temperature	41	Health/vigour assessment sample size (number assessed)
14	Minimum tow duration/soak time	28	Min air temperature	42	Health/vigour assessment sample selection

43	Health assessment mean size (cm)	61	Mark–recapture tagging mean size	79	Data storage tagging min size (cm)
44	Health assessment min size (cm)	62	Mark-recapture tagging min size	80	Data storage tagging (DST) max size (cm)
45	Health assessment max size (cm)	63	Mark-recapture tagging max size	81	Captive observation
46	Separate damage or injury assessment	64	Mark-recapture time period of data collection	82	Captive observation - observation period
47	Separate damage or injury assessment sample size (number assessed)	65	Acoustic tagging	83	Captive observation - sample size
48	Separate damage or injury assessment sample selection	66	Acoustic tagging sample size (numb	84	Captive observation - Sample selection
49	Separate damage or injury assessment mean size	67	Acoustic tagging Sample selection	85	Captive observation - mean size (cm)
50	Separate damage or injury assessment min size	68	Acoustic tagging mean size (cm)	86	Captive observation - min size (cm)
51	Separate damage or injury assessment max size	69	Acoustic tagging min size (cm)	87	Captive observation - mean size (cm)
52	Separate reflex assessment	70	Acoustic tagging max size (cm)	88	Analysis method deriving survival estimate 1
53	Separate reflex assessment sample size	71	Acoustic tagging time period of data collection (days)	89	Basis of survival estimate number alive
54	Separate reflex assessment sample selection	72	Data storage tagging (DST)	90	Basis of survival estimate number dead
55	Separate reflex assessment mean size (cm)	73	Data storage tagging (DST) - mean time at liberty	91	Proportion alive
56	Separate reflex assessment min size (cm)	74	Data storage tagging (DST) - min time at liberty	92	Standard error (or uncertainty estimate)
57	Separate reflex assessment mean size (	75	Data storage tagging (DST) - max time at liberty	93	Time (associated with estimate)
58	Mark-recapture tagging	76	Data storage tagging (DST) sample size (number tagged)	94	Identified explanatory variables
59	Mark-recapture tagging sample size	77	Data storage tagging (DST) sample selection	95	Quantified explanatory effect 1
60	Mark-recapture tagging sample selection	78	Data storage tagging (DST) mean size	96	Quantified explanatory effect 2
				97	Quantified explanatory effect 3

Appendix 3 – Protocol for Conducting a Meta-Analysis of Survival Data for inclusion in the WKMEDS Guidance Notes.

1.0 Introduction

Meta-analysis is a statistical technique to summarize the numerical results of a range of different studies and produce a summary statistic (together with its confidence interval), which gives the user a means of comparing the effect of an intervention (in this case discarding) compared with a baseline (or control). This section describes the principles and methods used to carry out a meta-analysis for the main types of data likely to be encountered in survival assessments.

In drafting these proposed protocols this group has relied substantially on the guidelines for meta-analysis provided by the Cochrane Collaboration (http://www.cochrane.org/) and published in their Handbook for Systematic Reviews of Interventions (eds. Higgins and Green, 2011). The primary reason for this is that because the Cochrane Collaboration focuses on the systematic review of medical studies it has very well established protocols for working with survival data.

1.1 Analysis Methods and Software

The Cochrane Collaboration support analysis software called Review Manager (RevMan5) that can perform a variety of meta-analyses. The formulae for all the methods used in RevMan and discussed in the Cochrane Handbook (Higgins and Green, 2011) are provided in Deeks and Higgins (2010).

Analogous to the RevMan software is an R package ("Meta") that includes all of the same methods discussed in the Cochrane handbook, along with some informative examples (Schwarzer, 2015; http://cran.r-project.org/web/packages/meta/meta.pdf).

In addition, a more comprehensive suite of methods is available in the R package ("Metafor"), which also provides additional methods in support of the "Meta" package (Viechtbauer, 2015; http://cran.r-project.org/web/packages/metafor/metafor.pdf).

1.2 Meta-analysis of Binary (dichotomous) Data

Survival assessments use the outcomes of individuals within each study to compare the effects of different interventions (e.g. discarding practices). Meta-analyses focus on pairwise comparisons of interventions, such as an experimental intervention vs. a control intervention, or the comparison of two experimental interventions.

