UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO IN SANITÀ E PRODUZIONI ANIMALI: SCIENZA, TECNOLOGIA E BIOTECNOLOGIE

DOTTORATO DI RICERCA IN PRODUZIONI ANIMALI

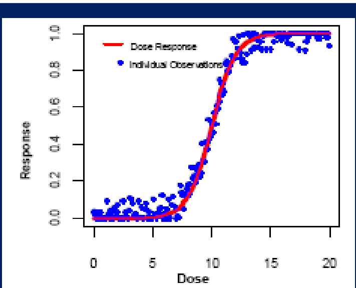
Risk assessment for bovine welfare at slaughter: a comparison between a risk assessment based on empirical data and a risk assessment based on expert opinion

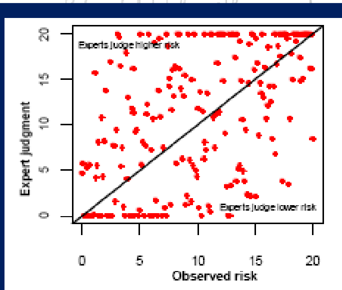
Tesi di: Elisa Aiassa

Docente guida: Prof.ssa Elisabetta Canali

Correlatore: Dr. Olaf Mosbach-Schulz

Ciclo: XXIII





Anno Accademico 2009 - 2010

UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO IN SANITÀ E PRODUZIONI ANIMALI: SCIENZA, TECNOLOGIA E BIOTECNOLOGIE

DOTTORATO DI RICERCA IN PRODUZIONI ANIMALI

Risk assessment for bovine welfare at slaughter: a comparison between a risk assessment based on empirical data and a risk assessment based on expert opinion

Tesi di: Elisa Aiassa

Docente guida: Prof.ssa Elisabetta Canali

Correlatore: Dr. Olaf Mosbach-Schulz

Ciclo: XXIII

Anno Accademico 2009 - 2010

Acknowledgments

Sono particolarmente grata al mio correlatore Dr. Olaf Mosbach-Schulz per il prezioso ed indispensabile contributo alla realizzazione di questa ricerca, la positività ed i continui incoraggiamenti.
Ringrazio la mia docente guida Prof. Elisabetta Canali per il supporto e la disponibilità durante lo svolgimento del lavoro.
Desidero inoltre esprimere la mia sincera gratitudine a tutti i colleghi Medici Veterinari che hanno accettato di partecipare a questo studio con tanta pazienza e dedizione e ai responsabili e tutto il personale dei macelli che mi hanno gentilmente ospitata durante la fase di raccolta dati.
Ringrazio di cuore i miei colleghi per avermi sostenuta nei periodi bui e tra loro in particolare Jane, Saghir e José per il prezioso aiuto e per avermi sempre incoraggiata con tanta solarità.
Un grazie speciale ai miei genitori e alla mia famiglia per esserci sempre e uno ancora più particolare al Griso, mio marito, senza il cui sostegno ed infinita pazienza non avrei mai potuto portare a termine questo lavoro.

Elisa Aiassa

Parma, Febbraio 2011

As for me, all I know is I know nothing

Socrates, *Phaedrus*, sec 235

SUMMARY

Background

Risk assessment is a systematic process that uses modelling to estimate the likelihood of adverse effects occurring from exposure to hazards. While risk assessment is widely used to support decision making in many areas of food and feed safety (e.g. veterinary epidemiology, toxicology, eco-toxicology), it is a relatively new concept for animal welfare. Currently there are no standardised guidelines for animal welfare risk assessment. Furthermore, very little research has been conducted to assess the reliability of the existing methods, which mostly rely on qualitative data and are based on expert opinion.

Objectives

The objectives of this research project were to assess the scientific robustness of existing risk assessment methods for animal welfare and identify any potential methodological flaws of these processes.

Methods

The currently available methods for animal welfare risk assessment were analysed in detail. Two risk assessments for beef cattle at slaughter (in northern Italy) were performed and compared. One based on empirical data (i.e. collected in slaughterhouses) and one based on expert opinion (gathered via a questionnaire submitted to a group of 11 experts).

The two new risk assessments were structured to be as similar as possible to the animal welfare risk assessments under appraisal. A list of 56 hazards potentially relevant to beef cattle at slaughter was produced via a literature review. The relevance of such hazards was assessed by a series of preliminary observations in abattoirs and by asking the 11 experts to assess it. Fourteen hazards were excluded from the subsequent analyses as never being observed during the on-site observations and indicated as not relevant by at least 5 out 11 experts.

For the risk assessment based on empirical data, a novel method for performing on-site exposure assessment and likelihood of the adverse effects (by severity levels) was developed. The method was based on a precise definition of the hazards and a differentiation between adverse welfare effects and measurable indicators of adverse effects. The latter were associated to different severity classes defined qualitatively on the basis of the intensity of the behavioural responses of the animals.

In the second risk assessment the approach to eliciting expert opinion was different from the existing animal welfare risk assessments (based on consensus opinion) as the experts answered the questionnaire independently. Thorough a series of risk assessment-tailored questions, the experts were asked to assess hazard exposure (for beef cattle at slaughter in northern Italy), characterise the adverse effects resulting from the exposure to the hazards and indicate the related uncertainty.

Exposure assessment based on empirical data and on expert opinion was compared for 42 hazards. As 18 hazards were never detected during the on-site observations (or the number of animals exposed was < 5), adverse effect characterisation and final risk were estimated and compared for 24 hazards.

Results

The results of exposure assessment based on empirical data and on expert opinion were inconsistent for 24 out of 42 hazards. Consistent results for all possible adverse effects resulting from the exposure to a hazard never occurred. Often the variability between the experts' responses on exposure assessment and adverse effect characterisation was high. In line with the results of exposure assessment and adverse effect characterisation, the two risk estimates rarely produced comparable results.

Discussion

The analysis of the available methods for animal welfare risk assessment performed in this study, the discordance of the results of the two risk assessments and the variability between the experts' responses highlighted some inherent flaws and requirements of existing risk assessments for animal welfare.

A more detailed and measurable description of the hazards should be

available. Further, a clear understanding of the animal welfare outcomes and their measurement is paramount. In addition, while performing the on-site observations it was clear that interactions between hazards and different hazards intensities and durations need more consideration.
The method developed for performing on-site exposure assessment and estimating the likelihood of the adverse effects proved to be very valuable to solve most of the highlighted limitations of existing animal welfare risk assessments.

Conclusions

A unique and useful approach to defining the hazards for animal welfare and to assessing animal welfare in a measurable and quantifiable way was developed. In particular the method for assessing animal welfare was based on a clear differentiation between adverse welfare effects and measurable indicators of poor welfare (classified by severity levels).
This approach to hazard description and welfare outcome definition and assessment is recommended for enhancing empirical research on animal welfare especially when there is a lack of empirical data for risk assessment. Furthermore, this method can lead to a standardised and harmonised approach for the evaluation of hazards and adverse effects between experts, leading to more robust risk assessments.
This study also proposed an alternative method for eliciting expert opinion based on independent scoring of the risk assessment parameters. The approach showed very useful implications for identifying sources of uncertainties that are normally overlooked in existing risk assessments for animal welfare, such as difficulties in assessing the risk assessment parameters, disagreements between the experts or lack of expert knowledge.
Finally this study highlighted that, independently of the data used for the risk assessment (i.e. empirical data or expert opinion), the method for either reviewing the literature or gathering expert opinion should be chosen in light of the best available practices. The process and any decisions taken should be documented to ensure greater transparency and reproducibility.

TABLE OF CONTENTS

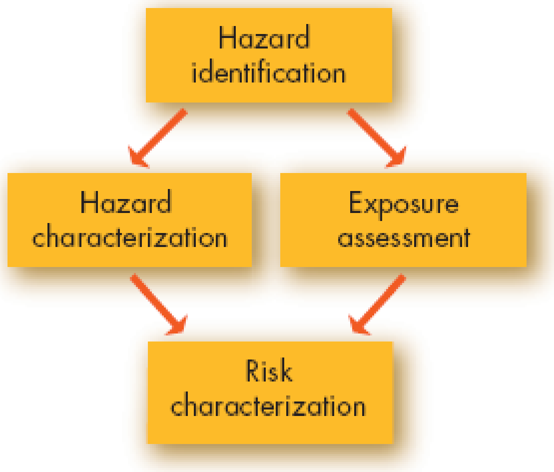
1. PART I. Background and objectives of the project

1.1 What is risk assessment?

Risk assessment is a process intended to calculate or estimate the risk to a given (sub)population, including the identification of attendant uncertainties, relating to exposure to a particular factor (i.e. hazard), taking into account the inherent characteristics of the factor of concern as well as the characteristics of the specific target system (IPCS, 2004). In a generic sense, risk assessment may be considered as a systematic process that uses multi-parameter models[1] for estimating the likelihood of occurrence of adverse effects resulting from exposure to hazards.
The risk assessment process consists of the following steps: i) hazard identification, ii) hazard (or adverse effect) characterisation, iii) exposure assessment and iv) risk characterisation (Figure 1) (CAC, 1999; OIE, 2004a and 2004b). The final objective is to describe each step in a transparent way and provide a quantitative or qualitative statement of the associated risk.

[1] Model: A (simplifying) representation of the essentials (parameters, relations, processes, or mechanisms) of an existing system (or a system to be constructed) which incorporates existing knowledge and/or assumptions about the relationship between all system components in an explicit form that can be investigated by systematic or manipulative experiments (EFSA, 2009a).

Figure 1. Steps of risk assessment (CAC, 1999)



1.2 Why using risk assessment?

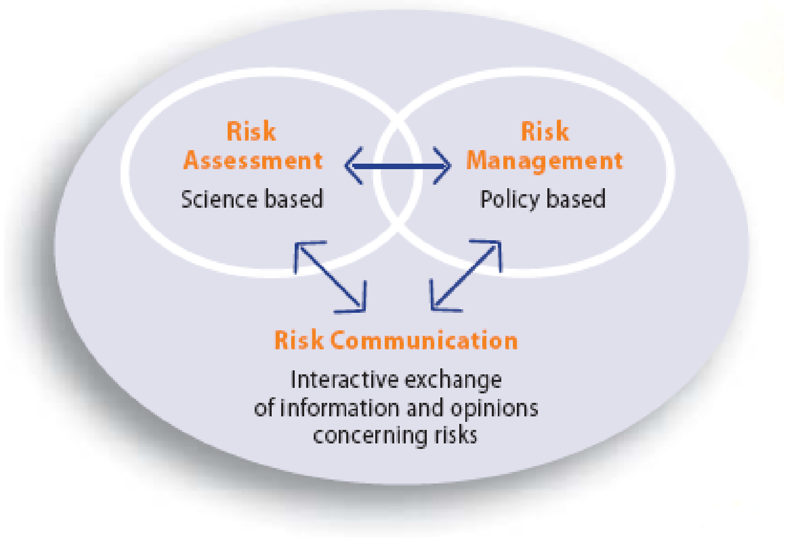
Risk analysis typically includes three parts: risk assessment, risk management and risk communication (Figure 2).
The overall scope of risk assessment is to provide, insofar as possible, a complete set of information to risk managers – so that the most systematic, comprehensive, accountable decision can be made concerning a potentially hazardous situation (K. Asante-Duah, 2002). Essentially, risk assessment provides information on the risk, and risk management is the action based on that information. Risk communication is an interactive process of exchange of information and opinion on risk among risk assessors, risk managers, and other interested parties.
The information developed in the risk assessment will typically facilitate decisions about the allocation of resources for safety improvements and hazard/risk reduction. Also, the analysis will generally provide decision makers with a more justifiable basis for determining risk acceptability, as

well as aid in choosing between possible corrective measures developed for risk mitigation programmes. Furthermore, risk assessment can be used to compare the risk reductions afforded by different remedial or risk control strategies.

Risk assessment is nowadays used to evaluate many forms of new products (e.g. chemical substances, including pesticides; pharmaceutical products, cosmetics; foods and additives; and consumer products); to set environmental standards (e.g. for air and water); to predict the health threat from contaminants in air, water and soils; to determine when a material is hazardous (i.e. to identify hazardous wastes and toxic industrial chemicals); to set occupational health and safety standards; and to evaluate soil and groundwater remediation efforts (K. Asante-Duah, 2002).

In the European Union risk assessment in food safety is under the responsibility of the European Food Safety Authority (EFSA), which deals with assessment of and communication on risks related to food and feed, plant health, animal health and welfare. The European Commission mainly deals with risk management and risk communication.

Figure 2. Risk analysis framework



Source: WHO/FAO

1.3 Quantitative vs qualitative risk assessment

Different models can be used for risk assessments, such as e.g. qualitative, semi-quantitative and quantitative approaches.

Qualitative models describe verbally the risk. The expression of the overall risk probabilities poses a specific challenge, since it has to be ensured that all the parties concerned (risk assessor, risk managers, risk communicators and the public) have the same understanding of the terms such as for example serious or moderate risk. Definitions of these qualitative terms may be useful and appropriate.

Quantitative models can be more transparent because of the numerical format and allow simulations and expressions of distributions of the input variables (their ranges) and risk estimates, but may sometimes give a wrong

impression of the precision without a helpful discussion of the model uncertainties (EFSA, 2009a).
Depending on the information available and the specific question it may be also useful to express information using scores, i.e. on a semi-quantitative scale.
The availability of an adequate and complete information set is an important prerequisite for producing a sound risk assessment and all risk assessment models depend on the quality of input data (rubbish in rubbish out).

Box 1. Different risk assessment models (EFSA, 2009a)

Qualitative Risk Assessment
An assessment that generates an estimate of categorical nature or based on an ordinal scoring system. The outcome of such an assessment is a classification of output into descriptive categories

Quantitative Risk Assessment
An assessment that generates an estimate of a numerical nature directly tied to a measurement or calculation. Depending on the type of model tool used, an indication of the associated uncertainties - expressed either as extreme values, confidence intervals or prediction intervals are needed

Semi-quantitative risk assessment
Within risk assessment, probabilities of an event are assessed and described textually on a scale from negligible, indicating that the probability of an event or the estimated risk cannot be differentiated from zero (and in practical terms can be ignored) to extremely high

1.4 Risk assessment for animal welfare

While risk assessment is widely used in many areas of food and feed safety (e.g. veterinary epidemiology, toxicology, eco-toxicology), in the area of animal welfare this method is relatively new.
The first attempt to use formal[2] risk assessment methodology for animal welfare was made in 2006 by the European Food Safety Authority (EFSA, 2006a, “The risks of poor welfare in intensive calf farming systems”). Since then, several reports using formal risk assessments for animal welfare have been developed by EFSA and other research groups at European level (Table 1 in section 1.5).
The methodology in place follows the approach proposed in the Codex Alimentarius (CAC, 1999), based on the four risk assessment pillars, namely hazard identification, hazard (adverse effect) characterisation, exposure assessment and risk characterisation. Thus, a risk in animal welfare is the result of the probability and magnitude of a negative animal welfare effect (i.e. the adverse effect) consequential to the exposure to a hazard(s). The probability or likelihood of the hazard at a population level is also considered (i.e. the proportion of the population which is exposed). The degree of confidence in the final estimation of risk would depend on the variability, uncertainty, and assumptions identified and integrated in the different risk assessment steps.
Despite animal welfare risk assessment has considerably improved since 2006, up to now there are no standardised processes available or generally applicable guidelines for risk assessment for animal welfare. Moreover, very little research has been conducted to assess the reliability of the existing methods.
Existing risk assessments for animal welfare are frequently based on qualitative data (due to the lack of quantitative data) and expert opinion has

[2] The term “formal RA” is used to exclude the less systematic assessment of the risks that is inherent in a good quality scientific review of any aspect of animal welfare, including animal health.

largely been used for estimating the frequency of exposure to the hazards and the likelihood of an individual being affected or for quantifying the severity and duration of the adverse welfare effects (adapted from Spoolder et al, 2010).

1.5 Objectives of this research project

This research project aimed at analysing existing methods for animal welfare risk assessments (Table 1), which often rely on qualitative data and/or expert opinion, in order to assess their scientific robustness and identify any potential methodological flaws.

Table 1. Existing reports on risk assessment for animal welfare

Report on risk assessment for animal welfare	Reference
The risks of poor welfare in intensive calf farming systems	EFSA, 2006a
Animal health and welfare risks associated with the import of wild birds other than poultry into the European Union	EFSA, 2006b
Welfare of pigs (sows and boars, fattening pigs and tail-biting) (3 reports)	EFSA, 2007a,b,c
Stunning and killing methods for seals	EFSA, 2007d
Animal welfare aspects of husbandry systems for farmed Atlantic salmon	EFSA, 2008
Stunning and killing methods of fish (salmon, trout, eel, sea bass-sea bream, carp, turbot and tuna) (7 reports)	EFSA, 2009b,c,d,e, f,g,h
Welfare assessment of dairy cow welfare (overall effects of farming systems on dairy cow welfare and disease; leg and locomotion; udder; metabolic and reproductive; and	EFSA, 2009j,k,l,m, n

Report on risk assessment for animal welfare	Reference
behaviour) (6 reports)	
Animal Welfare Risk Assessment Guidelines on Stunning and Killing*	Algers et al., 2009
Animal Welfare Risk Assessment Guidelines on Transport*	Dalla Villa et al., 2009
Animal Welfare Risk Assessment Guidelines on Housing and Management*	Spoolder et al., 2010
Broilers (2 reports on broilers: Genetic selection; Housing and Management)**	EFSA, 2010a,b
Harvesting feathers from live geese**	EFSA, 2010c

*These reports represent a first attempt to produce guidance on risk assessment for animal welfare in three systems: stunning/killing, transport and housing

**These reports were published when this project had already started and were not considered in this study

2. PART II. Data collection

2.1 Introduction

To appraise existing risk assessments for animal welfare, the available methods were analysed in detail.
Then two risk assessments for beef cattle at slaughter were performed and compared. One based on empirical data (i.e. data collected in slaughterhouses) and one based on expert opinion (gathered via a questionnaire submitted to a group of 11 experts).
This section illustrates the method developed and implemented in this study for performing hazard identification, exposure assessment and adverse effect characterisation and for collecting data for the two new risk assessments.
The two risk assessments were structured to be as similar as possible to the animal welfare risk assessments under appraisal.

2.2 Observations in abattoirs

To perform the risk assessment based on empirical data in this study, data were collected in abattoir.
In particular three cattle slaughterhouses in northern Italy were identified via the University of Milan, and contacted: two large abattoirs (>3000 and >1000 cattle slaughtered per week); and one small plant (≈80 animals slaughtered per week).
The observations in the abattoirs were performed in three stages:

a. Preliminary meetings with the responsible personnel of each slaughter plant.
b. Preliminary observations.
c. Definitive data collections.

The preliminary meetings were organised to gather overall information on the slaughter process (e.g. the duration; the procedures at arrival at the

slaughter plant; or number of animals slaughtered per day); the characteristics of the facilities (e.g. structure of the unloading bay, passageways, gates, or methodology for restraining and stunning the animals); the breed, age, weight, or gender of the animals slaughtered; and information on the personnel (e.g. frequency of the trainings; or working hours and breaks). The information collected during the meetings helped to define the scenario of the assessment (section 2.4) and to organise the subsequent on-site observations.
The preliminary observations in the abattoirs were performed by the same observer on a sample of 1171 animals in the three slaughter plants (Table 2). At least one preliminary observation of every phase of the slaughter process was performed (i.e. unloading; move along the corridors; passage from the corridors into the stunning box; and stunning and killing[3]). In some phases the animals were observed in group, in some others individually (i.e. one by one), to gain experience in the data collection exercise. The age of the animals observed ranged between 12 and 28 month; the breeds were mainly Charolais and Limousine; the gender included both males and females. In some slaughter plants information on age, breed, and gender was not always provided (Table 2).
The preliminary observations were carried out to further define the scenario and the target population of the definitive assessment (section 2.4); check the relevance of the hazards to beef cattle at slaughter identified via a literature review (section 2.5.2); precisely define all hazards and indicators of poor welfare, in such a way to allow the standardisation of the subsequent definitive data collections; and set the method for performing on-site exposure assessment and adverse effect characterisation (sections 2.6.2 and 2.7.2).
In addition, the preliminary observations helped to reduce the “observer effect” on the slaughter men, who seemed to become gradually used to the data-collection activity and act more naturally in presence of the observer.
As in none of the slaughter plants the use of video recording was allowed,

3 The phase lairage was not considered.

risk assessment-tailored data collection forms were developed using the information gathered during the preliminary meetings and the results of the preliminary observations (Appendix 1).

Table 2. Sample of animals observed during the preliminary on-site observations

Plant	Phase of the slaughter process	N	Gender (N)			Breed (N)			Age	
			M	**F**	**na**	**Char**	**Lim**	**na**	**Mo**	**na**
1	U	205	184	21	0	139	47	19	13 - 28	
	C	211	183	28	0	157	42	12	13 - 20	
	E	278	261	17	0	270	8	0	13 - 24	
	E*	173	173	0	0	0	0	173	14 - 24	
	S/K*	149	105	20	24	0	0	149	15 - 27	
2	U	124	61	8	55	100	24	0	0	124
3	U, C, E, S/K*	31	7	22	0	0	21		12 - 24	
TOT		**1171**	**974**	**116**	**79**	**666**	**142**	**353**	**12 - 28**	**124**

Slaughter plant 1: > 3000 cattle slaughtered per week
Slaughter plant 2: > 1000 cattle slaughtered per week
Slaughter plant 3: ≈80 cattle slaughtered per week
N: number of animals observed. na: not available

C: phase of the slaughter process when the animals move along the corridors. E: phase when the animals are in the area in front of the stunning box and enter it. S/K: stunning and killing phase. U: unloading phase
Char: Charolais. Lim: Limousine
Mo: Months

* In these phases the observer focussed on the animals one by one (individually)

The definitive data collections were carried out by the same observer on a sample of 1427 animals in one of the three slaughter plants (>3000 cattle slaughtered per week) (Table 3).

Table 3. Sample of animals observed during the definitive on-site observations

Phase of the slaughter process	Observation type	N
U	Group	273
U	Group	218
C	Group	190
C	Group	67
C	Group	140
E	Group	268
E	Group	21
E	Group	105
U, C, E*	Individual	6
U, C, E*	Individual	8
U, C, E*	Individual	7
U, C, E*	Individual	9
S/K	Individual	115
TOTAL		**1427**

N: number of animals observed
C: phase of the slaughter process when the animals move along the corridors
E: passage from the corridors into the stunning box
S/K: stunning and killing phase
U: unloading phase
* In these observations each animal was observed from unloading to the entrance of the stunning box

The observations were carried out by phase of the slaughter process:

1. Pre-slaughter handling:
    a. Unloading (U): starting when the animals step on the truck ramp, all along the unloading bay and ending when the animals enter the corridors.
    b. Corridors (C): phase when the calves move along the passageways.
    c. Entrance (E): phase when the animals are in the area in front of the stunning box and enter it.
2. Stunning and killing (S/K): inside the stunning box and during shackling, hoisting and bleeding.

During stunning and killing the animals were observed at individual level due the nature of the process (the animals enter the stunning box one by one).

On contrary, for each of the pre-slaughter handling phases the data collections were performed in two steps: first at group level; and subsequently at individual level.

This was done as in these phases the animals proceed in groups, rendering rather difficult the simultaneous detection of the hazards and the assessment of exposure for all animals observed, as well as the identification of indicators of poor welfare in every individual (section 2.7.2.1).

The results of the two-steps data collections were then analysed, compared and used in the final risk assessment as described in section 3.1.2.

2.3 Questionnaire for gathering expert opinion

To carry out the risk assessment based on expert opinion, a questionnaire on hazard identification, adverse effect characterisation and exposure assessment was prepared and distributed to a group of 11 experts from different regions of northern Italy, contacted via the University of Milan.

The experts were all veterinarians responsible for monitoring the slaughter process for cattle in various abattoirs or inspectors from the Public Health Service. It was not possible not gather information on any potential

additional expertise, e.g. if they had a background in ethology or knowledge of risk assessment methods. The gender of the experts was not considered as proved to be not relevant (Bracke et al., 2008).
In existing risk assessments for animal welfare expert opinion is always elicited within a group of experts who reach a *consensus* on the various estimates (e.g. on the frequency of exposure to the hazards; the likelihood of an individual being affected; or the quantification of the severity of the adverse welfare effects). The scientific arguments underpinning the consensus opinion are frequently not reported.
Differently from the existing animal welfare risk assessments, the experts replied to the questionnaire *independently*. Therefore, the questionnaire was accompanied by a detailed explanation of each step of the risk assessment process; a thorough description of the target population and the scenario under assessment (section 2.4); and technical instructions on how to compile the questionnaire. Upon request, phone calls were also made to clarify specific unclear aspects. However, during the calls attention was paid not to influence the answers.
The questionnaire, translated in Italian, was in Excel, divided in two parts (two spreadsheets), one for hazard identification and exposure assessment (part 1) and one for adverse effect characterisation (part 2). The method for performing hazard identification, exposure assessment and adverse effect characterisation in the risk assessment based on expert opinion is illustrated respectively in sections 2.5.2, 0 and 2.7.3.

2.4 Target population and scenario of the assessment

A risk assessment in animal welfare starts with a clear definition of the target population and the scenario subject to the assessment.
In this study these aspects were defined on the basis of the preliminary meetings and observations (section 2.2).

The target population was beef calves of about 1-2 years[4], both males and females. All beef breeds slaughtered in northern Italy were included (e.g. Charolais, Limousine, cross breeds, etc).

The characteristics of the scenario of the assessment are illustrated in Table 4.

When defined, the target population and the scenario of the assessment were exactly the same in the two risk assessments (i.e. the one based on empirical data and the one based on expert opinion), in order to allow the comparability of the results.

[4] Older cattle and adult milk cows at the end of the career or calves less than 1 year were not considered.

Table 4. Scenario of the risk assessments based on empirical data and expert opinion[5]

Item	Description	Comment
Geographical area	Northern Italy	Area where the on-site observations took place
Time of the year	Summer and autumn	This is the period when the on-site observations took place
Duration of transport (time farm-slaughter plant)	Non considered	Although acknowledging that the transport duration from farm to abattoir may influence the baseline welfare status of the animals at slaughter, it was not always easy to receive this information in the abattoirs. Thus for practical reasons this aspect was not considered (and hazards e.g. "too long transport time" were not included in the list of hazards – Appendix 2)
Duration of waiting time from arrival to unloading	Non considered	Same consideration as for transport duration. In addition, during the preliminary observations of the animals waiting for being unloaded, it was difficult observe the animals during this phase. Thus hazards e.g. "too long time between arrival and unloading" were not considered in the list of hazards (Appendix

5 Defined after the meetings at the abattoirs and the preliminary on-site observations on a sample of 1171 animals.

Item	Description	Comment
		2)
Duration of the slaughter process	The RA focussed on a slaughter process which: ➔ starts when each animal steps onto the truck ramp to get out of the truck ➔ ends when the animal is bled out	In previous works on risk assessment for animal welfare at slaughter (Algers et al., 2009) the unloading phase was not considered and the risk assessment started when the animals were already in the unloading bay. In this research project, after the preliminary on-site observations, it was decided to start the risk assessment when the animals stepped onto the truck ramp as unloading is an important part of pre-slaughter handling, where hazard(s) and adverse effect(s) may occur not only due to the transport conditions, but also to the slaughter plant (e.g. the personnel of the abattoir helping during unloading, or the structure of the unloading bay). On the other hand, the time for unloading and any hazard(s) and adverse effect(s) occurring on the truck during the unloading phase but before the animals stepped onto the ramp were not considered as difficult to detect and measure during the on-site observations
Lairage	Not considered	In the slaughter plants observed lairage was not common

Item	Description	Comment
	The RA focussed on a slaughter process in which the animals are handled towards the stunning box right after unloading, without any waiting periods in the holding pens	practice for 1-2 years beef calves
Stunning method	Stunning box and captive bolt	Most common method used in northern Italy
Overall duration of the slaughter process	Not indicated, assuming that the process is continuous from unloading to jugulation, without waiting time in the holding pens	In some risk assessment for animal welfare the duration (D) of the process under assessment is indicated, since D of the adverse effects occurring in the animals is measured considering the overall D of the process. In this research project, D of the adverse effects was calculated using a slightly different approach based on a 1-3 scale and in the risk assessment based on expert opinion only (section 2.7.3)

2.5 Hazard identification

Hazard identification (HI) takes place before starting the risk assessment and is the process to *identify*, *describe* and *clearly specify* changes in features (i.e. hazards) that can impair the welfare of the animals under consideration and in the given scenario, causing adverse effect(s).

2.5.1 Hazard identification in existing risk assessments for animal welfare

In existing risk assessments for animal welfare, the hazards that are relevant to the scenario (e.g. slaughter) and the population subject to assessment (e.g. beef cattle) are *identified* and *described* by reviewing the available scientific evidence. In the animal welfare field quantitative data often lack and frequently qualitative information or expert opinion are used. If relevant to the scenario and the target population, the hazards should also be *clearly specified* in terms of their *duration* and *intensity*. Hazard duration and intensity combined represent the *magnitude* of the hazard. The definition of hazard magnitude sets the input (or dose) which should be assessed with regard to its adverse effect(s) (or response) in terms of animal welfare. However, hazard intensities and durations are not always addressed in existing animal welfare risk assessments (Algers et al., 2009).
When identified, the hazards are inserted in risk tables that are then scored by groups of experts who reach a consensus on exposure assessment (i.e. hazard occurrence) and adverse effect characterisation (i.e. likelihood and magnitude of the adverse effects consequent to the hazards).
In most of the existing risk assessments for animal welfare the hazards listed in the risk tables are not thoroughly described and they are open for interpretation; in addition, the underlying evidence is frequently not reported and/or it is not indicated if and when expert opinion was used (Bracke et al., 2008; Algers et al. 2009; Spoolder et al, 2010).
The welfare of the animals may be impaired by hazards occurring at the same time and interacting among each others (e.g. a poor operator activity interacting with an inadequate facility). However, the question of how to

deal with interactions between hazards in animal welfare risk assessment has not been solved (Bracke, 1999; Spoolder et al, 2010) and interactions between hazards are normally not addressed in risk assessments for animal welfare.

2.5.2 Hazard identification in this research project

For the purposes of this study a comprehensive list of hazards to beef cattle at slaughter was produced by reviewing some major reports: EFSA (2004, 2006c); OIE (2005a, 2005b); Algers et al. 2009; and Grandin, 2010).

These reports represent the most comprehensive available works on animal welfare at slaughter and were considered sufficient for the scope of this study, which was the validation of a method and not the performance of a complete risk assessment or the assessment of beef cattle welfare at slaughter.

For the hazards related to the unloading phase (e.g. handling of animals during unloading and requirements for trucks) the report of Dalla Villa and colleagues (Dalla Villa et al., 2009).

The hazards were classified in hazards due to the management (e.g. “Inadequate handling technique”); the trucks (e.g. “Steep truck ramp”); or changes in the facilities and the situation (e.g. “Inadequate temperature” or “Dirty floor”). All hazards identified are reported in Appendix 2. For the risk assessment based on data collected on-site, the relevance of the hazards identified via the literature review was assessed during the preliminary on-site observations in the slaughter plants. Such observations highlighted the need for a more detailed description of those hazards that in the literature are described in a qualitative way (e.g. “noise”), in order to make them observable and measurable (Appendix 2).

For the risk assessment based on expert opinion, in part 1 of the questionnaire the experts were asked to assess the relevance of the hazards identified via the literature review.

Since the questionnaire was answered by the experts *independently*, an explanation of the hazards was provided when it was believed that misunderstanding may occur; however, the interpretation was left to the

experts for hazards that in the literature were described in a qualitative way (Appendix 2).
The experts were asked to provide a justification for any hazards that in their opinion should *not* be included (i.e. when they did not agree on the hazards identified via the literature review) and to integrate the list of hazards with any other hazard(s) they believed should be included and had not been considered.
Slaughter is by its nature a very limited process in time (especially if lairage is not considered) and thus animals' exposure to the any hazards occurring during slaughter, even to hazards due to the facilities, is rather short. For this reason, in both risk assessments performed in this research project, hazards' *duration was not considered*. For example, on farm a dirty floor covered with an excess of manure (hazard) may determine different adverse effects in the animals depending on the duration of the dirtiness (if it lasts e.g. for 5 days or 5 months) and thus the duration of this hazard should be considered when performing hazard identification. On contrary, in the slaughter process the animals would move in a rather fast flow towards the stunning box and thus hazards' duration and its impact on the welfare of the animals was considered as having no implications on the adverse effects in the animals.
As in existing animal welfare risk assessments, in the two risk assessments of this study *different hazards' intensities* or *interactions* were *not* considered. When detected during the on-site observations, interacting hazards were counted separately as illustrated in section 3.1.1.

2.6 Exposure assessment

Exposure assessment consists of the quantitative and/or qualitative evaluation of the frequency of exposure to the hazards in the *animal population*, or, in other words, likelihood of hazards to welfare occurring in a given animal population.

2.6.1 Exposure assessment in existing risk assessments for animal welfare

In existing risk assessments for animal welfare the likelihood of each exposure to the hazards is estimated specifying minimum, most likely and maximum values. The estimate is very often based on expert opinion as data on exposure to hazards to the welfare are often not available (Algers et al., 2009). Normally a qualitative assessment of the *uncertainty* related to exposure assessment is also estimated (section 2.8).

2.6.2 Measurement of exposure in the risk assessment based on data collected in abattoir

The method for performing exposure assessment based on empirical data was developed during the preliminary observations in 3 abattoirs.

Most of the hazards described in the literature *in a qualitative way* and thus difficult to measure objectively during the on-site observations (e.g. the hazards “noise” - when does a sound become a “noise”?; or “slippery floor” - how to define and measure “slipperiness”?) were measured on the basis of the experience gained during the preliminary observations.

For instance, “Dirty floor” was defined as “dirtiness (due to e.g. excess of manure) which at the preliminary observations clearly induced at least one animal to show at least one indicator of poor welfare" (see section 2.7.2.1 for indicators of poor welfare). The underlying assumption was that the observer was equally able to detect the same hazard in all following observations. Thus in the example of “Dirty floor”, it was assumed that during the definitive on-site data collections the observer was able to detect the same level of dirtiness, although maybe not causing any adverse effects to the animals. This was essential for performing exposure assessment, which consists of estimating the occurrence of the hazards independently of whether they cause an adverse effect or not.

The data collections of the pre-slaughter handing phases (i.e. unloading, corridors and entrance of the stunning box) were carried out in two steps (i.e. group and individual observations).

Some hazards were counted differently dependent upon the observation type. For instance, the hazard "sudden noise" in the group observations was counted as the *number of finished noise(s) detected*, whereas in the individual observations was measured as *the number of animals exposed to the noise(s) at least once*.

Furthermore, in the *group observations* of the pre-slaughter handling phases, the hazards were counted by different *units of observation*, considering that the cattle could be exposed to the hazards in groups, pairs/triples or individually (Table 5).

In the *individual observations*, every hazard detected was equal to "one animal exposed to the hazard", being "individual" the only unit of observation used. In the stunning and killing phase the animals were observed individually only, due to the nature of the process (the animals are stunned and jugulated one by one).

The above aspects are described in Appendix 2. List of hazardsand were considered in the analysis and use of the results of the two types of observation (section 3.1).

During the on-site observations the hazards detected were inserted in the data collections forms, which were different for the group and individual observations (Appendix 1).

Table 5. Units of observation for hazards in the *group* on-site observations*

Unit of observation	Explanation	Example
Group	Animals on the same truck or group of animals present in the corridors at the same time. When detected, these hazards were multiplied by the number of animals in that specific group	"Steep truck ramp", was multiplied by the N of animals on that truck "Too many animals in the flow" was multiplied by the number of animals present in the group exposed to the hazard
Pairs or triples	Two or three animals standing one next to the other or nearby in the flow. When detected, these hazards were multiplied by 2 or 3 respectively, assuming that when occurred at group level they affected respectively 2 or 3 animals simultaneously	"Inadequate handling technique" (pairs) "Noise" (triples)
Individual	Number of animals exposed to the hazard at least once. The number of hazards detected corresponded to the number of animals exposed to them	"Energised prod" counted as "number of animals prodded at least once"

*Carried out for pre-slaughter handling phases (i.e. unloading, corridors and entrance of the stunning box)
All hazards and related units of observations in the group observations are illustrated in Appendix 2

Since the definitive on-site observations were performed in one slaughter plant only, to allow the comparability of the two risk assessments the abattoir-related hazards (i.e. the “fixed” hazards due to the facilities) were not analysed[6].

Only the variable hazards were kept in the analysis, like those due to the personnel (e.g. “Inadequate handling technique”); the trucks (e.g. “Steep truck ramp”); or due to variable situations (e.g. “dirty floor”, which was often dependent upon the number of animals that had walked on the floor prior to the observation).

The decision on which hazards to analyse was made also on the basis of the preliminary observations in the 3 slaughter plants and the information gathered during the preliminary meetings with the responsible persons in the 3 abattoirs.

For instance it was noted that the trainings to the personnel would follow the same structure in the various slaughter plants, implying that any hazards due to the slaughter men would be due to the variability of the situations and not to different trainings.

Also, all observations were undertaken in absence of the inspector personnel, which excluded the possibility that the data collections may have been in some way biased by more (or less) severe controls.

The hazard “Use of sharp driving tools” was excluded from the analysis as it was forbidden in the slaughter plant where the definitive on-site observations occurred. As opposite, the hazard “Energised prod” was analysed as subject to the same internal rules in all slaughter plants visited (i.e. the prod should be as less used as possible and should never be applied on sensitive parts like head, legs, etc).

Although it may be considered abattoir-related, the hazard “Overall and continuous noise at the entrance of the stunning box” was included it in the assessment as was observed in all slaughter plants visited (the noise is

[6] Not all hazards due to the facilities were considered slaughter plant-related. For instance “too big/too small gates and corridors’” were kept in the analysis as depending upon the size of the animals. “Sharp barriers” or “faulty operating gates” were also not excluded as they may change due e.g. to damages.

present due to the nature of the captive bolt stunning process, i.e. the animals suddenly collapse and the material of stunning boxes). On the other hand, the same hazard in the corridors was excluded from the analysis as its presence would depend on the layout of the plant (e.g. in some abattoirs the corridors are separated from the stunning area by a barrier).

Some hazards were not included in the analysis as for different reasons it was not possible to assess them in the on-site observations.

Appendix 2 provides the list of all hazards identified; how they were defined for the on site observations; the related unit of observation for in the group observations; how they were asked in the questionnaire; and if and why they were excluded from the analysis.

2.6.3 Exposure assessment in the risk assessment based on expert opinion

In the risk assessment based on expert opinion the exposure was assessed by asking the experts to estimate, for each hazard identified as potentially occurring to beef cattle at slaughter, the minimum, most likely and maximum values of occurrence.

The question was posed in a way to avoid asking percentages and facilitate the reply, as follows:

"*Imagine a group of 100 beef calves slaughtered in northern Italy. Please tell, using a number from 0 to 100: the average, minimum and maximum number of animals that are normally exposed to each hazard listed*"

The hazards as described in the questionnaire are reported in Appendix 2.

2.7 Adverse effect characterisation

An adverse effect is the consequence of the exposure to the hazard(s), in terms of change in individual's welfare. It can also be defined as the failure of the animal to cope with the hazards.

Adverse effect characterisation consists of describing for all hazards indentified the related *adverse effect*(s); their *severity* and *duration* (i.e. *magnitud*e); and their *likelihood* to occur at individual level.

2.7.1 Adverse effect characterisation in existing risk assessments for animal welfare

One of the most crucial parts of animal welfare risk assessment is to express in quantitative terms the degree to which the animals fail to cope with the specific hazard(s), i.e. the adverse effects consequent to the hazards. Failure to cope might result in disease or injury, but also in adverse states like pain, fear and frustration. These are all different components of welfare which together, in some way, contribute to the overall level of welfare.

When setting out to assess and measure welfare, however, one runs into the difficulty of comparing values on different scales. Moreover, different hazards affect generally specific components of welfare and e.g. impairment with respect to one component of welfare is generally not accompanied by the same deterioration in all other aspects. The list of possible situations with incongruent welfare aspects is in fact practically endless. An animal might suffer from pain but still experience happiness, or it can be perfectly healthy but still feel frustration.

If some quantitative figures are available for likelihood of adverse effect, there still remains the problem of quantification of the severity and duration of the adverse welfare effects and at present these parameters are mainly based on expert opinion (adapted from Spoolder et al. 2010).

In the existing risk assessments for animal welfare, the following situations have been observed with regard to the description and estimate of the magnitude (i.e. combination of severity and duration) of the adverse effects:

- Often the difference between adverse effects and measurable indicators of adverse effects is not clearly or not at all expressed. The adverse effect is what actually happens in the animal when exposed to the hazards (i.e. the failure to cope with the hazards), whereas the indicator is a measurable variable of the adverse effect. For instance, if exposed to high temperature an animal will suffer from thermal distress (i.e. adverse effect), which may be indicated by panting, sweating and dehydratation (i.e. measurable indicators of thermal distress).

This differentiation is not clearly expressed in most of the existing animal welfare risk assessments, making it difficult to interpret how the experts characterised the adverse effects, especially in terms of severity and duration.

For instance, in a report on risk assessment for pig welfare the adverse effects “stress and fear” and “stress” were listed together with the indicators of adverse effects “leg injuries”, “lameness”, without explaining how these events were quantified in terms of duration and severity (EFSA 2007c).