Before conducting a meta-analysis, as part of a systematic review, careful consideration should be given to whether it is appropriate to combine the numerical results of all, or perhaps some, of the studies of interest. Therefore it is recommended that meta-analysis should only be used on data where: studies are related enough to allow meaningful comparison; there is low risk of bias in the primary studies; and where there is no evidence of publication and/or reporting bias (see Deeks *et al.*, 2008, section 9.1.4 for further discussion.

Performing a meta-analysis is a multi-step process, which is highly dependent on the systematic and critical review of the data sources before any analysis takes place. The first step consists of the problem identification (see section 3.2); then, it is necessary to elaborate a strategy for searching and selecting studies (find and read all relevant studies, disregarding those with obvious methodological flaws or that do not report needed information, such as sample size) (section 3.3). The next phase is the extraction of data

and information from the selected studies, and the creation of an appropriate database to be populated with the data (section 3.4). Finally, it is necessary to transform data into a 'common currency', the so called effect size (ES), which represent the input data for the meta-analysis.

Deeks *et al*., (2011) define the following principle elements on which most commonly used meta-analyses are based:

1. Calculating a measure of effect;
2. Calculating a summary (pooled) estimate of effect across studies;
3. Assumptions about the variance structure in the meta-analysis model;
4. Assessment of the precision of the summary estimate and the significance of the treatment effect; and
5. Assessment of the likelihood of non-random variation across studies.

2.0 Defining Measures of Effect

Data or test statistics to be combined in a meta-analysis should be transformed into a common format – i.e. effect size (ES). Discard survival data are considered dichotomous (binary) data, as the outcome can be one of two possibilities, for example, dead or alive. In medical statistics the most commonly encountered effect measures used with dichotomous data are:

- risk difference (RD) (also called the absolute risk reduction);
- risk ratio (RR) (also called the relative risk); and
- odds ratio (OR).

The calculation and interpretation of these measures of effect are discussed below, but for a more detailed description and discussion of these measures of effects for binary data see Hamilton, 1979; Scott, 2008; Newcombe, 2013.

These measures of effect all attempt to account for variations in baseline (i.e. control) survival levels between studies by adjusting the observed treatment survival relative to a corresponding control. Discussions in previous WKMEDS meetings have highlighted concerns over this approach, because it assumes a simple linear relationship between the treatment response and the underlying baseline survival – which may not necessarily be true (ICES, 2014). Although, it is recognized that within a meta-analysis there is a need to normalize the data with respect to baseline mortality across the range of studies and, indeed, this approach may prove informative in identifying the true relationship between the treatment and control mortality. However, it is important to remember that, without an informed insight into the relationship between treatment and control mortality, high levels of baseline mortality simply add to the level of uncertainty associated with the corresponding treatment effects. As such, survival estimates based on such studies, and any derived measures of effect, should be treated with caution.

Often different study designs are used to compute the same effect size. The question of whether or not it is appropriate to combine effect sizes from studies that used different metrics must be considered on a case by case basis (Borenstein *et al.*, 2009). The key issue is that it only makes sense to compute a summary effect from studies that are judged to be comparable in relevant ways.

It might be necessary to apply formulas to convert among effect sizes. Ways of converting data can be found in Borenstein *et al*., (2009). When converting between different measures certain assumptions about the nature of the underlying traits or effects are made (Borenstein *et al.*, 2009). Even if these assumptions do not hold exactly, the decision to use these conversions is often better than the alternative, which is to simply omit the studies that happened to use an alternate metric. This would involve loss of information, and possibly the systematic loss of information, resulting in a biased sample of studies. A sensitivity analysis to compare the meta-analysis results with and without the converted studies would be important.

2.1 Risk Difference (RD)

The risk difference (also known as attributable risk) is the difference between the observed risks (proportions of individuals with the outcome of interest) in the two groups.

Risk Difference = $P_1 - P_2$

The risk difference can be calculated for any study, even when there are no events in either group. The risk difference is straightforward to interpret: it describes the actual difference in the observed risk of events between experimental and control interventions; for an individual it describes the estimated difference in the probability of experiencing the event. However, it can be misleading when comparing the relative benefits of studies with different baseline survival rates. For example, with respect to the subject population, is an improvement in survival from 0.5 to 0.6 ($\Delta = 0.1$) as beneficial as in improvement from 0.1 to 0.2 ($\Delta = 0.1$)? In absolute terms, Δ is the same (0.1) in both cases, but when starting from different baseline survival probabilities an improvement of 0.1 from 0.5 produces only a 20% improvement in survival for the subject population, whereas the same absolute change from a baseline survival of 0.1 actually doubles the survival rate. From this we can see there is real merit in comparing different effects using relative measures, i.e. relative benefit and relative risk.