The difference between adverse effects and measurable indicators of adverse effects has partially been addressed in three recent reports on risk assessment for animal welfare (Dalla Villa et al. 2009; EFSA, 2009k; Spoolder et al., 2010). However, the adverse effects and the indicators of adverse effects were not described in detail and it is not clear how they were characterised in terms of severity and duration (see the next bullet point).

- The *severity* of the adverse effects is described on the basis of the level of behavioural responses demonstrating pain, frustration, fear, anxiety, changes in adrenal or behavioural reactions, using a 1-4 scale (i.e. limited, moderate, severe, critical). *Duration* is expressed as the period of time during which an animal is believed or expected to be experiencing the adverse effect once exposed to the hazard, and may be estimated on a continuous scale, e.g. expressed in minutes (e.g. Algers et al., 2009) or on a scale from 1 to x, indicating the time that the animal is believed or expected to be experiencing the adverse effect, once exposed to the hazard (e.g. Dalla Villa et al. 2009).

  Despite these explanations on how severity and duration are estimated, frequently the adverse effects or indicators of adverse effects are poorly described or grouped together under the same severity and duration class with no further justification.

For instance, in one report the adverse effects "Bruises, wounds, fractures" were grouped together under the same severity levels (either 2, "moderate" or 3 "severe" depending on the hazard causing them) leaving it unexplained how a wound can present the same severity and duration of a fracture (Dalla Villa et al. 2009).

In the same report other adverse effects were not defined in detail (e.g. "difficult movement" or "disorientation"), making it difficult to interpret how they were estimated by the experts in terms of severity and duration. In another risk assessment the same adverse effects were repeated for all hazards occurring in the pre-slaughter handling phases (i.e. "balking, aggression, bruising, wounds, fractures, crowding") and in almost all cases were given a severity level equal to 2 (i.e. moderate) (Algers et al., 2009).

Again in other reports (EFSA 2007a, b and c), the description of the adverse effects in the risk tables was not detailed and open for interpretation (e.g. "frustration ad stress").

- The magnitude of the adverse effects (response) results from the magnitude of the hazards (dose) and often may not be quantified directly using one single indicator but indirectly through cascades of indicators. For instance, if an animal is exposed to an energised prod (hazard), it may vocalise, balk or get wounded at the same time. This is also not clearly addressed in existing risk assessments for animal welfare, where just one adverse effect/indicator (or undefined group of adverse effects/indicators) per hazard is reported.
- Linearity and continuity of adverse effects is often assumed, because it is also technically easiest to deal with. This means that it is often implied that, e.g. severe pain (level 3) means a two times greater reduction in overall welfare than mild pain (level 1), comparing with no pain (level 0). This is of course a simplification. In reality, the association between pain and overall reduction in welfare might be anything else but linear and continuous. Unfortunately however, reliable data to support a specific type of relationship usually are lacking.

In existing animal welfare risk assessments the *likelihood* of the adverse

effects (given that the animal is exposed to the hazard) is estimated specifying minimum, most likely and maximum values.
The *uncertainty* about the adverse effects, their magnitude and likelihood is also assessed (section 2.8).
Adverse effect characterisations based on empirical data and on expert opinion were performed in this study taking into consideration the structure of the risk assessments for animal welfare under appraisal.
However, some adaptations were made to comply with the specific requirements of the on-site data collections and the questionnaire submitted to the experts. This is described in the following sections.

2.7.2 Adverse effect characterisation in the risk assessment based on data collected in abattoir

The main problem inherent to adverse effect characterisation based on empirical data was to measure cattle welfare at slaughter, i.e. to detect the adverse effects due to the hazards and simultaneously assess their severity and duration.
The following method was developed and applied:

- *Use of measurable indicators of adverse effects* (section 2.7.2.1). This approach considered that in the context of the relatively short slaughter process, the most relevant *adverse effects* seem to be negative emotional states like pain, distress, fear and frustration, anxiety, changes in adrenal or behavioural reactions as well as physical injuries (adapted from Algers et al., 2009). Depending on the severity of these statuses, the situation may range from a complete absence of them to a condition that, if persisting, would be life-threatening. These adverse effects and their intrinsic severity are difficult to measure, especially at slaughter. Thus for the on-site data collections a series of animal-based *measurable indicators* of adverse effect(s) was identified and used (section 2.7.2.1, Table 6).
- To quantify the *severity* of the adverse effects, the indicators of poor welfare were associated to specific severity levels, in such a way that every time that an indicator of poor welfare was detected, the intrinsic

severity of the related adverse effect was also estimated (section 2.7.2.2 and Table 7).

- The *duration* of the adverse effects was not measured, for the reasons explained in section 2.7.2.3.
- The use of the indicators of the adverse effects permitted the measurement of the *likelihood of the adverse effects by severity levels* (the duration was not estimated), as described in section 2.7.2.4.

The method proved to be very effective despite the constraints of the on-site observations (e.g. the speed of the process; the layout of the slaughter plants; the fact that in some phases the animals proceed in group; and the not-authorised use of video recording).

2.7.2.1 Indicators of adverse welfare effects

The indicators of poor welfare at slaughter used in this study are illustrated in Table 6. They were defined using and adapting the approach described the Welfare Quality® project (WQ®, 2009) and the method for auditing bovine welfare at slaughter proposed by Grandin (2010).

The Welfare Quality® project proposes a scoring method for the assessment of behaviours when the animals are driven into the stunning box (such as fear behaviours: running; moving backwards; turning around; freezing/balking; vocalising; and injurious behaviours: slipping, falling, stepping on other mates); a method for assessing a poor or undefined stun, based on specific symptoms; a review of the scoring methods for lameness and a method for scoring it; and an assessment system for skin alterations in cattle (developed for on-farm welfare assessment).

The difference between the Welfare Quality® approach and other monitoring systems is that it focuses on animal based measures (e.g. directly related to animal body condition, health aspects, injuries, behaviour, etc.) instead of on resource based measures (e.g. size of a cage, feeding space, etc.), posing the attention on the “outcome” of the

interaction between the animal and its environment[7].
Grandin (2010) suggests the use of practical scoring systems at slaughter (e.g. the energised prod, the falls, or symptoms of effective stunning).
In this study all above scoring systems were integrated with the information gathered during the preliminary observations, in a way to precisely define each welfare indicator and make it objectively measurable and distinguishable from the others. Where necessary precise start and end times (or total duration) of the indicator were described (e.g. the indicator "hesitating" referred to the animal stopping and faltering, lasting less than 3 seconds) (Table 6). The preliminary observations served also to the observer to become familiar with the list of indicators and their use, enhancing the repeatability of the data-collections.
Easy-to-remember codes were assigned to the indicators, in order to obtain good-format data collection forms and enhance the data collections in the abattoir (Appendix 1).

[7] The Welfare Quality® system comprises four welfare principles (i.e. good feeding, good housing, good health and appropriate behaviour). Each of these four principles corresponded to several criteria, with an overall total of 12 criteria. Each criterion represents a separate aspect of animal welfare: 1.Absence of prolonged hunger; 2.Absence of prolonged thirst; 3.Comfort around resting; 4.Thermal comfort; 5.Ease of locomotion; 6.Absence of injuries; 7.Absence of disease; 8.Absence of other pain; 9.Expression of social behaviours; 10.Expression of other behaviours; 11.Good human-animal relationship; 12.Absence of fear.

Table 6. Measurable indicators of adverse welfare effects for beef calves at slaughter[8]

Indicator	Description	Code[9]
Attempt to escape	The animal attempts to jump over the barriers by putting at least one leg over them and return to the standing position	**Esc**
Balking (or freezing)	The animal stops and refuses to proceed for more than 3sec, in spite of the handler physically trying to drive it forward. An animal that stops but continues to walk when the handler drives it forward is not balking	**Balk**
Bruises/ wounds	Any bruises or wounds occurring during the slaughter process (only skin lesions of a minimum diameter of 2cm at the largest extent and only injuries/wounds occurring at head, neck, or back level, as only them could be detected during the data collections, due to the distance of the observer from the animals) (This indicator was counted as "number of animals bruised or injured at least once")	**Bru/Wou**
Falling	The animal loses the upright position suddenly and a part of its body other than the limbs touches the ground (e.g. udder, sternum, whole flank, or abdomen). It succeeds in standing up again	**Fall**

8 These definitions were adapted from Welfare Quality® Project, 2009 and Grandin, 2010 and on the basis of the preliminary observations. Some information was also taken from Algers et al. (2009) and Dalla Villa et al. (2009).

9 These codes were inserted in the data collection forms (Appendix 1. Data collections forms).

Indicator	Description	Code[9]
Falling without recovering	The animal falls (see "Falling" in this table) and is not able to return to the standing position (e.g. it falls on its back in a narrow corridor)	**Fall no rec**
Getting (severely) lame (fractures)	The animal becomes lame during the slaughter process. Only severe lameness is counted as mild lameness was difficult to detect (due to the distance of the observer from the animals during the observations), defined as: strong reluctance to bear weight on one limb, or more than one limb affected	**Lame**
Going back into the corridors from the stunning box	The animal moves back from the stunning box to the passageways	**Back to C**
Hesitating	The animal stops and falter, lasting less than 3 sec	**Hes**
Jumping	The animal jumps fully clear of the ground (e.g. when moving from the truck ramp onto the unloading bay)	**Jump**
Mounting	The animal lifts itself up on its hind legs and jumps with its forelegs onto another animal. It finishes when the animal steps back on the ground	**Mount**
Moving backwards	The animal stops, moves at least 1m backwards (without turning around) by itself or as a reaction to the handler trying to drive it forward, *and stops*. This may happen both when the animal moves backwards "on its own" or as a consequence of the animal(s) in front moving backwards	**Mov back**

Indicator	Description	Code[9]
	(Important: in the group observations, when detected, this indicator was always multiplied by 2 – see the explanation in section 2.7.2.4)	
Bad stunned	No-loss/incomplete loss of consciousness after stunning, as shown by the following selected indicators[10]: 1. arched back righting reflex with the head bent straight back (sideways neck flexion that relaxed in a few seconds is not a righting reflex); 2. rhythmic breathing (ribs moved in and out at least twice); 3. vocalisations while hanging on the bleed rail; 4. stiff, curled tongue; 5. corneal reflex (detected by touching the eyes); 6. spontaneous blinking (like alive animals in the yards); 7. second shot (the animal is shot twice as the operator detects the inappropriate stun) (indirect indicator)	**Bad stun**
Pain, frustration, etc due to the use of energised prod	All stress reactions due to at least one touch with the energised prod. This indicator in the on-site data collections was counted as “animal touched (any part of its body) at least once with an energised prod or any sharp/pointed handling tool”	**Prod**
Pushing against	The animal pushes backward against the gate of the stunning box and stops	**Push**

[10] For practical reasons during the on-site observations *undefined* stuns were not assessed (e.g. looking at eyeball rotation up to sticking with no other symptoms; nystagmus; gasping/groaning; and excessive kicking or struggling at sticking in combination with eyeball rotation, nystagmus, and gasping/groaning) and only good or bad stuns were counted.

Indicator	Description	Code[9]
the stunning box door		
Running	The animal runs *as a reaction to the handler driving it forward*[11]	**Run**
Slipping	If occurring *on the truck ramp, in the unloading bay or in the corridors*: when a portion of the leg other than the foot touches the ground If occurring *in the stunning box*: when a portion of the leg other than the foot touches the floor of the stunning box, OR a foot loses contact with the floor of the stunning box in a non-walking manner OR the animal experiences a loss of balance that alters its steadiness in the stunning box	**Slip**
Turning around	If occurring on *the truck ramp, in the unloading bay/corridors/ in front of the entrance of the stunning box*: the animal turns around, moves back at least 1m against the flow and turns around again into the flow If occurring *in the stunning box*: the animal turns back (when stun box is big enough to allow it or the animal is small)	**Turn around**
Vocalising alone[12]	*In the stunning box only*, the animal vocalises at least once without any other indicator, due to fear, pain, etc	**Voc alone**

11 In Welfare Quality® (2009) "running" is defined as the animal running also "by itself"; this aspect was not considered to differentiate from "too fast animal flow", which in this research project was listed as a hazard.

12 The indicator "Vocalisations" was not used in other areas of the slaughter plant as difficult to detect in both the group and the individual observations, often due to the noise.

2.7.2.2 Measurement of adverse effect severity

To assess the *severity* of the adverse effects, each indicator of poor welfare was *pre-assigned* to a specific category of severity, using the 1-4 level scale illustrated in Table 7 (negligible effects - level 0 - were not considered).

Thus during the on-site data collections, when a specific indicator of poor welfare was detected (and ticked in the appropriate cell in the data collection form), the related adverse effect and its severity were also counted at the same time.

For instance, if one animal was observed “balking in the unloading bay” (i.e. indicator), this meant “one animal showing moderate distress, pain, frustration, fear or anxiety and clear change in adrenal or behavioural reactions” (i.e. adverse effect of severity 2 level) (Table 7).

The levels of severity associated to the various indicators of poor welfare were established upon discussion and agreement within the research group; the results of the preliminary observations; and check by an external expert (Table 7).

When a cascade of indicators of poor welfare was observed in an animal exposed to a hazard (e.g. vocalising and balking due to hitting by the operators), the indicator reflecting the most intense severity was recorded, since it was considered as corresponding to the respective intensity of the hazard. Of course this was a simplification; however it was considered acceptable for the scope of this study, as in existing animal welfare risk assessments cascades of indicators are not considered (section 2.7.1)[13].

The preliminary observations proved that due to the nature of the slaughter process, the animals tend to show more severe adverse welfare effects towards the entrance of the stunning box.

For instance, an animal “moving backwards” (indicator) in the area in front of the stunning box would clearly show a level of stress, fear, frustration *more severe* than an animal “moving backwards” in the unloading bay.

However, such increase in stress reactions (etc), although obvious, was not

[13] In addition it is reasonable to think that the experts that answered the questionnaire followed the same approach, i.e. scored the probability of occurrence of the various adverse effects considering the events that indicate the worst possible adverse effects.

easy to quantify using the indicators described in Table 6 or other measurable indicators manageable during the slaughter process[14]. In other words, “moving backwards” was easy to measure (i.e. “*The animal stops, moves at least 1m backwards (without turning around) by itself or as a reaction to the handler trying to drive it forward, and stops*”), but it was not possible to measure “moving backwards *plus increased fear*”.

The problem was solved by associating some indicators to different severity levels depending on the phase of the slaughter process where they occurred.

Thus “moving backwards” in the unloading bay was given a severity level equal to 1 (limited), whereas the same indicator occurring in front of the entrance of the stunning box was associated to a severity equal to 2 (moderate). Another example was “turning around”, which was assigned to a severity level equal to 1 (limited), 2 (moderate) and 3 (acute), depending if it occurred respectively in the unloading bay, in front of the stunning box or inside the stunning box (Table 7).

Moreover, the preliminary on-site observations highlighted that some hazards clearly induce in the animals adverse effects for which it is difficult to define easy-to-detect specific indicators.

For instance, the use of the energised prod (i.e. hazard) induces a series of stress reactions which may not be reflected by a specific indicator. In existing reports on animal welfare risk assessments the energised prod was associated to the indicators “bruising and wounds” (Dalla Villa et al., 2009). However, in the on-site observations bruises or wounds due to the use of such tool were never detected. Thus it was decided to associate to this hazard the indicator “pain, frustration, etc *due to the use of energised prod*”, with a severity level equal to 3, corresponding to “*explicit* pain, frustration, fear or anxiety” (Table 7). This was not in contrast with the available reports where “bruising and wounds” were also given a severity lever equal to 3.

[14] Fear may be assessed using e.g. the hearth rate or specific hormones in blood; however this was not considered feasible during the slaughter process.

During the on-site data collections at *individual level*[15], it was noticed that an indicator of adverse effect (e.g. “mounting”) occurring more than once in the same animal (and in the same phase of the slaughter process under observation) may reflect either (A) a severity higher than the one of the indicator occurring just once; or (B) no increase in severity. Case (A) indicators were named “cumulative-severity indicators”, case (B) “non-cumulative-severity indicators”.

The choice on whether to increase or not the level of severity was made on a case by case basis and in general considering the *time occurring between the various events of the same indicator*.

For instance, an animal exposed to the hazard “Too rapid/uncontrolled funnelling into single file when entering the corridors from the unloading bay” mounted (indicator) 3 times almost *consecutively*. In this case, the indicator was considered as a cumulative-severity type and the overall related severity of this adverse effect was estimated as equal to 3 (which is 1 point higher that the severity level of “mounting” occurring just once). Conversely, the same hazard provoked in another animal under observation “moving backwards” (indicator) several times rather *separate in time*. In this case the severity level was not increased.

Indicators of cumulative-severity could not be detected during the group observations.

In addition, often severity 3 indicators occurring more than once consequently could not be increased in severity as this would not match the definition of severity 4, which implied the “incompatibility with life” (i.e. Severity 4: *Extreme distress, pain, fear or anxiety, that if persisting would be life-threatening*).

For instance, the severity 3 indicator “Pushing against the stunning box door” when occurred more than once consequently in the same animal was not increased to severity 4 as it was considered that pushing against the stunning box over and over again, although indicating a high severity

[15] Individual observations were performed for the stunning/killing phase; and in a second stage for the three phases of pre-slaughter handling (i.e. unloading, corridors and entrance of the stunning box).

distress, would lead the animal to die.

All strengths and limitations of the method applied were highlighted in the conclusions.

Table 7. Adverse welfare effects at slaughter, related behavioural indicators and associated category of severity[16].

Adverse welfare effect	Indicators	Severity	Score
No distress, pain, frustration, fear or anxiety	Normal behaviour	**Negligible**	**0**
Minor distress, pain, frustration, fear or anxiety	Minor behavioural changes such as: 1. Hesitating 2. Moving backwards in unloading bay/corridors 3. Running 4. Turning around on the truck ramp, in the unloading bay or in the corridors	**Limited**	**1**
Some distress, pain, frustration, fear or anxiety. Clear change in adrenal or behavioural reactions	1. Balking in the unloading bay/corridors 2. Jumping 3. Mounting 4. Moving backwards (occurring in the area close to the entrance of the stunning box) 5. Pain, frustration, etc due to e.g. the use of sharp handling	**Moderat**	**2**

[16] The levels of severity associated with the various poor welfare indicators were established were established upon discussion and agreement within the research group; the results of the preliminary observations; and check by an external expert.

Adverse welfare effect	Indicators	Severity	Score
	tools 6. Slipping on the truck ramp, in the unloading bay/corridors/ the area close to the entrance of the stunning box 7. Turning around in the area in front of the entrance of the stunning box		
Explicit distress, pain, frustration, fear or anxiety. Strong change in adrenal or behavioural reactions	1. Attempt to escape 2. Balking the area close to the entrance of the stunning box 3. Bruises/wounds 4. Falling 5. Going back into the corridors from the stunning box 6. Pain, frustration, etc due to e.g. the use of energised prod 7. Pushing against the stunning box door 8. Slipping in the stunning box 9. Turning around in the stunning box (all can be together with vocalising) 10. Vocalisation alone in the stunning box	**Severe**	**3**
Extreme distress, pain, fear or anxiety, that if persisting would be life-threatening	Extreme changes from normality that if persisting would be life-threatening, e.g.: 1. Bad stun (no-loss/incomplete loss of consciousness after stunning) 2. Falling without recovering 3. Getting (severely) lame (fractures)	**Critical**	**4**

2.7.2.3 Duration of the adverse effects

In existing animal welfare risk assessments *duration* of the adverse effects is estimated by the experts as the period of time during which an animal is believed or expected to be experiencing the adverse effect once exposed to the hazard.

In the on-site data collections, the observations focussed on the indicators of adverse effects (e.g. slipping) and not the adverse effects (e.g. distress). Thus it was noticed that measuring the duration of the indicator (e.g. the duration of "slipping") would not imply that the related adverse effect (e.g. distress) had the same duration of the indicator showing it[17]. Moreover, a manageable and reliable method for measuring the duration of the adverse effects at slaughter was not identified in the literature.

Thus the duration of the adverse effects was not measured in the on-site data collections and for the final risk assessment the duration estimated by the experts in the questionnaire was used (section 3.3). This decision did not negatively affect the comparison of the results of the two risk assessments and the overall scope of the study (section 5.3).

2.7.2.4 Measurement of adverse effect likelihood

As it was not possible to assess the duration of the adverse effects, during the on-site observations the *likelihood (occurrence) of the adverse effects by severity levels* was estimated.

In the *individual* observations[18], this parameter was estimated by counting the *number of indicator(s)* of adverse effect(s) observed during the on-site observations, which corresponded to the *number of animals* suffering the adverse welfare effect(s), out of all animals exposed to the hazard(s).For example, 3 "Falling", 4 "Wounds" and 6 "Slipping" would correspond to

---

[17] A few examples of indicators whose duration could be measurable are e.g. "wounds", which (at slaughter) last from the moment when they occur until the end of the slaughter process.

[18] Individual observations were performed for the stunning/killing phase; and in a second stage for the three phases of pre-slaughter handling (i.e. unloading, corridors and entrance of the stunning box).

respectively "3 animals falling at least once", "4 animals wounded at least once" and "6 animals slipping at least once". The indicators occurring more than once in the same animal as a consequence of the exposure to the *same hazard* were always counted as "1 indicator" and, depending on the time occurring between the indicators, either of cumulative or non-cumulative severity (section 2.7.2.2).

On contrary, in the *group* observations of the phases "corridors" and "entrance of the stunning box", the observer counted the total number of indicators of adverse effects and not the number of animals showing them. For instance, a slipping score equal to 4 could be equal to: the same animal slipping 4 times; 2 animals slipping twice each; etc, and not necessarily to 4 animals slipping once each. Thus the results of the group observations were analysed and used for the final risk characterisation on a case by case basis as described in section 3.1.2.

In the *group observations*, one particular case was represented by the indicator "moving backwards". When the animals are in a row (e.g. in the corridors), if the animal in front moves backwards, the one behind is also induced to move backwards. In addition, cattle are herd animals and often fear behaviours such moving backwards, associated to a frightening situation, are likely to influence the group mates (WQ®, 2009).

Thus, when detected in the group observations, "moving backwards" was always multiplied by 2, assuming that at least two animals were always involved[19]. This assumption may be criticised for not reflecting the real situations and it may be argued that the indicator "moving backwards" could have been considered in the individual observations only. However, this was not considered as this indicator was one of the most frequent and excluding it from the group observations would have reduced the precision of the estimates far more than applying the multiplication by 2. Moreover, the results of the group and individual observations were always compared and the use of ones or the others thoroughly evaluated on a case by case basis (section 3.1.2).

[19] The results were limited to the maximum number of animals.

2.7.3 Adverse effect characterisation in the risk assessment based on expert opinion

Part 2 of the questionnaire contained a set of questions aiming to perform adverse effect characterisation. The various parameters of adverse effect characterisation (description of the adverse effects and definition of their severity, duration, likelihood) were asked as follows:

1. *Description* of the adverse effects. The same assumption as in the on-site adverse effect characterisation was given in the questionnaire, i.e. adverse welfare effects occurring at slaughter are mainly represented by *distress, pain, frustration, fear, anxiety, changes in adrenal or behavioural reactions, with severity levels being limited, moderate, severe and critical* (1-4 scale, with 0 equal to "negligible adverse effects"). This statement was explained and it was not asked to comment on it.
2. *Severity* and *likelihood* of the adverse effect. It was asked to indicate, for each hazard listed in the questionnaire (and for which hazard identification was agreed), the *most likely probability of occurrence* of *each of the 4 severity levels* adverse effects (Table 7 above, first column on the left).

   The question was posed in a way to avoid asking percentages, as follows:

   Considering a group of 100 beef calves[20] exposed to the hazards previously listed and agreed upon, please indicate how many animals would show:*

   1. *Adverse effects of negligible Severity (0)***
   2. *Adverse effects of limited Severity (level 1)*
   3. *Adverse effects of moderate Severity (level 2)*
   4. *Adverse effects of severe Severity (level 3)*
   5. *Adverse effects of critical Severity (level 4)*

   **the experts were asked to ensure the sum of the 5 likelihoods was equal to 100; **this was described as "no adverse effect"*

[20] The target population was described in the premise of the questionnaire.

By answering this question, the experts estimated the *likelihood of the adverse effects by severity level*.

The way to estimate the likelihood of the adverse effects was slightly different from existing animal welfare risk assessments, where the experts do not assess the probability of occurrence of *all* 4 severity levels for each hazard and indicate only some adverse effects and/or indicators of adverse effects (section 2.7.1). This was made to make the exercise clearer for the experts.

Another difference from existing risk assessments was that the experts in this study were asked to assess the most likely probability of the adverse effects only (and not also the minimum and maximum probabilities). The decision was made for simplifying the questionnaire as much as possible and it did not negatively affect the comparability of the two final risk assessments.

To help understanding the four severity levels, some examples of potential indicators of poor welfare were given. However, the classification of welfare indicators by severity levels used in the on-site data collections was not illustrated in the questionnaire. The few examples reported were “Hesitating” (example of a status of stress of limited severity); “Balking” (associated to moderate severity); “Fractures” (linked to severe pain); and “Bad stun” (example of critical severity).

Two alternative methods were considered and discarded:

- The experts could have been asked to indicate what actually happens to the animals when exposed to the hazards (i.e. to illustrate the “indicators” of adverse effects). For instance the question could have been: “Considering a group of 100 beef calves exposed to the hazard “Dirty floor”, please tell how many slip, fall, move backwards, mount, etc”. The indicators illustrated in Table 6 could have been listed and the experts asked to add as many indicators as they believed necessary). In this way however, the severity level meant by the experts by replying e.g. “10 animals fall”, “30 animals slip, etc”, would have remained unexplained.
- The experts could have been asked to indicate, for each of the 4

severity levels, both the probability of occurrence of the adverse effect(s) *and* the indicator(s) showing it. This would have certainly been the most comprehensive way to address the problem and would have also been useful to assess and (maybe) validate the classification of welfare indicators produced for the on-site observations. However, the questionnaire would have become extremely complex and difficult to answer. In addition, the scope of the project was to appraise existing animal welfare risk assessments which do not rely, for performing adverse effect characterisation, on agreed and validated welfare assessments methods (section 2.7.1).

Thus the method selected was assumed as the most practical and effective considering the available tool (i.e. a questionnaire) for eliciting expert opinion. Nonetheless the choice was thoroughly discussed in the comparison of the results of the two risk assessments (section 5.2) and in drawing the conclusions of the overall study.

3. *Duration* of the adverse effects. As in existing animal welfare risk assessments[21], the experts were requested to estimate the duration of the 4 adverse effects (by severity levels) consequent to the exposure to the hazards. A 1-3 scale was proposed, reflecting the time that the animal was believed or expected to be experiencing the adverse effect, once exposed to the hazard (Table 8).

Table 8. Scale for calculating the duration of the adverse effects in the questionnaire

Score	Criteria for duration of the adverse effect
1	The adverse effect lasts less than one minute and ends
2	The adverse effect lasts for some minutes and ends
3	The adverse effect is not reversible and lasts until the end of the

[21] And differently from the empirical risk assessment, where adverse effect duration was not estimated.

Score	Criteria for duration of the adverse effect
	slaughter process
(0)	Duration “not applicable”, when the experts indicated that “no adverse effect” would occur (see text above in this chapter)

2.8 Uncertainty

Since risk assessment constitutes a fundamental part of the decision making process, it is essential to estimate the degree of uncertainty on the various risk assessment parameters and assumptions (CAC, 1999).

Two principal factors of uncertainty influence the outcome of risk assessment, i.e. the likelihood and magnitude of the adverse welfare effects assuming exposure to the particular hazard, and the proportion of the animal population subjected to the hazard in a geographical region under a certain specifically defined scenario (i.e. exposure assessment).

While the former may sometimes be assessed by scrutinising available scientific studies, few if any published data are available on the degree of exposure of animal populations to a particular hazard. In addition, considerable differences in exposure exist across Europe, dependent on the various animal production or processing systems and welfare experts involved in a particular area are not necessarily aware of the situation outside their immediate activity arena.

In existing animal welfare risk assessments uncertainty is generally expressed on a 1-3 scale, on the basis of the available evidence, where low, medium of high uncertainty refers respectively to: “solid and complete data available; strong evidence provided in multiple scientific publications; authors report similar conclusions”; “some but no complete data available; evidence provided in small number of scientific publications; authors’ conclusions vary from one to another. Solid and complete data available from other species which can be extrapolated to the species considered”; and “scarce or no data available; rather evidence provided in unpublished reports, based on observations or personal communications; authors’ conclusions vary considerably between them" (Algers et al., 2009).

2.8.1 Uncertainty in the risk assessment based on data collected in abattoir

Uncertainty on the on-site exposure assessment and likelihood of the adverse effects was reflected by the confidence intervals of each estimate, which indicated the uncertainty related to the true values estimated (section 3.1).

2.8.2 Uncertainty in the risk assessment based on expert opinion

In the questionnaire the experts were asked to indicate their level of uncertainty on exposure assessment and adverse effect characterisation.

In particular on the latter the experts had to indicate two types of uncertainty: one related to the likelihood of the adverse effects; and one related to the duration of the adverse effects. This was different from the existing risk assessments, where only the uncertainty related to overall adverse effect characterisation is measured. This decision was made to enhance precision in uncertainty assessment. In fact, it was foreseen that an expert may be sure about e.g. the likelihood of occurrence of an adverse effect as resulting from the exposure to a hazard and be rather unsure about its duration. It was considered that this would not hamper the comparison of the results of the two risk assessments of this research project (as uncertainty in the risk assessment based on empirical data was indicated by the confidence intervals of the estimates).

Each uncertainty was estimated by the experts using a 1-3 scale, where the various levels of uncertainty were defined on the basis of the availability and quality of scientific evidence *and the level of personal experience on the topi*c (Table 9). The new concept of personal experience (never considered in other risk assessments for animal welfare) was included as considered a fundamental requirement of uncertainty assessment in risk assessment for animal welfare. In fact, in this field in some cases the assessments do not necessarily necessitate to be supported by published

evidence.
For instance, it is unarguable that if an animal is skinned when still alive it will suffer from extreme pain. However, according to the definition of uncertainty used in existing risk assessments for animal welfare, if nothing has been published on pain due to skinning, the uncertainty related to this subject should be high.
The concept of personal experience in the topic was thus added to avoid such unrealistic situations. In addition, it provided more flexibility to the experts that they could to reply although maybe not aware of specific publications on all topics asked in the questionnaire.
The method for assessing uncertainty was considered in the analysis and discussion of the results of the two risk assessments and for drawing the conclusions of the study.

Table 9. Qualitative uncertainty scores in the questionnaire submitted to the experts

Uncertainty Score	Description	Explanation
1. Low	Robust scientific evidence available OR high personal experience on the topic	Robust scientific evidence: solid and complete data available; strong evidence provided in multiple references; authors report similar conclusions OR the experts have observed the situation in many occasions
2. Medium	Some but non-robust scientific evidence available AND/OR some personal experience on the topic	Non-robust scientific evidence: no complete data available; authors' conclusions vary from one to another; solid and complete data available from other species which can be extrapolated to the species considered AND/OR the experts have observed the situation in some cases

Uncertainty Score	Description	Explanation
3. High	Scarce or no scientific evidence available AND/OR no personal experience in the topic	Only evidence provided in unpublished reports, based on observations or personal communications; authors' conclusions vary considerably between them AND/OR the experts have never observed the situation

3. PART III. Data analysis

3.1 Analysis of the on-site estimates

For the analysis of the on-site collected data, the confidence intervals (CI) of the point estimates were calculated[22]. The confidence intervals indicated the uncertainty on the true values estimated.

For each hazard one probability of occurrence (P) and related confidence interval were calculated.

For instance, in one observation the “Energised prod at unloading” (i.e. hazard) was used on 29 out of 491 animals observed. The related P and CI are reported here below:

N	K	P (%)	Confidence interval (95%)	
			Lower (%)	Upper (%)
491	29	6	4	8

N: number of observed animals
K: number of animals exposed to the hazard
P: best estimate for hazard probability
Lower: lower limit
Upper: upper limit

Since the likelihood of the adverse effects was assessed by (4) severity levels, for each hazard 4 adverse effect probabilities (P1, P2, P3 and P4) (and 4 confidence intervals: CI1, CI2, CI3 and CI4) were estimated.

For instance in an observation out of 24 animals exposed to the hazard “Too many animals in the flow”, 6 and 3 animals showed respectively indicators of severity 2 and 3 adverse effects[23]. The probabilities of occurrence of the adverse effects by severity levels and the related confidence intervals are illustrated here below:

---

22 Using the method proposed by Johnson et al., 2002.

23 See section 2.7.2.1 and Table 6 for the explanation of the indicators.

K	L	Observed indicator*	Severity level	Probability (%)		Confidence (95%)		
							Lower (%)	Upper (%)
24	**0**	/	S1	**P1**	0	**CI1**	0	12
24	**6**	3 Mounting more than once (no cumulative-severity) 2 Moving backwards more than once (cumulative-severity) 1 Mounting alone	S2	**P2**	25	**CI2**	10	47
24	**3**	2 Mounting more than once (cumulative-severity) 1 Falling	S3	**P3**	13	**CI3**	3%	32
24	**0**	/	S4	**P4**	0	**CI4**	0	12

K: Number of animals exposed to the hazard
L: number of animals showing the adverse effect(s)
S1, S2, S3 and S4: severity 1, 2, 3 and 4
* See section 2.7.2.1 and Table 6 for the explanation of the indicators

Lower: lower limit; Upper: upper limit
P1, P2, P3 and P4: probability of occurrence of severity 1, 2, 3 and 4 adverse effects
CI1, CI2, CI3 and CI4: confidence intervals of P1, P2, P3 and P4

3.1.1 Interacting hazards observed in abattoir

To reproduce the structure of the existing risk assessments for animal welfare under appraisal (which do not consider interactions between hazards[24]), when detected in the on-site observations, interacting hazards were analysed as follows:

1. When performing exposure assessment: interacting hazards were counted as if happening separately (i.e. no interaction was counted). For instance, the hazards “Dirty floor” and “Energised prod” occur at the same time and produce 1 falling animal. The two hazards are counted separately as 1 “Dirty floor” and 1 “Energised prod”.
2. When assessing the likelihood of the adverse effects: the adverse effects caused by interacting hazards were allocated to one hazard only, by running a random selection. Using the above example, a random selection is run and the adverse effect “Falling” (severity 3) allocated to the randomly selected hazard (either “Dirty floor” or “Energised prod”).

Thus the on-site observations highlighted how the exclusion of the interacting hazards from the analysis is likely to cause an underestimation (or overestimation) of the related risk (see the conclusions in section 6.2).

In the on-site observations, hazards that were almost always present (e.g. continuous noise in the corridors) were not interpreted as interacting with other hazards.

3.1.2 Analysis of on-site estimates of group and individual observations

The on-site observations of the pre-slaughter handling phases (i.e. unloading, moving along the corridors and entrance in the stunning box) were carried out in two steps, i.e. firstly at group and secondly at individual level.

During stunning and killing, due to the nature of the process, the animals were observed individually only.

---

[24] Interacting hazards are hazards that occur at the same time and have a combined action, producing adverse effects that are different from those produced by the hazards occurring separately.

Sampling the animals in group had the advantage to allow the observation of a higher number of animals at the same time and the limitation that some items (e.g. hazards and/or indicators of adverse effects) could not be detected with the same precision as in the individual observations. However, the results of the group observations were corrected using the data collected at individual level and discarded when considered not reliable.

The results of the group and individual observations for exposure assessment and likelihood of the adverse effects were *merged* when *both* the following requirements were met:

1. Requirement 1. The confidence intervals (CI)[25] overlapped and thus the data showed no significant difference in the estimated likelihood between group and individual observations.
2. Requirement 2: The data was collected in the same way, i.e.:

   ➔For exposure assessment:

   - *The unit of observation was the same in the group and individual observations* (i.e. "Individuals/Individuals"). When the unit of observation at group level was e.g. "triples", "pairs" or "group", it was not possible to merge the results, as in the individual observations the unit of observation was always "individual" (Table 5 in section 2.6.2).
   - *The number of hazards detected corresponded to the number of animals exposed to them, both in the group and individual observations*. For instance, "3 energised prods" corresponded to "3 animals exposed to energised prod at least once" both at group and individual level; in this case, if the CI overlapped, the results of exposure assessment of the group and individual observations were merged. On the other hand, "3 noises" in the group observations corresponded to the actual "number of noises" detected, whereas at individual level it was equal to the "number of animals exposed to a noise at least once"; in this

25 The CI was interpreted as the "uncertainty" on the true values estimated.

case the results of exposure assessment could not be merged (section 2.6.2)[26].

➔For likelihood of the adverse effect (by severity levels), *the number of indicators detected corresponded to the number of animals showing those indicators, both in the group and individual observations*. For instance, at unloading 7 "slipping" or 4 "mounting" corresponded to "7 animals slipping" or "4 animals mounting" both at group and individual level.

However, this would not always happen in the group observations of the phases "corridors" and "entrance of the stunning box", as in such phases the observer counted the total number of indicators of adverse effects and not the number of animals showing them. For instance, a slipping score equal to 4 could be equal to: the same animal slipping 4 times; 2 animals slipping twice each; etc, and not necessarily to 4 animals slipping once each[27]. In the unloading phase this did not happen, as in the slaughterhouse observed the layout of the unloading bay allowed the detection of the indicators at individual level.

When it was not possible to merge the results of the group and individual observations, the analysis of the results was performed as follows:

1. When the *CI overlapped*, but *data at group and individual level were collected in a different way*, the results of the observation with a smaller CI were used.
2. When the CI did *not overlap*, but the *data at group and individual level were in collected in the same way* (and it always occurred that the value estimated at group level was bigger than the estimate at individual level): the individual observations were used, as normally more precise (see above in this section).

26 It must be noted that the results of exposure assessment for the hazard "noise" could not be merged also because the hazard presents different units of observations in the group and individual observations (i.e. "triples" against "individuals").

27 When no adverse effect was observed in one or both observations, the results of the group and individual observations were merged also for the phases "corridors" and "entrance".

3. When *none of the two requirements was met* (i.e. the CI did not overlap and the data at group and individual level were collected in a different way) the results of the individual observations were used, as in general more precise.

Some examples of the analysis of the results of the group and individual observations are illustrated in sections 3.1.2.1 and 3.1.2.2.

The same analysis was performed for all on-site estimated values and for practical reasons is not documented in the project[28]. However, the results of the analysis of the group and individual observations for all hazards are reported in Appendix 3.

[28] The analysis is available upon request.

3.1.2.1 Examples on exposure assessment

Example 1: use the results of group and individual observations merged

H2. Energised prod at unloading

	Group observations					Individual observations				
	Unit of observation: Individual					Unit of observation: Individual				
	N	K	P (%)	CI (95%) Lower (%)	CI (95%) Upper (%)	N	K	P (%)	CI (95%) Lower (%)	CI (95%) Upper (%)
Exposure assessment	491	29	6	4	8	30	6	20	8	39

K: number of animals exposed to the hazard
P: best estimate for hazard probability
CI: confidence interval
Lower: lower limit
Upper: upper limit

a. the 2 CI overlapped;
b. the unit of observation was the same in the group and individual observations and the number of hazards detected corresponded to the number of animals exposed to those hazards in both observations.

→ Merge the results of the group and individual observations:

	Group and Individual observations *merged*				
	N	K	P	CI (95%)	
				Lower (%)	Upper (%)
Exposure assessment	**521**	**35**	**7**	**5**	**9**

K: number of animals exposed to the hazard
P: best estimate for hazard probability
CI: confidence interval
Lower: lower limit
Upper: upper limit

Example 2: use the results of the observation with smaller CI

H76. Sudden noise at the entrance of the stunning box

	Group observations					Individual observations				
	Unit of observation: Triples					Unit of observation: Individual				
	N	**K**	**P (%)**	**CI (95%)**		**N**	**K**	**P (%)**	**CI (95%)**	
				Lower (%)	**Upper (%)**				**Lower (%)**	**Upper (%)**
Exposure assessment	131	9	7	3	13	30	2	7	1	22

K: number of animals exposed to the hazard
P: best estimate for hazard probability
CI: confidence interval
Lower: lower limit
Upper: upper limit

a. The 2 CI overlapped;
b. Different unit of observation (triples against individual) and hazards not counted in the same way in the group and individual observations.

→ Use the results of group observations as with smaller CI

Example 3: use results of individual observations only, as more precise

H79. Dirty floor at the entrance of the stunning box

	Group observations					**Individual observations**				
	Unit of observation: Individual					Unit of observation: Individual				
	N	**K**	**P (%)**	**CI (95%)**		**N**	**K**	**P (%)**	**CI (95%)**	
				Lower (%)	**Upper (%)**				**Lower (%)**	**Upper (%)**
Exposure assessment	394	394	100	99	100	30	26	87	69	96

K: number of animals exposed to the hazard
P: best estimate for hazard probability
CI: confidence interval
Lower: lower limit
Upper: upper limit

a. The 2 CI did not overlap;
b. Same unit of observation at group and individual level (and estimate at group level > estimate at individual level).
→ Use the results of individual observations as more precise

3.1.2.2 Examples on likelihood of the adverse effects

Example 1: use the results of group and individual observations merged

H79 Dirty floor at the entrance of the stunning box

Severity level	Group observations K	Group observations L	Group observations P (%)	Group CI (95%) Lower (%)	Group CI (95%) Upper (%)	Individual observations K	Individual observations L	Individual observations P	Individual CI (95%) Lower (%)	Individual CI (95%) Upper (%)
S1	394	0	0	0	0.8	26	0	0	0	11
S2	394	0	0	0	0.8	26	0	0	0	11
S3	394	1*	0.25	0	1.4	26	0	0	0	11
S4	394	0	0	0	0.8	26	0	0	0	11

K: Number of animals exposed to the hazard
L: number of animals showing the adverse effect(s)
P: best estimate for adverse effect probability
CI: confidence interval
Lower: lower limit
Upper: upper limit
S1, S2, S3 and S4: severity 1, 2, 3 and 4
*1 animal fell

a. The 4 CI overlapped;
b. In principle in the corridors and at the entrance of the stunning box the indicators of adverse effects were counted differently in the group and individual observations (and thus the results could be merged). However, in this case no adverse effect was observed in the individual observations.