Table 8.3.2.1.1 – A 2 x 2 Contingency Table and how the associated risks for each group (treatment and control) are calculated.

	Outcome (Death)		Total
Exposure	Yes	No	
Yes	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	N

$$P(\text{Dead} \mid \text{Test}) = P_1 = \frac{a}{a+b}$$

$$P(\text{Dead} \mid \text{Control}) = P_2 = \frac{c}{c+d}$$

2.2 Risk Ratios

These measures of relative effect express the outcome (Death = Risk; Survival = Benefit) in one group relative to that in the other. The risk ratio (or relative risk) is the ratio of

the risk of an event in the two groups (i.e. the proportion of deaths in the treatment: proportion of deaths in the controls).

$$\text{Relative Risk} = \frac{P_1}{P_2}$$

$$\text{Relative Benefit} = \frac{1-P_1}{1-P_2}$$

For both measures a value of 1 indicates that the estimated effects are the same for both interventions. Risk ratios describe the multiplication of the risk that occurs, relative to the baseline, with the application of a treatment. For example, a risk ratio of 3 for a treatment implies that an event (e.g. death) with the treatment (e.g. discarding) is three times more likely than without the treatment (i.e. control). Alternatively we can say that the treatment increases the risk of events by 100 × (RR – 1)% (e.g. 100 x (3-1) = 200%).

2.3 Odds Ratios (OR)

The odds ratio is the ratio of the odds of an event (i.e. the odds of surviving in the treatment: odds of surviving in the controls).

$$\text{Odds Ratio} = \frac{a/b}{c/d} = \frac{a \cdot d}{c \cdot b}$$

As with the Risk Ratios, Odds Ratios (OR) describe the multiplication of the odds of the outcome that occur with use of intervention. But they are more difficult to interpret than RR, to ease their interpretation is useful to convert between the two (see InfoBox 1 – Notes on Odds and Risk).

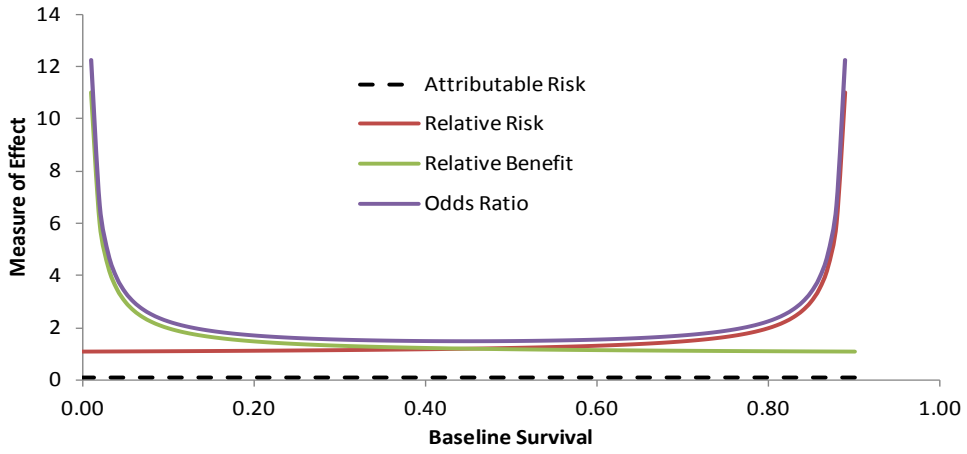

Figure 8.3.2.3.1: A comparison of the values of four different measures of effects for binary survival data (Attributable Risk; Relative Risk; Relative Benefit; and Odds Ratio) based upon the same fixed absolute difference in survival (i.e. Δ = 0.1) over a range of baseline survival (0–0.9).

Note on "Risk" vs. "Odds"

Risk describes the probability with which an outcome will occur. Risk is commonly expressed as a decimal number between 0 and 1, although it is occasionally converted into a percentage.

Odds is the ratio of the probability that a particular event will occur to the probability that it will not occur, and can be any number between zero and infinity. It is commonly expressed as a ratio of two integers. For example, an odds of 0.01 is often written as 1:100, odds of 0.33 as 1:3, odds of 3 as 3:1, etc.

Risk	Odds
0.80	4.0 ⌂⌂⌂⌂/⌂
0.67	2.0 ⌂⌂/⌂
0.50	1.0 ⌂/⌂
0.20	0.25 ⌂/⌂⌂⌂⌂
0.10	0.11 ⌂/⌂⌂⌂⌂⌂⌂⌂⌂⌂

The interpretation of an odds is more complicated than for a risk. The simplest way to ensure that the interpretation is correct is to first convert the odds into a risk.

$$\text{Odds} = \text{Risk}/(1-\text{Risk})$$

$$\text{Risk} = \text{Odds} / (1 + \text{Odds})$$

For example, when the odds are 1:10, or 0.1, one individual will experience the event for every 10 who do not, and the risk of the event is 0.1/(1+0.1) = 0.091. In a sample of 100 specimens, about 9 individuals will have the event and 91 will not. When the odds is equal to 1, one individual will experience the event for everyone who does not, so in a sample of 100, 100 × 1/(1+1) = 50 will have the event and 50 will not.

2.4 Which measure of effect?

When selecting which measure of effect to use, from a statistical perspective, the Odds Ratio is seen to have a number of advantages over both RD and RR (Newcombe, 2013):

1. Irrespective of the baseline level of risk it is always possible to calculate meaningful OR, unlike RD and RR which may encounter ceiling effects. (For example, starting at a baseline risk of 0.4, it is not possible to achieve an absolute difference of greater than 0.6, or a relative increase of more than 2.5).

2. The OR has symmetrical distribution (because it incorporates contributions from both the observed deaths and survivors), while RR tends towards infinity at either high or low baseline survival levels (depending on the underlying outcome, i.e. death or survival) (Figure 8.3.2.3.1). This implies that OR is better suited to providing more balanced comparisons between studies with a wide range of baseline survival rates.

3. The odds ratio is a direct output from logistic regression, as its natural logarithm is identical with the regression coefficient.

However, OR does have some disadvantages and inherent paradoxes, not least of all the difficulty that users may have in interpreting the output (Newcombe, 2013). Despite this, WKMEDS III proposes that Odd Ratio should be considered the primary measure of effect for meta-analysis of survival assessment data. Although, following the guidance of Schechman (2002), it is recognized that it may be informative to compare the output of both an absolute (e.g. RD) and a relative (e.g. OR) measure. Furthermore, Deeks and Higgins (2010) advise using sensitivity analysis to investigate whether choice of summary statistic is critical to the conclusions of the meta-analysis.

Note – when using ratio measures of effect (RR and OR) in meta-analysis, it is important to use the natural logarithm of the measure to ensure that scores along the scale are compared with appropriate symmetry. This ensures that effects of the same magnitude but opposite directions on the ratio scale (for example odds ratios of 0.5 and 2) are interpreted as being equidistant from 1.0.

3.0 Calculating a summary (pooled) estimate of effect across studies

Techniques for the meta-analysis of binary data are available in the R "meta" package (function = "metabin"). These techniques are synonymous with the Cochrane's RevMan software, and so the Cochrane Manual provides very useful support material, including the underlying algorithms detailed in Deeks and Higgins (2010).

The required data are sent to the function in the form of vectors:

- event.e - Number of events in experimental group.
- n.e - Number of observations in experimental group.
- event.c - Number of events in control group.
- n.c - Number of observations in control group.

or as an appropriately structured dataframe. In addition, optional vectors may be included to provide information on potential explanatory variables.

Note – to avoid computational problems where cells have zero counts entered 0.5 is added by default during the meta-analysis modelling. This can be adjusted through a number of arguments, including: incr, allincr, addincr, allstudies and MH.exact.