→ Merge the results of the group and individual observations:

	Group and individual observations merged				
				CI (95%)	
Severity level	**K**	**L**	**P (%)**	**Lower (%)**	**Upper (%)**
S1	420	0	0	0	0.7
S2	420	0	0	0	0.7
S3	420	1	0.24	0	1.3
S4	420	0	0	0	0.7

K: Number of animals exposed to the hazard
L: number of animals showing the adverse effect(s)
P: best estimate for adverse effect probability
CI: confidence interval

Lower: lower limit
Upper: upper limit

S1, S2, S3 and S4: severity 1, 2, 3 and 4

Example 2: use the results of the individual observations only, as more precise

H94. Overall and continuous noise at the entrance of the stunning box

	Group observations					Individual observations				
				CI (95%)					CI (95%)	
Severity level	**K**	**L**	**P (%)**	**Lower (%)**	**Upper (%)**	**K**	**L**	**P**	**Lower (%)**	**Upper (%)**
S1	394	3*	1	0	2	30	0	0	0	10
S2	394	148**	38	33	43	30	13[1]	43	26	63
S3	394	19***	5	3	8	30	17[2]	57	37	75
S4	394	0	0	0	0.8	30	0	0	0	0
	*3 Hesitating ** 148 Mounting/Moving backwards in front of the entrance of the stunning box ***19 Balking (Here the number of indicators - not the number of animals - was counted)					[1] 8 animals moving backwards (in front of the stunning box) and 5 animals mounting [2] 13 animals moving backwards more than once (in front of the stunning box) (cumulative-severity); 1 animal mounting more than once (in front of the stunning box) (cumulative-severity); 3 animals balking (in front of the stunning box)				

K: Number of animals exposed to the hazard
L: number of animals showing the adverse effect(s)
P: best estimate for adverse effect probability
CI: confidence interval

Lower: lower limit
Upper: upper limit

S1, S2, S3 and S4: severity 1, 2, 3 and 4

a. Not all 4 CI overlapped;
b. In the group observations, the number of indicators detected did not correspond to the number of animals showing them.
   → Use the results of *individual* observations as more precise (e.g. adverse effects of cumulative-severity could be detected)

3.2 Analysis of the experts' responses

The response rate was 100% (i.e. all 11 experts answered the questionnaire).

During the phone calls several experts expressed doubts about the vague description of the hazards and the lack of data on exposure assessment for beef cattle at slaughter in northern Italy.

Perhaps these respondents had not fully realised that in this study they were the experimental subjects and that these aspects were unavoidable as the study had to follow the method of the animal welfare risk assessments under appraisal.

By virtue of being knowledgeable experts, their opinion was inherently valid, considering also that uncertainties about the scores were to be considered. As the respondents were familiar with the abattoirs and hazards, it is reasonable to assume that in most cases the scores were given for typical, representative examples of plants and hazards.

These respondents' doubts, however, showed a legitimate concern about a risk of misinterpretation of the outcomes of this study.

Another important caveat with respect to the interpretation of the experts' scores was that the scores were given on the basis of the expert personal interpretation of welfare. Even though the adverse welfare effects were defined in the survey as distress, pain, frustration, fear, anxiety, changes in adrenal or behavioural reactions (with 4 severity levels i.e. limited, moderate, severe and critical), differences in interpretation may have occurred and contributed to variation in the scores (see sections 5.2 and 6.3 on the discussion and conclusions).

3.2.1 Experts' weighted average

In the questionnaire the experts were asked to indicate their level of uncertainty on exposure assessment and likelihood of the adverse effects by severity levels, using a 1-3 scale (section 2.8.2).

To include the uncertainties of experts' judgements the uncertainty levels indicated by the experts were used as weights in the calculation of the

averages and standard deviation. An indicated low uncertainty was translated to a double weight, a high uncertainty as half weight compared to the medium uncertainty category. The sum of all weights was adjusted to the number of expert answers.
The experts' weighted averages indicated the assumed agreement between the experts on hazard probability (i.e. exposure assessment) and likelihood of the adverse effects by severity levels.
The confidence intervals were calculated assuming an asymptotic normal distribution. They indicated the uncertainty in selecting the experts.
Finally the results were restricted to the interval from 0 to 100%.

3.2.2 Duration of the adverse effects

The experts were asked to estimate the duration of the adverse effects (by severity levels) consequent to the exposure to the hazards (section 2.7.3) and indicate their related uncertainty[29].
To calculate the average duration category, the median of weighted answers was used in an analogue way as for the weighted probabilities of occurrence. Answers of experts indicating low uncertainty were counted double, with high uncertainty as half weight compared to experts indicating the medium uncertainty category.

3.2.3 Variability between the experts

The standard deviations of the probabilities of different hazards and adverse effects by severity levels are not directly comparable. Answers on probabilities around 50% may vary more than answers on more extreme probabilities.
To compare different standard deviations a coefficient of dispersion Di was calculated, which expresses the standard deviation in relation to the

[29] None of the experts scored different levels of uncertainty for likelihood and duration of the adverse effects.

deviation of fully random answers. So the Di value reflected the variability of the experts compared to random answers.

In particular, Di values indicated the following situations:

- Di < 1 (underdispersion): this value reflected a sort of coherent thinking/knowledge between the experts and homogeneity in the replies.
- Di > 1 (overdispersion): this value indicated that an element would interfere with the experts' replies, e.g. additional sources of variation, like different interests.

In view of the above, a Dispersion value close to 1 was interpreted as a "don't know" answer.

All above aspects were considered in the analysis of the experts' replies and in the comparison of the experts' weighted averages with the on-site estimates.

3.3 Risk characterisation

The final step of risk assessment is risk characterisation.

A risk is a function of the probability of an adverse effect and the duration and severity of that effect, consequent to a hazard to welfare.

Risk characterisation is the process of qualitative estimation of the probability of occurrence, severity and duration of poor animal welfare in a population.

The risk estimation was calculated using the following formula:

$$R = P(\text{Hazard}) \cdot \left[\sum_{\text{Effect}} P(\text{Adverse Effect} \mid \text{Hazard}) \cdot \frac{\text{Severity}(\text{Effect})}{4} \cdot \frac{\text{Duration}(\text{Effect})}{3} \right] \cdot 100\%$$

The overall risk estimator is the expected (average) product of the relative severity and the relative duration weighted by the likelihood of the hazard. It can be seen as expected relative harm (in percent of maximum 100%) to the animal related to a specific hazard.

As during the on-site observations it was not possible to estimate the duration of the adverse effects, for the final risk estimate the duration indicated by the experts was used[30].

An uncertainty analysis was performed to explore the overall uncertainty of this estimator. The influencing factors (e.g. the likelihood of the hazard and adverse effects) were substituted by distributions reflecting the confidence intervals of their measurement. A series of random replications showed the combined influence on the overall risk estimator. The results were again expressed as 95% prediction intervals around the most likely estimator.

[30] In some cases all experts indicated a 0% probability of an adverse effect and thus did not estimate the related duration. If the adverse effect was observed in the on-site observations, then the duration of the adverse effect was estimated on the basis of the knowledge of the observer.

4. PART IV. Comparison of the results of the two risk assessments

4.1 Introduction

The hazards identified in the literature as potentially occurring to beef cattle at slaughter and analysed for the two risk assessments are listed and described in Appendix 2. In table 10 the total hazards analysed and compared are reported by phase of the slaughter process.

Table 10. Number of hazards analysed and compared in the two risk assessment, per phase of the slaughter process

Phase	Hazards analysed	Hazards assessed as not relevant[1]	Hazards for which exposure assessment was compared[2]	Hazards for which risk estimate was compared[3]
U	19	5	14	8
C	14	6	8	5
E	13	4	9	5
S/K	10	1	9	6
Total	56	14	42	24

[1] When at least 5 out of 11 experts replied that the hazard does not exist *and* the hazard was never observed during the on-site observations
[2] Section 4.2.
[3] Sections 4.3 and 4.4
U: unloading
C: corridors
S/K: stunning and killing

When a hazard was appraised as not relevant by at least 5 out of 11experts *and* was never detected during the on-site observations, the hazard was defined as *not relevant* and excluded from the following analyses (i.e. exposure assessment, adverse effect characterisation and risk characterisation).
This occurred for 14 out of 56 hazards identified via the literature review (

Table 10 above).
Out of these 14 hazards, for 8 of them the (few) experts who replied that they were relevant assigned to these hazards a low probability of occurrence (max P=11%) with overall very homogeneous replies (max Di=0.23) (Table 11).
For 6 hazards (i.e. Inadequate temperature and Inadequate humidity in the various phases of the slaughter process), the variability between the replies of the experts who appraised these hazards as relevant was very high (Di close to 1), showing high difficulty in assessing the actual likelihood of those hazards (Table 12).
For all other hazards that were included in the following analyses, it often occurred that a few experts (always less than 5 out of 11 experts) replied that the hazards were not relevant and should not be included in the list of hazards to beef cattle at slaughter.
Overall, looking at all the reasons given for excluding the hazards, it was noted that in the vast majority of the cases the experts actually disagreed on the questionnaire and in a very few cases they seemed to misinterpret the questions.
For instance, one expert replied that “Energised prod at unloading” is not a hazard to beef cattle, as according to his/her experience the use of such tool at unloading has never been necessary. In another example two experts replied that “Overall and continuous noise originating in the stunning box” in general does not bother the animals and that in the chaotic situation the continuous noise is not a hazard. These replies showed an actual disagreement with the questionnaire.
In another case 1 expert replied that the hazard “Too many animals in the flow at the entrance of the stunning box” is not relevant, due to the layout of the corridors, which makes the animals progress one-by-one. The answer

showed a sort of misinterpretation as the fact that the animals proceed in a row does not necessarily implies that the space allowance is always adequate.
For one hazard only (i.e. “Too small stunning box”) it occurred that 5 experts replied that the hazard is not relevant, but the hazard was observed in the on-site observations.
None of the experts suggested new hazards to include in the list.

Table 11. Hazards estimated as non-relevant* and thus excluded from the analysis (and low Di between the replies of the experts who estimated that they are relevant)

	H16		H17		H28		H57	
	Narrow truck ramp		Sharp protrusions (truck ramp)		Sharp barriers at U		Sharp barriers in C	
	Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	0	6	0	5	0	4	0	6
Low	0	2	0	2	0	1	0	0
Up	10	10	9.5	9	1	8	1	13
N	30*	/	30*	/	521*	/	427*	/
K	0	/	0	/	0	/	0	/
Di	/	0.1	/	0.14	/	0.15	/	0.23
"No"	/	7	/	5	/	6	/	6
	H88		H105		H58		H89	
	Sharp barriers in E		Sharp barriers (stunning box)		Too wide corridors in C		Too wide corridors in E	
	Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	0	4	0	3	0	8	0	10
Low	0	1	0	1	0	1	0	9
Up	1	7	3	6	1	14	1	11
N	424*	/	115*	/	427*	/	424*	/
K	0	/	0	/	0	/	0	/
Di	/	0.14	/	0.11	/	0.18	/	0.02
"No"	/	6	/	6	/	6	/	6

* At least 5 out of 11 experts replied that a hazard does not exist and the hazard was never detected during the on-site observations
H: hazard
Est: on-site estimates (%)
Exp: experts' weighted values (%) P: probability of occurrence of the hazard, namely:

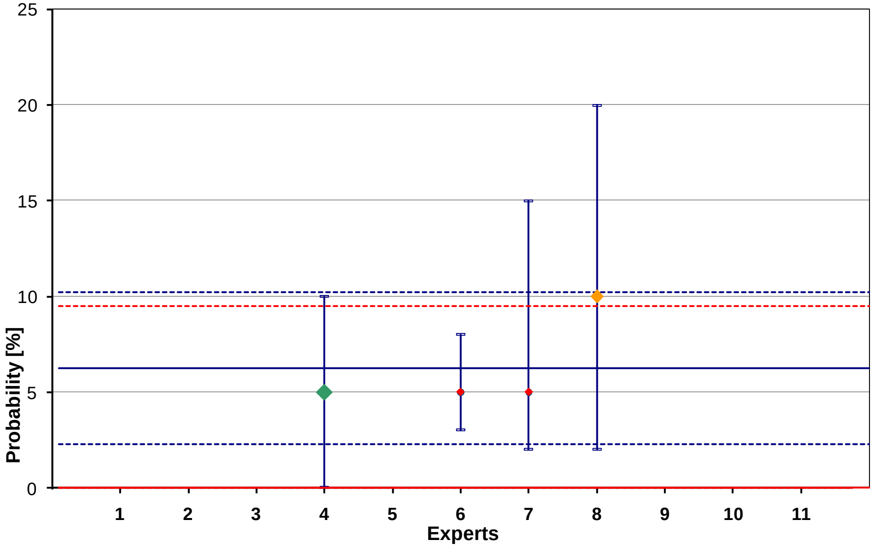
- In Est: on-site estimated average
- In Exp: weighted average

Low: lower limit; Up: upper limit

N: number of units of observations (e.g. *individuals, **groups, ***subgroups)
K: number of animals exposed to the hazard
Di: dispersion coefficient of the expert judgement
"No": number of experts that replied that the H does not exist

Hazard 16 (Narrow truck ramp) (Figure 3). The hazard was never detected during the on-site observations (P=0%; CI=0-10%) and 7 out of 11 experts considered this hazard as not relevant. The experts who agreed on including the hazard assigned to it a rather low probability of occurrence (P=6%, CI=2-10%), with homogeneous replies (Di=0.1).

Figure 3. On-site and expert opinion-based exposure assessment for the H16 "Narrow truck ramp"



Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

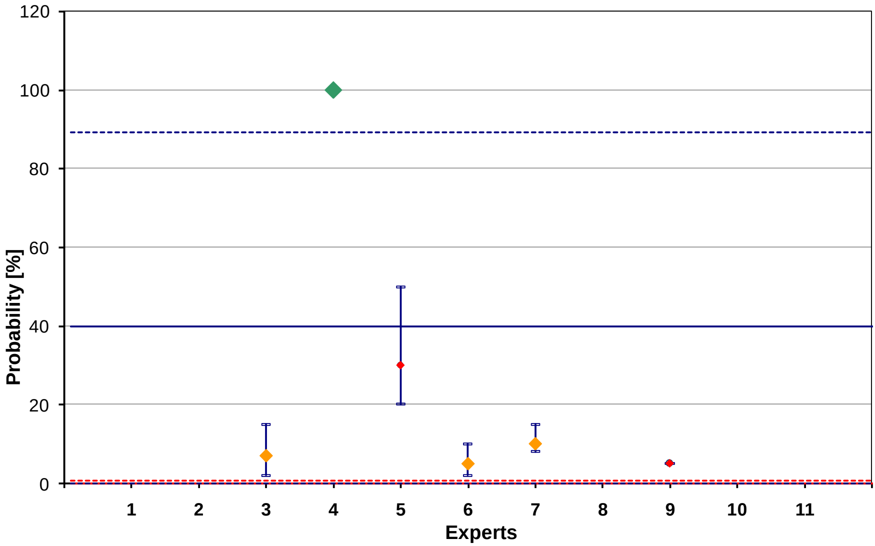
Table 12. Hazards estimated as non-relevant* and thus excluded from the analysis (and high Di between the replies of the experts who estimated that they are relevant)

	H25		H54		H85	
	Inadequate temperature at U		Inadequate temperature in C		Inadequate temperature in E	
	Est	Exp	Est	Exp	Est	Exp
P	0	40	0	40	0	43
Low	0	0	0	0	0	0
Up	1	89	1	87	1	100
N	521*	/	427*	/	424*	/
K	0	/	0	/	0	/
Di	/	0.96	/	0.90	/	1.08
"No"	/	5	/	5	/	7
	H26		**H55**		**H86**	
	Inadequate humidity at U		**Inadequate humidity in C**		**Inadequate humidity in E**	
	Est	**Exp**	**Est**	**Exp**	**Est**	**Exp**
P	0	41	0	45	0	45
Low	0	0	0	0	0	0
Up	1	96	1	97	1	97
N	521*	/	427*	/	424*	/
K	0	/	0	/	0	/
Di	/	0.99	/	1.00	/	1.00
"No"	/	5	/	5	/	5

Same legend as for Table 11

Hazard 25 (Inadequate temperature at unloading) (Figure 4). This hazard was never observed in the on-site data collections (P=0%, CI=0-0.6%) and 5 experts replied that the hazard does not exist at unloading. The variability between the replies of the experts who replied that this is a hazard was very high, close to 1 (Di=0.96).

Figure 4. On-site and expert opinion-based exposure assessment for the H25 "Inadequate temperature at unloading"



Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:

Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

4.2 Comparison of exposure assessment based on empirical data and on expert opinion

The results of on-site and expert-opinion based exposure assessments were compared for all hazards identified in the literature and whose relevance was confirmed by the on-site observations and the experts (section 4.1).
Analysing the experts' replies, a variability (i.e. Dispersion - Di) between the experts' responses ≥ 0.6 was considered high (see the following sub-chapters).
The results of the comparison of on-site and expert-opinion-based exposure assessments can be summarised as follows (Table 13):

1. *Not-conflicting/adhering* results. For 14 hazards the experts' replies were rather homogeneous (i.e. Di < 0.6) *and* the confidence intervals of the on-site estimated exposure overlapped with the confidence intervals of the weighted averages of the experts[31]. This showed how the random choice of the experts gave a chance that the agreement between the experts (i.e. the weighted average) was not in conflict with the on-site estimated values (section 4.2.1, Table 14 and Table 15).

2. *Inconsistent* results. For 24 the experts showed homogeneous replies (i.e. Di < 0.6) *and* the confidence intervals of the on-site and experts' estimates did not overlap. In these cases expert opinion was not confirmed by the on-site observations hazards (section 4.2.2, Table 16, Table 17 and Table 18).

3. For 4 hazards the variability between the experts' responses was high (i.e. Di ≥ 0.6). This was interpreted as if something interfered with the assessment (section 4.2.3, Table 19).

[31] Sometimes the confidence interval of the on-site observations was broad, due to the small sample size.

All above cases were discussed in the relevant section (5.1). The results of the comparison of exposure assessment in the two risk assessments performed in this study are illustrated in detail in the following sub-chapters.