3.1 Defining Measures of Effect

The "metabin" function will calculate the selected measure of effect for the meta-analysis, which is indicated in the "sm" argument:

- • Risk ratio (sm="RR")
- • Odds ratio (sm="OR")
- • Risk difference (sm="RD")
- • Arcsine difference (sm="ASD")

3.2 Pooling / Weighting the Data

A weighted average of the measures of effect is applied across to the data, where the weighted average is defined as:

$$\text{weighted average} = \frac{\text{sum of (estimate} \times \text{weight)}}{\text{sum of weights}} = \frac{\sum Y_i W_i}{\sum W_i}$$

where Yi is the intervention effect estimated in the ith study, Wi is the weight given to the ith study, and the summation is across all studies. So the more weight given to the ith study, the more it will contribute to the weighted average. The weights are therefore chosen to reflect the amount of (and confidence in) information that each study contains. For example, a typical weighting in meta-analysis is the inverse of the variance of the measure of effect, which reflects the statistical uncertainty associated with each study estimate.

In the R function "metabin", the method for weighting (or pooling) the various studies within the meta-analysis are selected using the "method" argument, and the following options are available: three fixed-effect methods (Mantel-Haenszel, Peto and inverse variance) and one random-effects method (DerSimonian and Laird)(see Table 3.3.2.1). If method is "MH" (default), the Mantel-Haenszel method is used to calculate the fixed effect estimate; if method is "Inverse", inverse variance weighting is used for pooling; finally, if method is "Peto", the Peto method is used for pooling.

Table 3.3.2.1: Summary of meta-analysis methods for dichotomous data (Source: Deeks and Higgins, 2010; Higgins and Green, 2011).

Effect measure	Fixed–effect methods	Random–effects methods
Odds ratio (OR)	Mantel-Haenszel (M-H) Inverse variance (IV) Peto	Mantel-Haenszel (M-H) Inverse variance (IV)
Risk ratio (RR)	Mantel-Haenszel (M-H) Inverse variance (IV)	Mantel-Haenszel (M-H) Inverse variance (IV)
Risk difference (RD)	Mantel-Haenszel (M-H) Inverse variance (IV)	Mantel-Haenszel (M-H) Inverse variance (IV)

3.3 Fixed–effect Methods

The inverse-variance is seen as the generic weighting method and simply uses the inverse square of the standard error (SE) of the measure of effect for each study to estimate the weighting:

$$w_i = \frac{1}{\left(\mathrm{SE}\left\{\hat{\theta}_i\right\}\right)^2}.$$

The Mantel-Haenszel method is the default method in "metabin" and uses a different weighting scheme that depends upon which effect measure (e.g. risk ratio, odds ratio, risk difference) is being used (see Deeks and Higgins, 2010). This approach has been shown to have better statistical properties than inverse variance methods when there are few events (e.g. dead or alive, as in discards survival studies).

Peto's method can only be used to pool odds ratios. It uses an inverse variance approach but utilizes an approximate method of estimating the log odds ratio, and uses different weights. The approximation used in the computation of the log odds ratio works well when intervention effects are small (odds ratios are close to one), events are not particularly common and the studies have similar numbers in experimental and control groups. In other situations it has been shown to give biased answers. As

these criteria are not always fulfilled, Peto's method is not recommended as a default approach for meta-analysis.

3.4 Random-effect Methods

The random-effects method incorporates an assumption that the different studies are estimating different, yet related, intervention effects. This method is primarily based on the inverse-variance approach, making an adjustment to the study weights according to the extent of variation, or heterogeneity, among the varying intervention effects. The random-effects method and the fixed-effect method will give identical results when there is no heterogeneity among the studies. Where there is heterogeneity, confidence intervals for the average intervention effect will be wider, if the random-effects method is used rather than a fixed-effect method, and corresponding claims of statistical significance will be more conservative.

In the R function "metabin", internally, both fixed effect and random effects models are calculated regardless of whether fixed or random methods are selected in the appropriate arguements (c.f comb.fixed and comb.random). Therefore, the estimate for the random effects model can be extracted from TE.random. By default, the DerSimonian-Laird estimate (1986) is used in the random effects model (method.tau="DL"). Other methods available include: the Peto method (sm="Peto") Mantel-Haenszel method (sm="MH") and the inverse variance (sm="Inverse"). The iterative Paule-Mandel method (1982) may also be used to estimate the between-study variance (method.tau="PM"). If R package "metaphor" (Viechtbauer, 2010) is installed, the following methods to estimate the between-study variance τ^2 (argument = method.tau) are also available:

- Restricted maximum-likelihood estimator (method.tau="REML")
- Maximum-likelihood estimator (method.tau="ML")
- Hunter-Schmidt estimator (method.tau="HS")
- Sidik-Jonkman estimator (method.tau="SJ")
- Hedges estimator (method.tau="HE")
- Empirical Bayes estimator (method.tau="EB").

4.0 Assumptions about the variance structure in the meta-analysis model

Meta-analysis brings together a variety different studies, which will inevitability introduce variability (or heterogeneity) into the observed measures of effect; and the wider the scope of a meta-analysis the greater the potential for heterogeneity. The sources of this variability can, in the case of discard survival, come from the fishing activity being observed (in the population studies, the fishing gear used, the handling procedures, as well as environmental conditions) and from the scientific methods used to make those observations (for examples, the method of survival estimation, the occurrence of captive effects, the use of blind controls, etc.) (see ICES 2014 for further discussion). To obtain a reliable pooled estimate of the measure of effect this heterogeneity needs to be addressed in some way.

At the simplest level, the pooling of the measures of effect may assume that each study is estimating the same intervention effect (i.e. with directly comparable baseline effects and modifiers), in which case a **fixed-effect meta-analysis** is used. If there is heterogeneity in the data, clearly this assumption is undermined. Alternatively, it may be assumed that the different studies are not all estimating the same intervention effect, but that the estimated intervention effects follow a similar distribution across studies; in

which case a **random-effects meta-analysis** is performed. These random-effects are typically assumed to be normally distributed around the underlying "fixed" effect. (NB - This is likely to be the most suitable meta-analysis method for many discard survival assessments).

Note - Deeks *et al*., (2011) highlight that a random-effects model does not 'take account' of the heterogeneity, in the sense that it is no longer an issue. They advise that it is always advisable to explore possible causes of heterogeneity, although there may be too few studies to do this adequately.

5.0 Assessment of the precision of the summary estimate and the significance of the treatment effect

Ultimately, the desired output of a meta-analysis is a single estimate for the chosen measure of effect (with appropriate confidence interval), based on the weighted analysis of the pooled dataset. The width of the associated confidence gives a clear indication of the uncertainty associated with the estimate, with wider intervals showing a higher degree of uncertainty, while its location relative to a baseline value (in the case of OR: 1) provides information on the strength of evidence of the treatment effect. Formal hypothesis testing using the standard error of the pooled estimate can be used to define a P value for the null hypothesis (for example the Chi-squared test for the OR).

The estimated effect (and CI) should not be viewed in isolation, however, and it is useful to compare that pooled estimate with the original data (and their confidence intervals). This is conveniently achieved using a "forest plot", which can be generated in R "meta" package using the function "forest" (see Figure 8.3.5.1).

The forest plot in figure 2 displays the pooled estimate (and CI) of the Odds Ratio in the form of a diamond at the bottom of the plot. Above it, the OR (and CIs) for the individual studies are displayed, alongside the raw data for the treatment and control groups. It is clear that the pooled estimate (from its CI) is more precise than any of the individual studies, moreover there appears to be a significantly decreased risk associated with treatment (although no P value is presented to support this inference).