Table 13. Results of the comparison of on-site and expert opinion-based exposure assessment

Results	Hazards at unloading	Hazards in the corridors	Hazards at the entrance of the stunning box	Hazards at stunning and killing	Tot H
Adhering results (CI overlapped and Di<0.6)	H2 (Prod at U); H11 (Sudden noise due to other than the truck); H12 (Too rapid/uncontrolled funnelling into single file at U); H14 (Slippery truck ramp); H15 (Steep truck ramp); H20 (Too many animals in the flow at U);	H51 (Too fast animal flow in C)	H80 (Too many animals in the flow in E)	H96 (Inadequate handling technique at S/K); H111 (Too small stunning box); H115 (Wrong shot); H116 (Delayed second shot); H106 (Stunning box gates' opening size too small); H112(Too big stunning box)	14
Inconsistent results (CI did not overlapped and Di<0.6)	H 3 (Inadequate handling technique at U); H10 (Sudden change of light intensity at U); H18 (Dirty floor at U); H22 (Too fast animal flow at U); H31 (Gates' opening size too small at U); H33	H37 (Prod in C); H38 (Inadequate handling technique in C); H44 (Sudden change of light intensity in C); H45 (Sudden noise in C); H47 (Dirty floor in C); H49 (Too many animals in the flow	H68 (Prod in E); H69 (Inadequate handling technique in E); H75 (Sudden change of light intensity in E); H79 (Dirty floor in E); H82 (Too fast animal flow in E); H90 (Too narrow corridors in E)	H108 (Faulty operating non-return gate in the stunning box); H114 (Delayed stunning); H120 (No check after stunning)	24

Results	Hazards at unloading	Hazards in the corridors	Hazards at the entrance of the stunning box	Hazards at stunning and killing	Tot H
	(Faulty operating non-return gates at U);	in C); H59 (Too narrow corridors in C); H62 (Gates' opening size too small in C); H64 (Faulty operating non-return gates in C)			
High variability between the experts (Di≥0.6)	H9 (Truck noise at U); H13 (Open/low side barriers on the truck ramp)	/	H76 (Sudden noise in E); H94 (Overall and continuous noise originating in the stunning box)	/	4
Total hazards for which the results of empirical and expert opinion-based exposure assessment were compared					42

Tot H: total hazards
CI: confidence interval
Di: dispersion coefficient of the expert judgement

4.2.1 Consistent results

Adhering results occurred when the confidence intervals of the on-site and experts' estimates overlapped and the variability between the experts' replies was low (Di<0.6)

Table 14. Consistent results for exposure assessment

	H2		H11		H12		H14	
	Prod at U		Noise due to other than the truck		Too rapid/ uncontr. funnelling		Slippery truck ramp	
	Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	7	11	8	19	8	19	17	9
Low	5	5	4	6	4	13	6	4
Up	9	16	13	31	15	25	35	14
N	521*	/	164***	/	109***	/	30*	/
K	521	/	13	/	9	/	5	/
Di	/	0.25	/	0.41	/	0.22	/	0.25
"No"	/	1	/	2	/	0	/	1
	H15		H20		H80		H51	
	Steep truck ramp		Too many animals in the flow at U		Too many animals in the flow in E		Too fast animal flow in C	
	Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	17	10	10	19	27	12	0	16
Low	6	5	2	14	12	5	0	8
Up	35	14	27	24	46	18	10	23
N	30*	/	30*	/	30*	/	30*	/
K	5	/	3	/	7	/	0	/
Di	/	0.17	/	0.19	/	0.23	/	0.27
"No"	/	4	/	1	/	3	/	2

H: hazard
Est: on-site estimates (%)
Exp: experts' weighted values (%)
P: probability of occurrence of the hazard, namely: in Est: on-site estimated average; in Exp: weighted average
Low: lower limit; Up: upper limit
N: number of units of observations (e.g. *individuals, **groups, ***subgroups)
K: number of animals exposed to the hazard
Di: dispersion coefficient of the expert judgement
"No": number of experts that replied that the H does not exist

Table 15. Consistent results for exposure assessment (cont')

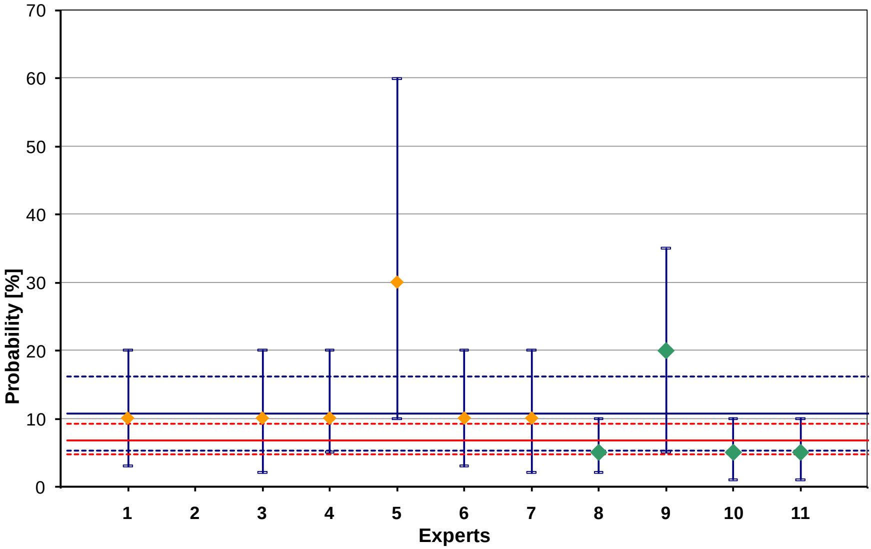
	H111		H112		H115	
	Too small stunning box		Too big stunning box		Wrong shot	
	Est	Exp	Est	Exp	Est	Exp
P	5	11	1	6	22	14
Low	2	0	0	3	15	9
Up	11	28	5	9	30	18
N	115*	/	115*	/	115*	/
K	6	/	1	/	25	/
Di	/	0.47	/	0.13	/	0.18
"No"	/	5	/	3	/	0
	H96		H106		H116	
	Inadequate handling technique at stunning		Stunning box gates' opening size too small		Delayed second shot	
	Est	Exp	Est	Exp	Est	Exp
P	10	13	0	13	10	10
Low	6	5	0	0	6	6
Up	18	22	3	28	18	15
N	115*	/	115*	/	115*	/
K	12	/	0	/	12	/
Di	/	0.35	/	0.48	/	0.22
"No"	/	1	/	4	/	0

Same legend as above table

Some examples of hazards for which the results of exposure assessment for the two risk assessments were adhering are illustrated here below.

Hazard 2 (Energised prod at unloading) (Figure 5). Both the on-site observations and the experts' replies confirmed that the energised prod is rarely used to drive the animals out of the truck and towards the exit of the unloading bay (on-site estimated probability of occurrence equal to 7%, CI=5-9%; expert-opinion-based P=11%, CI=5-16%, with low variability between the experts (Di=0.25)).

Figure 5. On-site and expert opinion-based exposure assessment for the H2 "Energised prod at unloading"

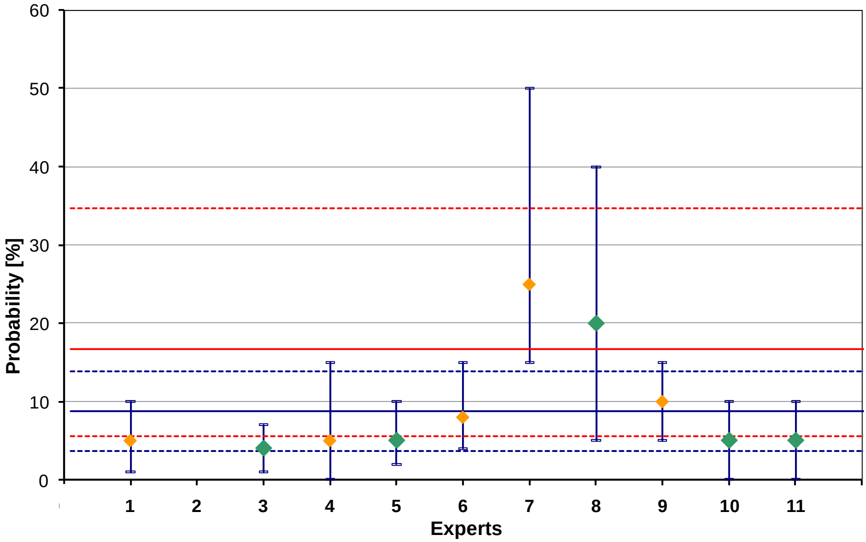


Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

Hazard 14 (Slippery truck ramp) (Figure 6). Adhering results (on-site estimated P=17%, CI-6-35%; experts' weighted average equal to 9%, CI=4-14%, with low variability between the experts). One expert replied that this hazard was not relevant, as normally the ramps are compliant with the design requirements.

Figure 6. On-site and expert opinion-based exposure assessment for the H14 "Slippery truck ramp"

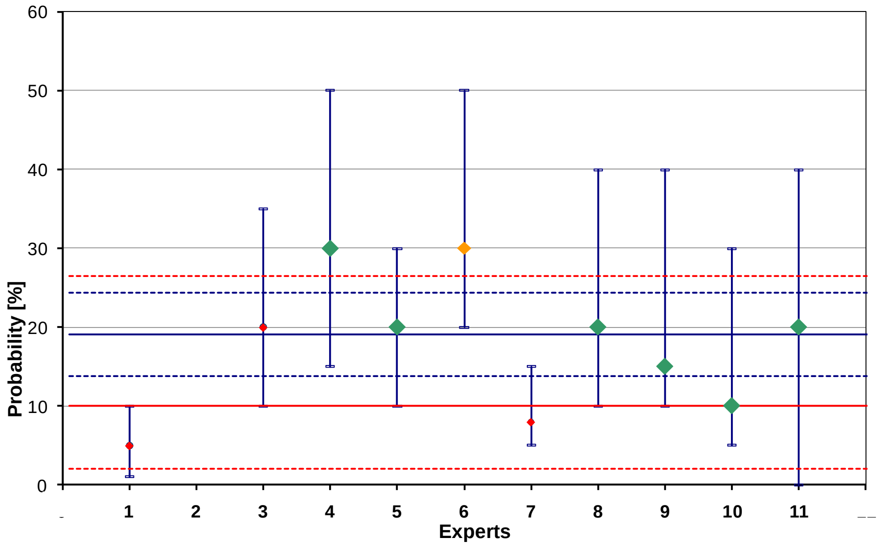


Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

Hazard 20 (Too many animals in the flow at unloading) (Figure 7). The 2 CI overlapped and the experts' replies were rather homogeneous. The on-site estimated P was equal to 10% (CI=2-27%); experts' weighted average P=19% (CI=14-24%). One expert replied that this hazard does was not relevant as normally the number of animals is adequately managed.

Figure 7. On-site and expert-opinion-based exposure assessment for the H20 "Too many animals in the flow at unloading"

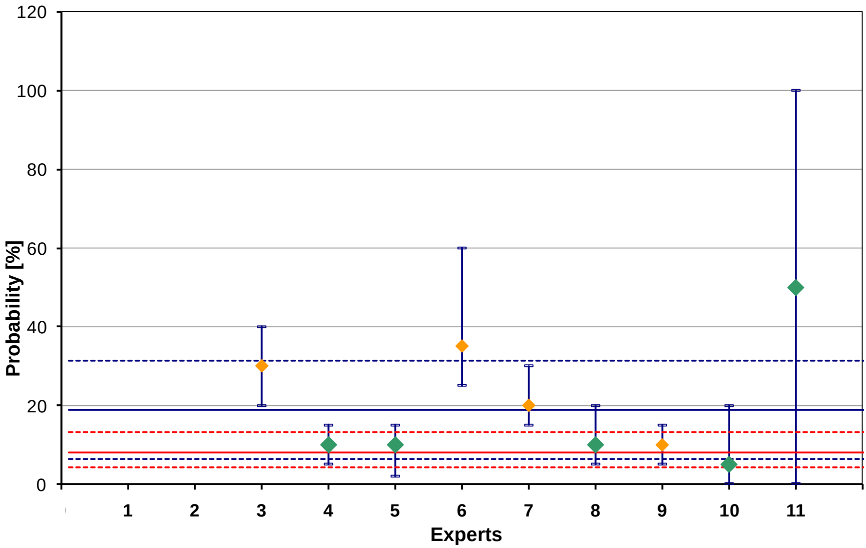


Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

Hazard 11 (Sudden noise due to other than the truck at unloading) (Figure 8). The 2 CI overlapped and the variability between the experts (Di=0.41) was relatively higher than in all other cases where the results were adhering (Di in all other cases was ≤ 0.27). This may show certain difficulty in assessing this hazard.

Figure 8. On-site and expert-opinion-based exposure assessment for the H11 "Sudden noise due to other than the truck at unloading"



Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

4.2.2 Inconsistent results

Inconsistent results occurred when the confidence intervals of the on-site and experts' estimates did not overlap and the experts replies were homogeneous (Di<0.6).

Table 16. Inconsistent results for exposure assessment

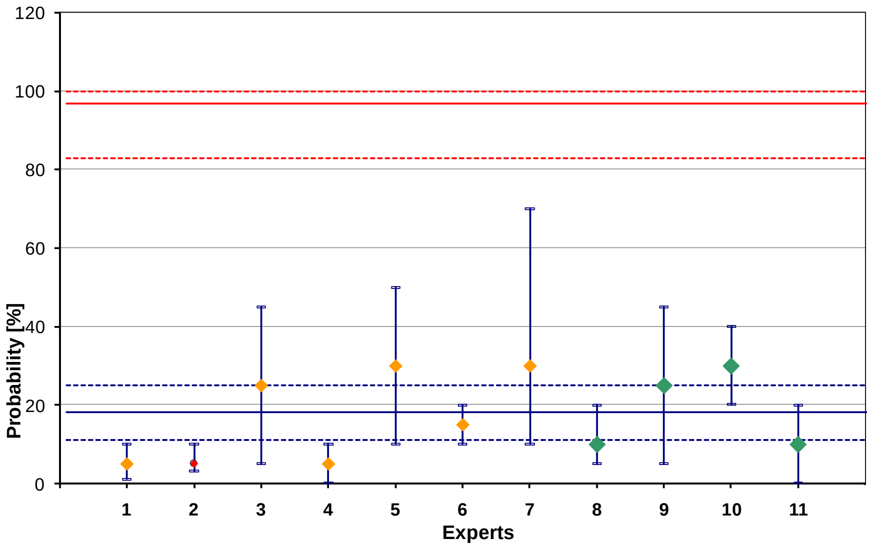
	H3		H38		H69		H37	
	Inadequate handling technique at U		Inadequate handling technique in C		Inadequate handling technique in E		Prod in C	
	Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	47	17	97	18	40	20	7	15
Low	28	7	83	11	32	13	4	10
Up	66	27	100	25	46	27	9	20
N	30*	/	30*	/	197***	/	427*	/
K	14	/	29	/	77	/	28	/
Di	/	0.37	/	0.27	/	0.26	/	0.20
"No"	/	1	/	0	/	0	/	0
	H68		**H82**		**H22**		**H114**	
	Prod in E		**Too fast animal flow in E**		**Too fast animal flow at U**		**Delayed stunning**	
	Est	**Exp**	**Est**	**Exp**	**Est**	**Exp**	**Est**	**Exp**
P	64	17	0	12	0	26	51	15
Low	59	13	0	7	0	17	42	9
Up	68	21	1	17	10	35	61	21
N	424*	/	424*	/	30*	/	115*	/
K	270	/	0	/	0	/	59	/
Di	/	0.17	/	0.20	/	0.29	/	0.24
"No"	/	0	/	2	/	1	/	1

H: hazard
Est: on-site estimates (%)
Exp: experts' weighted values (%)

P: probability of occurrence of the hazard, namely: in Est: on-site estimated average; in Exp: weighted average
Low: lower limit; Up: upper limit
N: number of units of observations (e.g. *individuals, **groups, ***subgroups)
K: number of animals exposed to the hazard
Di: dispersion coefficient of the expert judgement
"No": number of experts that replied that the H does not exist

Hazard 38 (Inadequate handling technique in the corridors) (Figure 9).This hazard was observed with a much higher frequency in the on-site observations (on-site estimated P=97%, CI=82-100% against experts' P=18%, CI=11-25%). The 2 CI did not overlap and the experts' variability was low (Di=0.27).

Figure 9. On-site and expert-opinion-based exposure assessment for the H38 "Inadequate handling technique in the corridors"

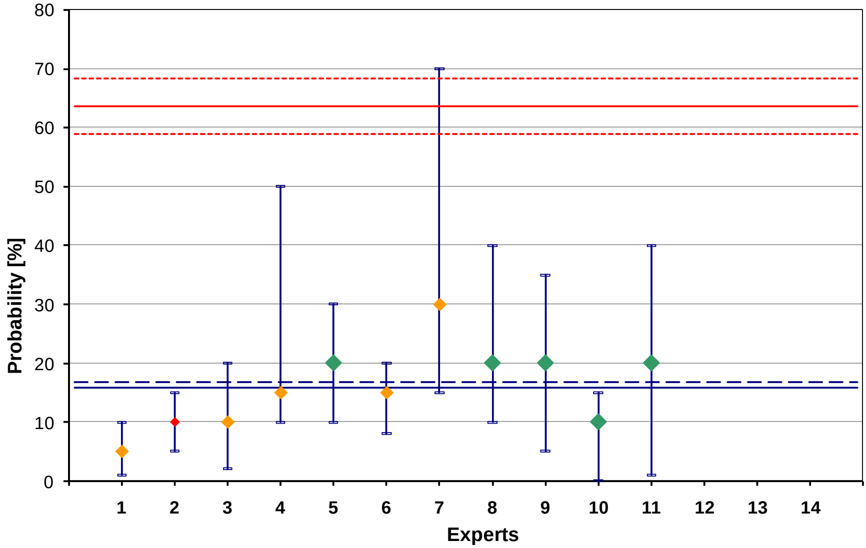


Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

Hazard 68 (Prod at the entrance of the stunning box) (Figure 10). On-site estimated P=64% (CI=59-68%) much higher than experts' weighted average (P=17%; CI=13-21%). The 2 CI did not overlap; low variability between the experts (Di=0.17).

Figure 10. On-site and expert opinion-based exposure assessment for the H68 "Energised prod at the entrance of the stunning box"

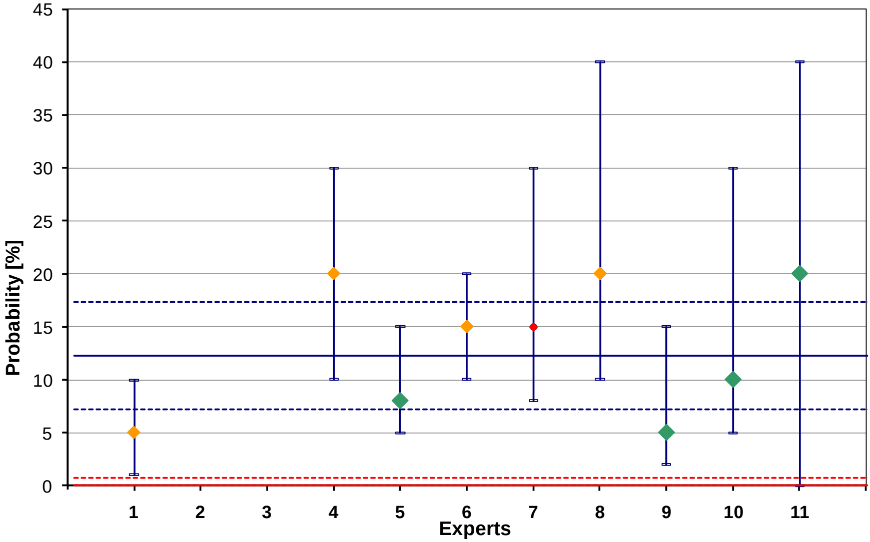


Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

Hazard 82 (Too fast animal flow at unloading at the entrance of the stunning box). This hazard was never observed in the on-site observations (P=0%; CI=0-1%). As opposite, the experts estimated P=12% (CI=7-17%) and low variability (Di=0.20).

Figure 11. On-site and expert opinion-based exposure assessment for the H82 "Too fast animal flow at unloading at the entrance of the stunning box"



Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

Table 17. Inconsistent results for exposure assessment (cont')

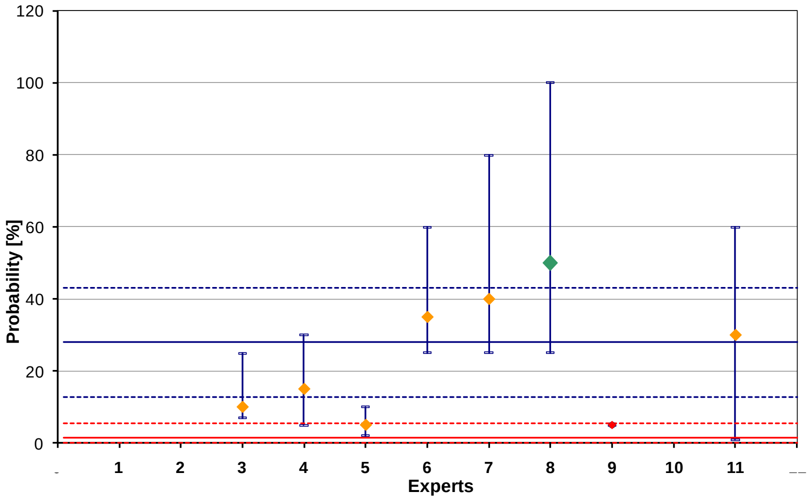
	H90		H59		H45		H31	
	Too narrow corridors in E		Too narrow corridors in C		Sudden noise in C		Gates' opening size too small at U	
	Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	0	12	0	9	2	28	0	14
Low	0	8	0	5	0	13	0	10
Up	1	16	1	13	5	43	1	18
N	424*	/	427*	/	132***	/	521*	/
K	0	/	0	/	2	/	0	/
Di	/	0.15	/	0.17	/	0.40	/	0.15
"No"	/	3	/	3	/	3	/	2

	H62		H33		H64		H108	
	Gates' opening size too small in C		Faulty operating non-return gates at U		Faulty operating non-return gates in C		Faulty operating non-return gate in the stunning box	
	Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	0	12	0	7	0	11	2	10
Low	0	9	0	4	0	6	0	8
Up	1	15	1	10	1	17	6	12
N	427*	/	521*	/	427*	/	115*	/
K	0	/	0	/	0	/	2	/
Di	/	0.11	/	0.13	/	0.21	/	0.07
"No"	/	3	/	3	/	4	/	2

Same legend as Table 16

H45 (Sudden noise in the corridors) (Figure 12). Although less than for the other hazards in this category of inconsistent results[32], the experts' replies were homogenous (Di=0.40) and the 2 CI did not overlap (on-site estimated P=2%, CI=0.2-5%; experts' weighted average= 28%, CI=13-43%).