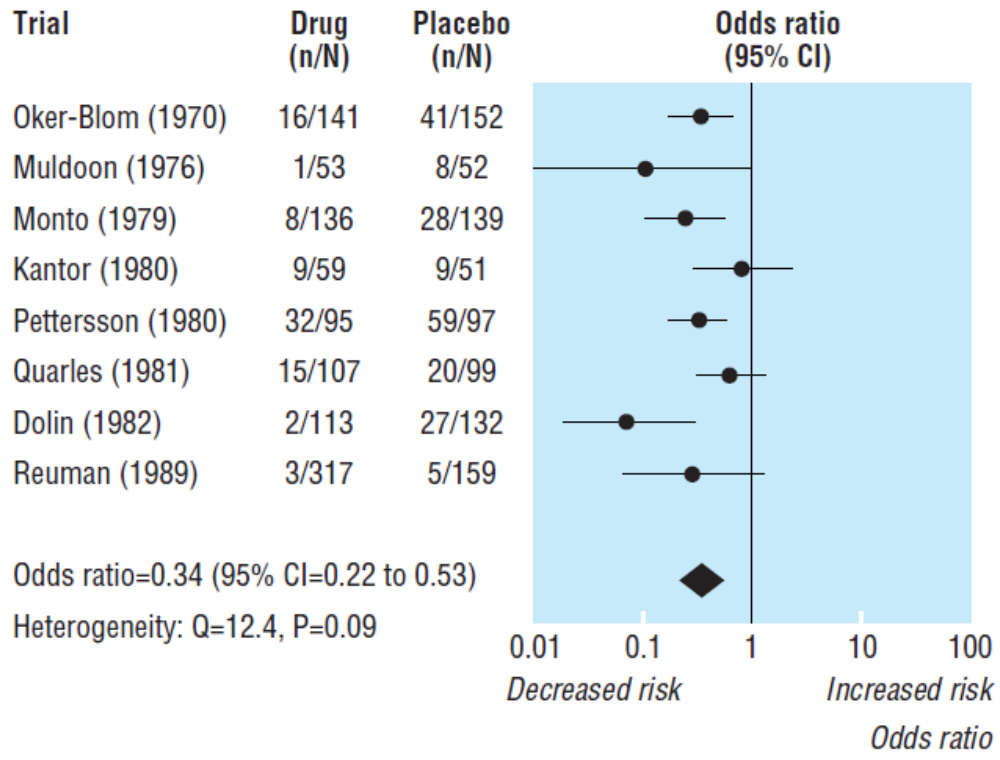

Figure 8.3.5.1: An example of the output from a meta-analysis of binary data (Jefferson *et al.*, 2002; reproduced from Higgins *et al.*, 2003)

6.0 Assessment of the likelihood of non-random variation (heterogeneity) across studies

There are a number of tools available for assessing the likelihood of non-random heterogeneity within the meta-analysis data. The first is simply to compare the measures of effect (and CI) for each of the contributing studies and assess how consist these estimates are. In the example presented in Figure 8.3.5.1, all estimates imply there is a decreased risk associated with the treatment. Moreover, all of the confidence intervals overlap each other, however some of these CIs are particularly wide (in the extreme case covering from 0.01 to 1). This suggests there is reasonable consistency within the data, at least with respect to the direction of the effect, although there is less certainty over the magnitude of the effect.

Various tests are also available, including Q (Chi-squared) and H tests, to determine whether there are genuine differences underlying the results of the studies (heterogeneity), or whether the variation in the data is driven by random chance (homogeneity). Concerns have been expressed about the reliability of these tests for identifying non-random heterogeneity within a meta-analysis dataset, particularly when the number of studies is small (Higgins *et al.*, 2003). However, an alternative measure (I^2) has been proposed that is independent of the number of studies in a meta-analysis (Higgins *et al.*, 2002 and 2003).

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$$

From the example in Figure 8.3.5.1, there is a measure of heterogeneity (Q = 12.4) with an associated P value (P = 0.09). This, and the examination of raw data, suggest there

may be some evidence of heterogeneity within the dataset, but that it is not significant (at least at the 95% confidence level). By inputting the Q value into the above equation we can calculate a value for I^2 of 35.5%.

$$I^2 = 100 \times (12.4 - 8)/12.4 = 35.5\%$$

This also indicates there is no conclusive evidence of heterogeneity within this dataset (i.e. less than the nominal 40% level) (Higgins *et al.*, 2003). Although, Deeks *et al.*, 2011 advise that even with this measure this is some room for interpretation, and the importance of the observed value of I^2 depends on (i) magnitude and direction of effects and (ii) strength of evidence of heterogeneity (e.g. P value from the chi-squared test (Q), and/or a confidence interval for I^2).

In the R package "meta", the output from a meta-analysis can be extracted using the function "print.meta". As well as the estimate of the measure of effect (and it's CI), this function also outputs various measures of heterogeneity (including Q, H and I^2) and with their uncertainty intervals.

6.1 Addressing Heterogeneity

Once heterogeneity has been identified in the dataset, there are a number of possible options for addressing it, and these are discussed in detail in Deeks *et al*., (2011) and Borenstein *et al*., (2009):

1. Check that the data are correct
2. Change the effect measure
3. Explore heterogeneity using subgroup analysis and meta-regression
4. Incorporating heterogeneity into random-effects models
5. Ignore heterogeneity and just use a fixed-effect meta-analysis
6. Exclude studies.
7. Abandon the meta-analysis

With respect to the data arising from the discard survival assessments, it is highly likely that these data will be affected by both heterogeneity and the issue of limited numbers of studies. Every effort will be made to explore the nature of the heterogeneity, and address it. To that end, R packages "meta" and "metafor" provide suitable functions for such analysis, including: "metabin" and "metareg".