Figure 12. On-site and expert opinion-based exposure assessment for the H45 "Sudden noise in the corridors"



Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

[32] For the majority of the hazards in the category inconsistent results, variability between the experts was ≤0.29.

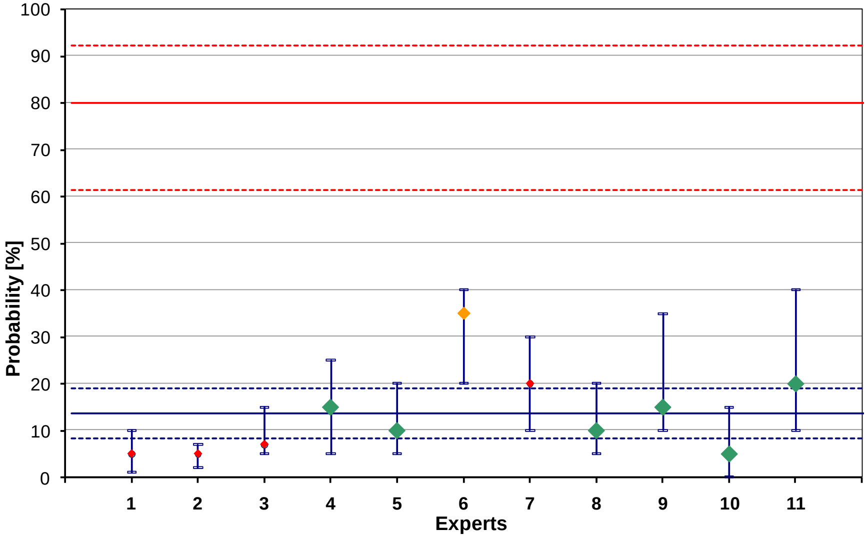
Table 18. Inconsistent results for exposure assessment (cont')

	H49		H44		H75		H10	
	Too many animals in the flow in C		Sudden change of light intensity in C		Sudden change of light intensity in E		Sudden change of light intensity at U	
	Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	80	14	0	16	0	23	0	23
Low	61	8	0	6	0	9	0	8
Up	92	19	1	25	1	37	1	37
N	30*	/	427*	/	424*	/	248***	/
K	24	/	0	/	0	/	0	/
Di	/	0.23	/	0.34	/	0.40	/	0.41
"No"	/	0	/	2	/	0	/	3
	H18		**H47**		**H79**		**H120**	
	Dirty floor at U		Dirty floor in C		Dirty floor in E		No check after stunning	
	Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	60	21	73	20	87	22	52	20
Low	41	11	54	9	69	9	43	14
Up	77	32	88	31	96	35	62	27
N	30*	/	30*	/	30*	/	115*	/
K	18	/	22	/	26	/	60	/
Di	/	0.36	/	0.39	/	0.38	/	0.20
"No"	/	1	/	1	/	3	/	2

Same legend as Table 16

Hazard 49 (Too many animals in the flow in the corridors) (Figure 13). Although for all experts this hazard was relevant (i.e. they agreed on including it), the weighted average P was rather low (P=14%, CI=8-19%), compared to the on-site estimated occurrence (P=80%, CI=61-92%). The variability among the experts was low (Di=0.23). This reflected inconsistent results.

Figure 13. On-site and expert opinion-based exposure assessment for the H49 "Too fast animal flow at unloading at the entrance of the stunning box"

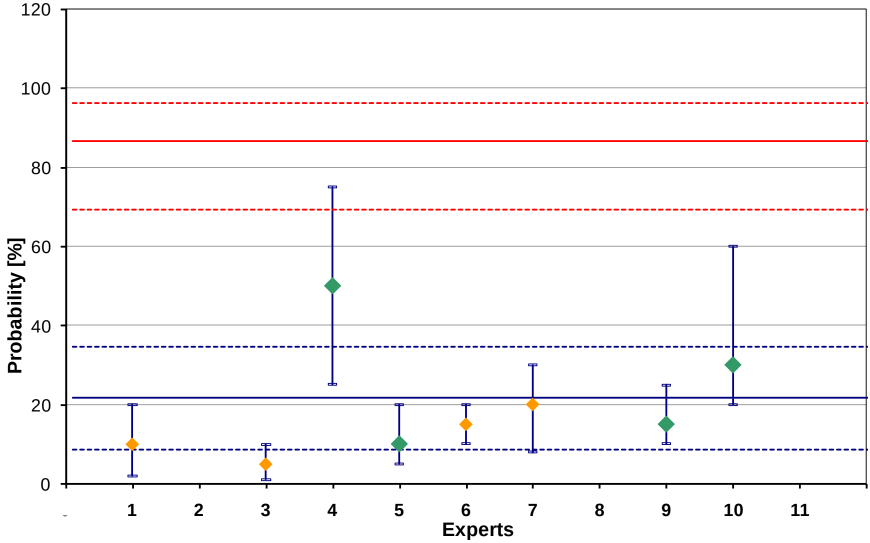


Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

Hazard 79 (Dirty floor at the entrance of the stunning box) (Figure 14). The 2 CI did not overlap (on-site estimated P=87%, CI=70-96%; experts' weighted average equal to 22%, CI=9-35%) and the experts' replies were somehow homogeneous (Di=0.38).

Figure 14. On-site and expert opinion-based exposure assessment for the H79 "Dirty floor at the entrance of the stunning box"



Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

4.2.3 High variability between experts' replies

The hazards for which the variability between the experts' replies was very high (Di ≥ 0.6) are illustrated in the Table and below. Some examples are presented in Figure 15 and Figure 16.

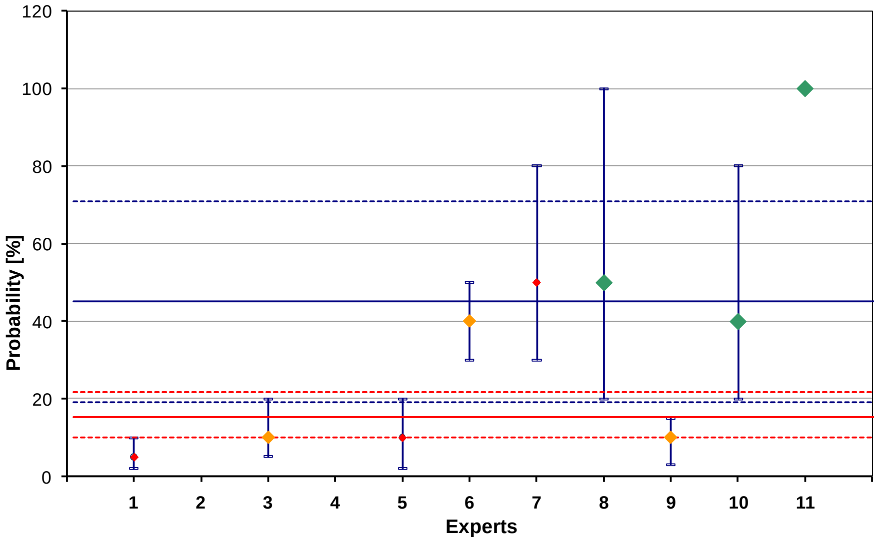
Table 19. Hazards for which the variability between the experts on exposure assessment was very high (Di≥0.6)

	H9		H76		H94		H13	
	Truck noise		Sudden noise in E		Overall and continuous noise		Open / low side barriers on the truck	
	Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	15	45	7	32	100	51	7	33
Low	10	19	3	7	93	27	1	6
Up	22	71	13	56	100	75	22	60
N	164***	/	131***	/	424*	/	30*	/
K	25	/	9	/	424	/	2	/
Di	/	0.66	/	0.73	/	0.62	/	0.72
"No"	/	2	/	1	/	2	/	2

H: hazard
Est: on-site estimates (%)
Exp: experts' weighted values (%)
P: probability of occurrence of the hazard, namely: in Est: on-site estimated average; in Exp: weighted average
Low: lower limit; Up: upper limit
N: number of units of observations (e.g. *individuals, **groups, ***subgroups)
K: number of animals exposed to the hazard
Di: dispersion coefficient of the expert judgement
"No": number of experts that replied that the H does not exist

Hazard 9 (Truck noise) (Figure 15). The experts' weighted average was equal to 45% (CI=19-71%), with very high variability (Di=0.66). For 2 experts this hazard was not relevant as according to their knowledge the animals do not seem to be bothered by the noise of the truck.

Figure 15. On-site and expert opinion-based exposure assessment for H9 "Noise due to the truck at unloading"

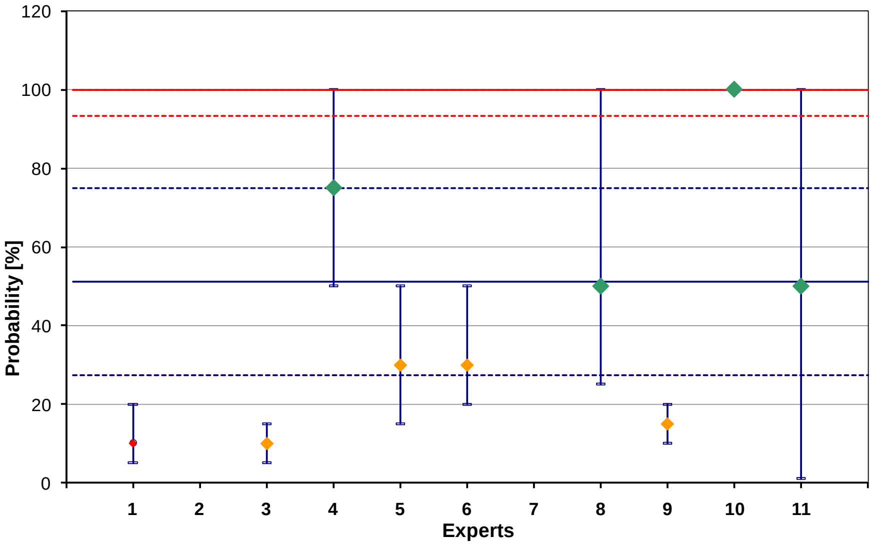


Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI

Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

Hazard 94 (Overall and continuous noise originating in the stunning box) (Figure 16). The experts' replies showed a rather high variability (Di=0.62). The on-site estimated P was equal to 100% (CI=93-100%); the experts' weighted average was equal to51% (CI=27-75%). It is interesting to note that out of all hazards analysed, this is the hazard that was scored by the experts with the highest probability of occurrence.

Figure 16. On-site and expert opinion-based exposure assessment for H94 "Overall and continuous noise originating in the stunning box"



Continuous red horizontal line: on-site estimate
Dashed red horizontal lines: confidence bands
Blue continuous horizontal line: experts' weighted average
Blue dashed horizontal lines: confidence bands
Vertical blue lines: single expert point estimate and CI
Experts' uncertainty:
Big green: low uncertainty (weight=2); **Orange**: medium uncertainty (weight=1); **Small red**: high uncertainty (weight=0.5)

4.3 Comparison of likelihood of the adverse effects based on data collected on-site and on expert opinion

When an animal is exposed to a hazard, a range of adverse effects of different severity levels may occur (i.e. severity 1, limited; severity 2, moderate; severity 3, acute; and severity 4, critical) (section 2.7.2.2).

For each hazard, likelihood of causing the various adverse effects was estimated as described in sections 2.7.2.4 (using data collected on-site) and 2.7.3 (using expert opinion). In this section the results of such assessments are compared[33].

Out of the 42 hazards assessed as relevant to beef cattle at slaughter, 18 hazards were either never observed during the on-site observations or the number of animals exposed to them was < 5 (section 4.2). Thus the results of on-site-estimated and expert opinion-based likelihood of the adverse effects were compared for 24 hazards.

The analysis was carried out using the same principles as for the comparison of the results of exposure assessment, i.e. considering the overlaps of the confidence intervals of the estimates and the variability between the experts' responses. Looking at the results, the variability between the experts was considered high when ≥ 0.5 (Table 20 - Table 25).

The results of the comparison of the on-site and expert-opinion-based likelihood of the adverse effects (by severity levels) and can be summarised as follows:

1. For no hazard the results of all 4 severity levels were adhering (i.e. it never occurred that all 4 confidence intervals overlapped and Di<0.5 for all 4 severity levels).
2. In no case the results of all 4 severity levels were conflicting (i.e. it never occurred that the confidence bands did not overlap for all 4 severity levels and Di<0.5).
3. In most cases (23 out of 24 hazards) the variability between the

[33] Duration of the adverse effects was not compared since it was not assessed in the empirical risk assessment. Duration was considered in the final risk estimate only.

experts' replies was relatively high (Di≥0.5) for severity 1 and severity 2 adverse effects.

4. For some hazards the experts showed certain variability also for probability of occurrence of severity 3 adverse effects (H96 "Inadequate handling technique at stunning" and H120 "No check of unconsciousness after stunning"), and for severity 3 and 4 adverse effects (H115 "Wrong shot" and H116 "Delayed second shot").
5. Low variability between the experts responses for all 4 probabilities of occurrence of the adverse effects occurred for 3 hazards only (H38 "Inadequate handling technique in the corridors"; H80 "Too many animals in the flow at the entrance of the stunning box"; and H94 "Overall and continuous noise in at the entrance of the stunning box").
6. The experts assigned to most of the hazards (23 out of 24) a higher probability of causing adverse effects of limited or moderate severity (levels 1 and 2) and a gradually lower probability of determining severity 3 and 4 adverse effects and this was not always reflected by the on-site observations. The only exceptions were represented by the hazards: H120 (No check of unconsciousness after stunning), for which the highest probability of occurrence was assigned to severity 3 adverse effects, followed by severity 4, 2 and 1 adverse effects; and H115 (Wrong shot) and H116 (Delayed second shot), for which the maximum P of occurrence was assigned to severity 3 (H115) and 4 (H116) adverse effects.

The above aspects were discussed in the relevant section (5.2). All results and related comparisons are reported in Table 20 - Table 25 and some specific examples are reported immediately following the tables.

Table 20. Comparison of the results of empirical and expert opinion-based likelihood of the adverse effects (by severity levels)

H number		H2		H37		H68		H15	
H name		Prod at U		Prod in C		Prod in E		Steep truck ramp	
Estimates		Est	Exp	Est	Exp	Est	Exp	Est	Exp
S1	Av	0	43	/	33	/	31	0	31
	Lo	na	24	/	15	/	14	0	15
	Up	na	63	/	52	/	48	5	46
	K	/	/	/	/	/	/	61	/
	L	/	/	/	/	/	/	0	/
S2	Av	0	12	0	13	0	13	0	19
	Lo	0	8	0	5	0	4	0	2
	Up	8	17	10	20	1	22	5	36
	K	35	/	28	/	270	/	61	/
	L	0	/	0	/	0	/	0	/
S3	Av	100	1	100	0	100	1	0	6
	Lo	92	0	90	0	99	0	0	2
	Up	100	2	100	1	100	5	5	10
	K	35	/	28	/	270	/	61	/
	L	35	/	28	/	270	/	0	/
S4	Av	0	0	/	0	/	0	0	0
	Lo	na	0	/	0	/	0	0	0
	Up	na	0	/	0	/	0	5	1
	K	/	/	/	/	/	/	61	/
	L	/	/	/	/	/	/	0	/
	Di1	/	0.55	/	0.59	/	0.55	/	0.41
	Di2	/	0.20	/	0.35	/	0.40	/	0.51
	Di3	/	0.18	/	0.23	/	0.45	/	0.19
	Di4	/	0.00	/	0.00	/	0.14	/	0.13

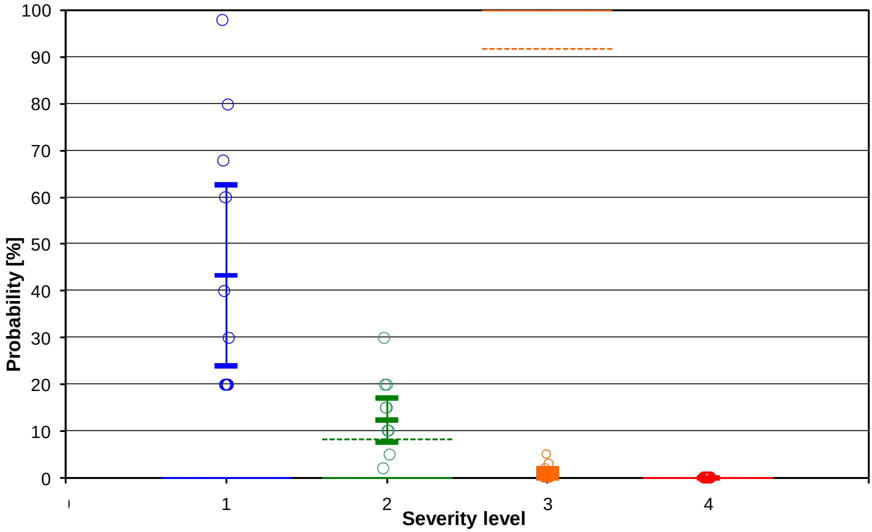
H: hazard
Est: on-site estimates (%)
Exp: experts' weighted values (%)

S1, S2, S3, S4: probability of occurrence of the adverse effects by severity levels, namely: in Est: on-site estimated average; in Exp: weighted average
Av: average; Lo: lower limit; Up: upper limit
K: number of animals exposed to the hazard
L: number of animals showing the adverse effect
Di1, Di2, Di3, Di4: dispersion coefficient of the expert judgement for the 4 probabilities of occurrence of the adverse effects (by severity levels)

Hazard 2 (Energise prod at unloading) (Figure 17). This hazard represented a particular case. Indeed for the on-site observations this hazard was always pre-assigned (by assumption) to a severity level equal to 3; this was also supported by the (section 2.7.2.2). As opposite, none of the experts agreed on this value (severity 3 adverse effects P=0%), with very low variability (Di=0.18). Only 3 experts agreed that Severity 3 adverse effects may occur when the animals are prodded, although assigning to it a rather low value (one expert scored severity 3 P=2%; one expert scored severity 3 P=5%; and one expert scored severity 3 P=3%). According to the majority of the experts, when the animals are prodded, severity 1 adverse effects had the highest probability to occur (weighted average P=43%, CI=24-63%), although with relatively high variability between them (severity 1 Di=0.55). All experts agreed that severity level 4 is never caused by the energised prod[34].

[34] Similar considerations apply for this hazard in the corridors (H37) and at the entrance of the stunning box (H68).

Figure 17. On-site and expert opinion-based likelihood of the adverse effect (by severity levels) for the H2 "Energised prod at unloading"



Di1	Di2	Di3	Di4
0.55	**0.20**	**0.18**	**0.00**

On site observations:

- **Wide continuous horizontal line**: on-site estimate
- **Wide dashed horizontal lines**: confidence bands

Di1, Di2, Di3, and Di4: Dispersion coefficient of the expert judgement related to the 4 severity levels

Experts' replies:

- **Narrow intermediate horizontal bar**: experts' weighted average
- **Narrow horizontal bars at the extremes**: confidence bands
- **Circles**: experts' point estimates

Experts' uncertainty calculated in the weighted average

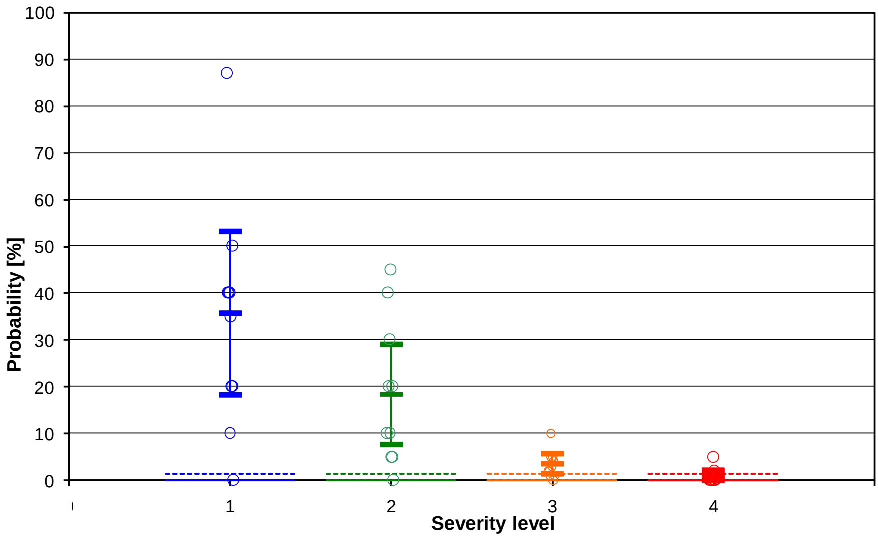
Table 21. Comparison of the results of empirical and expert opinion-based likelihood of the adverse effects (by severity levels) (cont')

H number		H18		H47		H79		H111	
H name		Dirty floor at U		Dirty floor in C		Dirty floor in E		Too small stunning box	
Estimates		Est	Exp	Est	Exp	Est	Exp	Est	Exp
S1	Av	0	36	0	30	0	34	0	36
	Lo	0	18	0	12	0	10	0	11
	Up	1	53	1	49	1	58	39	61
	K	223	/	22	/	420	/	6	/
	L	0	/	0	/	0	/	0	/
S2	Av	0	18	0	12	0	13	100	25
	Lo	0	8	0	7	0	7	61	11
	Up	1	29	2	18	1	19	100	39
	K	223	/	299	/	420	/	6	/
	L	0	/	0	/	0	/	6	/
S3	Av	0	4	0	1	0	2	0	5
	Lo	0	1	0	0	0	0	0	0
	Up	1	6	1	3	1	4	39	15
	K	223	/	22	/	420	/	6	/
	L	0	/	0	/	1	/	0	/
S4	Av	0	1	0	0	0	0	0	0
	Lo	0	0	0	0	0	0	0	0
	Up	1	2	1	1	1	1	39	0
	K	223	/	22	/	420	/	6	/
	L	0	/	0	/	0	/	0	/
	Di1	/	0.50	/	0.55	/	0.59	/	0.50
	Di2	/	0.38	/	0.24	/	0.21	/	0.32
	Di3	/	0.16	/	0.18	/	0.15	/	0.40
	Di4	/	0.18	/	0.14	/	0.17	/	0.00

Same legend as Table 20

Hazard 18 (Dirty floor in the unloading bay) (Figure 18**)**. No adverse effects were observed on-site as a consequence to this hazard at unloading. The experts estimates were adhering to the on-site observations for severity levels 3 and 4, (i.e. overlapping CI and low variability). For severity 2 the results were inconsistent whereas for severity 1 the experts showed certain variability (Di1=0.50).

Figure 18. On-site and expert opinion-based likelihood of the adverse effect (by severity levels) for the H18 "Dirty floor in the unloading bay"



Di1	Di2	Di3	Di4
0.50	0.38	0.16	0.18

On site observations:

- **Wide continuous horizontal line**: on-site estimate
- **Wide dashed horizontal lines**: confidence bands

Di1, Di2, Di3, and Di4: Dispersion coefficient of the expert judgement related to the 4 severity levels

Experts' replies:

- **Narrow intermediate horizontal bar**: experts' weighted average
- **Narrow horizontal bars at the extremes**: confidence bands
- **Circles**: experts' point estimates

Experts' uncertainty calculated in the weighted average

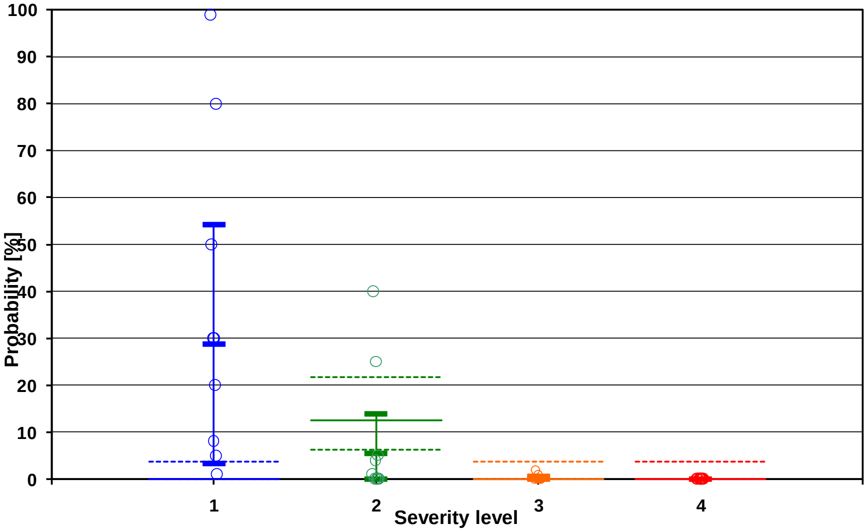
Table 22. Comparison of the results of empirical and expert opinion-based likelihood of the adverse effects (by severity levels) (cont')

H number		H9		H38		H69		H3	
H name		Truck noise		Inadequate handling technique in C		Inadequate handling technique in E		Inadequate handling technique	
Estimates		Est	Exp	Est	Exp	Est	Exp	Est	Exp
S1	Av	0	29	72	28	10	26	7	34
	Lo	0	3	53	14	0	11	0	19
	Up	4	54	87	41	14	40	34	48
	K	80	/	29	/	154	/	14	/
	L	0	/	21	/	14	/	1	/
S2	Av	13	6	3	12	48	18	7	17
	Lo	6	0	0	2	40	5	0	3
	Up	22	14	18	22	56	32	34	32
	K	80	/	29	/	154	/	14	/
	L	10	/	1	/	74	/	1	/
S3	Av	0	0	0	2	12	2	0	2
	Lo	0	0	0	0	8	0	0	0
	Up	4	1	10	4	19	6	19	5
	K	80	/	29	/	154	/	14	/
	L	0	/	0	/	19	/	0	/
S4	Av	0	0	0	0	0	0	0	0
	Lo	0	0	0	0	0	0	0	0
	Up	4	0	10	0	2	0	19	0
	K	80	/	29	/	154	/	14	/
	L	0	/	0	/	0	/	0	/
	Di1	/	0.71	/	0.45	/	0.50	/	0.43
	Di2	/	0.46	/	0.45	/	0.52	/	0.53
	Di3	/	0.12	/	0.21	/	0.33	/	0.21
	Di4	/	0.00	/	0.00	/	0.14	/	0.00

Same legend as Table 20

Hazard 9 (Truck noise) (Figure 19). The CI of severity levels 2, 3 and 4 overlapped and the experts' replies were homogeneous, showing adhering results. The experts' replies on severity 1 adverse effects showed a considerably high variability (Di1=0.71). This may be due to the fact that this hazard was difficult to assess as such (as reflected also by the results of exposure assessment, for which experts' variability was also high); and/or to the difficulty to quantify adverse effects, especially those of low severity (section 5.2).

Figure 19. On-site and expert opinion-based likelihood of the adverse effect (by severity levels) for the H9 "Truck noise"



Di1	Di2	Di3	Di4
0.71	0.46	0.12	0.00

On site observations:

- **Wide continuous horizontal line**: on-site estimate
- **Wide dashed horizontal lines**: confidence bands

Di1, Di2, Di3, and Di4: Dispersion coefficient of the expert judgement related to the 4 severity levels

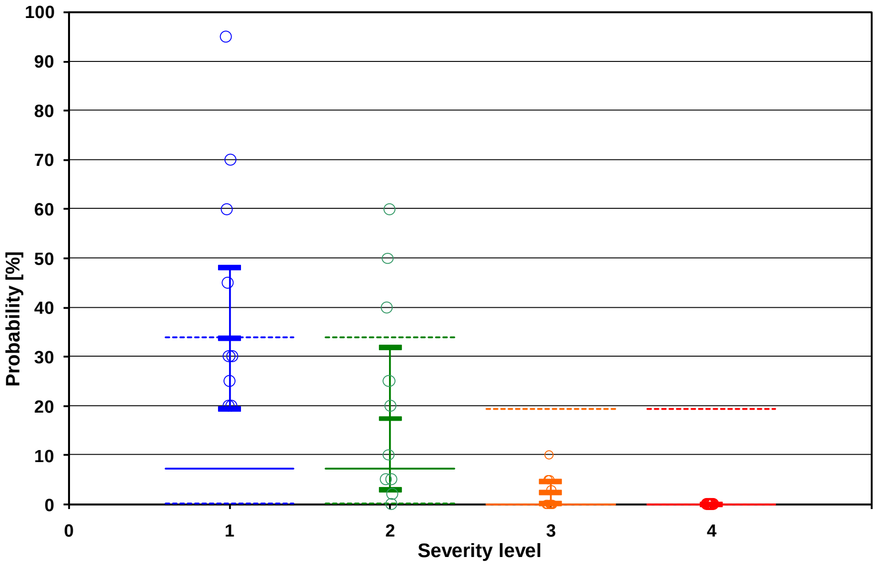
Experts' replies:

- **Narrow intermediate horizontal bar**: experts' weighted average
- **Narrow horizontal bars at the extremes**: confidence bands
- **Circles**: experts' point estimates

Experts' uncertainty calculated in the weighted average

Hazard 3 (Inadequate handling technique at unloading) (Figure 20). For P of occurrence of severity 1, 3 and 4 the results were adhering (i.e. overlapping CI and low variability between the experts). The variability between the experts' replies was relatively higher for Severity 2 adverse effects (Di2=0.53). All experts agreed that severity 4 adverse effects as a consequence of the exposure to this hazard would never occur (P=0%, CI=0-0%).

Figure 20. On-site and expert opinion-based likelihood of the adverse effect (by severity levels) for the H3 "Inadequate handling technique at unloading"



Di1	Di2	Di3	Di4
0.43	0.53	0.21	0.00

On site observations:

- **Wide continuous horizontal line**: on-site estimate
- **Wide dashed horizontal lines**: confidence bands

Di1, Di2, Di3, and Di4: Dispersion coefficient of the expert judgement related to the 4 severity levels

Experts' replies:

- **Narrow intermediate horizontal bar**: experts' weighted average
- **Narrow horizontal bars at the extremes**: confidence bands
- **Circles**: experts' point estimates

Experts' uncertainty calculated in the weighted average

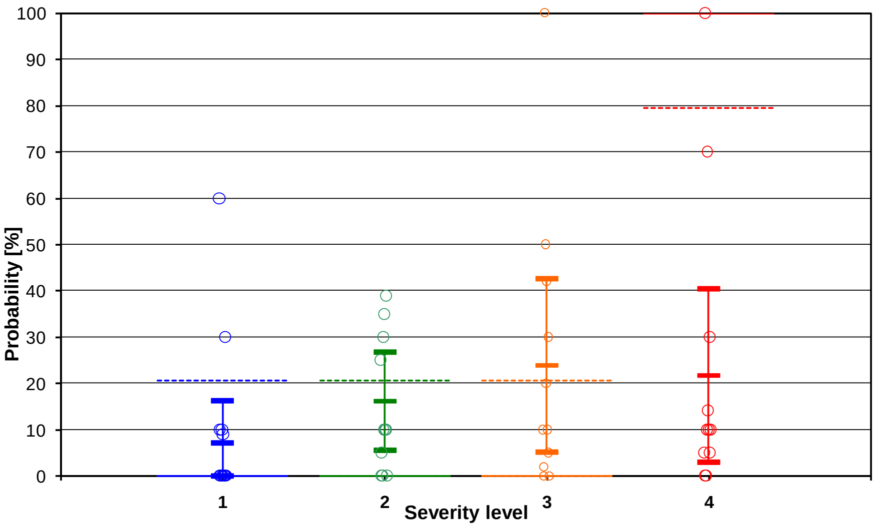
Table 23. Comparison of the results of empirical and expert opinion-based likelihood of the adverse effects (by severity levels) (cont')

H number		H114		H115		H116		H11	
H name		Delayed stunning		Wrong shot		Delayed second shot		Noise due to other than the truck	
Estimates		Est	Exp	Est	Exp	Est	Exp	Est	Exp
S1	Av	0	39	0	7	0	12	2	22
	Lo	0	20	0	0	0	0	0	12
	Up	5	59	21	16	22	30	13	32
	K	59	/	13	/	12	/	41	/
	L	0	/	0	/	0	/	1	/
S2	Av	0	17	0	16	0	10	0	7
	Lo	0	10	0	5	0	0	0	0
	Up	5	25	21	27	22	21	7	20
	K	59	/	13	/	12	/	41	/
	L	0	/	0	/	0	/	0	/
S3	Av	10	1	0	24	0	14	0	0
	Lo	4	0	0	5	0	0	0	0
	Up	21	4	21	43	22	33	7	1
	K	59	/	13	/	12	/	41	/
	L	6	/	0	/	0	/	0	/
S4	Av	0	0	100	22	100	29	0	0
	Lo	0	0	79	3	78	0	0	0
	Up	5	0	100	41	100	58	7	0
	K	59	/	13	/	12	/	41	/
	L	0	/	13	/	0	/	0	/
	Di1	/	0.53	/	0.51	/	0.80	/	0.31
	Di2	/	0.25	/	0.42	/	0.54	/	0.67
	Di3	/	0.31	/	0.64	/	0.79	/	0.12
	Di4	/	0.00	/	0.67	/	0.95	/	0.00

Same legend as Table 20

Hazard 115 (Wrong shot) (Figure 21). Every time an inadequate shot was detected in the on-site observations, the related adverse effect was classified as severity 4 (i.e. *Extreme distress, pain, fear or anxiety, that if persist would be life-threatening*). This was not agreed by all the experts, whose replies were not homogeneous at all, not only for severity 4, but also for severity 3 and 1 adverse effects. This was interpreted as a discrepancy in interpreting the question in the questionnaire (section on discussion 5.2). Similar considerations apply to **Hazard 116 (Second shot delayed)**.

Figure 21. On-site and expert opinion-based likelihood of the adverse effect (by severity levels) for the H115 "Wrong shot"



Di1	Di2	Di3	Di4
0.80	0.42	0.79	0.95

On site observations:

- **Wide continuous horizontal line**: on-site estimate
- **Wide dashed horizontal lines**: confidence bands

Di1, Di2, Di3, and Di4: Dispersion coefficient of the expert judgement related to the 4 severity levels

Experts' replies:

- **Narrow intermediate horizontal bar**: experts' weighted average
- **Narrow horizontal bars at the extremes**: confidence bands
- **Circles**: experts' point estimates

Experts' uncertainty calculated in the weighted average

Hazard 114 (Delayed stunning). As for the majority of the other hazards, the experts assigned a gradually decreasing probability of occurrence from severity 1 to severity 4 adverse effects. This was not confirmed by the on-site observations, as when this hazard occurred the adverse effects detected were “Falling”, “Slipping” or “Pushing against the stunning box” (all associated to a severity level equal to 3). On the other hand, no adverse effects of severity 1 or 2 were detected in the on-site data collections. As in most of the other cases the variability between the experts was relatively high for severity 1.

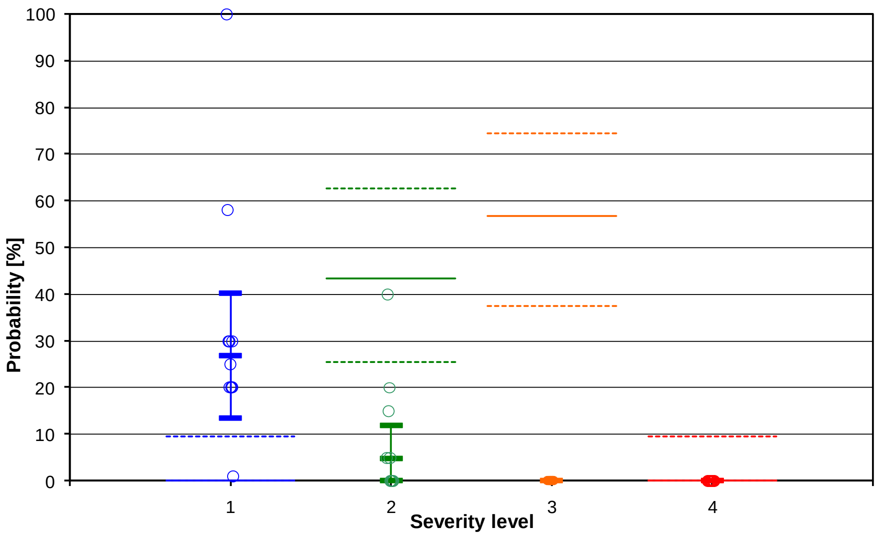
Table 24. Comparison of the results of empirical and expert opinion-based likelihood of the adverse effects (by severity levels) (cont')

H number		H49		H80		H94		H12	
H name		Too many animals in the flow in C		Too many animals in the flow in E		Overall and continuous noise in E		Too rapid / uncontrolled funnelling at U	
Estimates		Est	Exp	Est	Exp	Est	Exp	Est	Exp
S1	Av	0	26	0	17	0	27	15	27
	Lo	0	7	0	8	0	14	4	12
	Up	1	45	1	27	10	40	34	42
	K	24	/	227	/	30	/	27	/
	L	0	/	0	/	0	/	4	/
S2	Av	25	7	26	7	43	5	11	10
	Lo	10	2	20	0	26	0	2	3
	Up	47	12	32	14	63	12	29	16
	K	24	/	227	/	30	/	27	/
	L	6	/	58	/	13	/	3	/
S3	Av	13	0	2	1	57	0	0	1
	Lo	3	0	1	0	37	0	0	0
	Up	32	1	5	3	75	0	11	1
	K	24	/	227	/	30	/	27	/
	L	3	/	4	/	17	/	0	/
S4	Av	0	0	0	0	0	0	0	0
	Lo	0	0	0	0	0	0	0	0
	Up	12	1	1	0	10	0	11	0
	K	24	/	227	/	30	/	27	/
	L	0	/	0	/	0	/	0	/
	Di1	/	0.64	/	0.29	/	0.42	/	0.52
	Di2	/	0.30	/	0.33	/	0.47	/	0.34
	Di3	/	0.15	/	0.22	/	0.00	/	0.20
	Di4	/	0.13	/	0.00	/	0.00	/	0.10

Same legend as Table 20

Hazard 94 (Overall and continuous noise in the area close to the entrance of the stunning box) (Figure 22). The results were inconsistent for all severity levels apart from 4. The experts estimated a much lower P of occurrence, with low variability, for severity 2 and 3 adverse effects. As opposite the experts' weighted average for severity 1 adverse effects was much higher than the on-site estimated (P=0%).

Figure 22. On-site and expert opinion-based likelihood of the adverse effect (by severity levels) for the H94 "Overall and continuous noise at the entrance of the stunning box"



Di1	Di2	Di3	Di4
0.42	0.47	0.00	0.00

On site observations:

- **Wide continuous horizontal line**: on-site estimate
- **Wide dashed horizontal lines**: confidence bands

Di1, Di2, Di3, and Di4: Dispersion coefficient of the expert judgement related to the 4 severity levels

Experts' replies:

- **Narrow intermediate horizontal bar**: experts' weighted average
- **Narrow horizontal bars at the extremes**: confidence bands
- **Circles**: experts' point estimates

Experts' uncertainty calculated in the weighted average

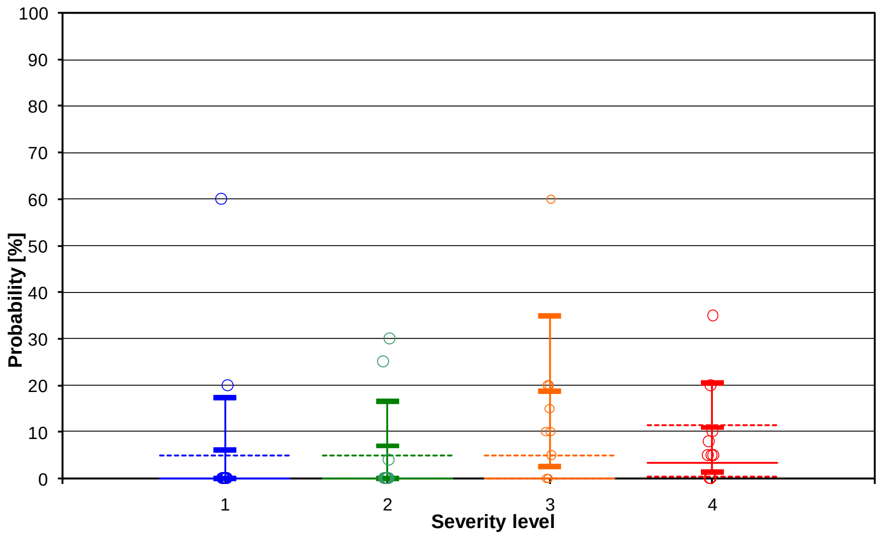
Table 25. Comparison of the results of empirical and expert opinion-based likelihood of the adverse effects (by severity levels) (cont')

H number		H14		H45		H96		H120	
H name		Slippery truck ramp		Sudden noise in C		Inadequate handling technique at stunning		No check after stunning	
Estimates		Est	Exp	Est	Exp	Est	Exp	Est	Exp
S1	Av	0	28	6	18	0	32	0	6
	Lo	0	12	0	5	0	13	0	0
	Up	13	44	29	32	22	51	5	17
	K	21	/	17	/	12	/	60	/
	L	0	/	1	/	0	/	0	/
S2	Av	0	15	0	7	0	12	0	7
	Lo	0	4	0	0	0	1	0	0
	Up	13	26	16	19	22	23	5	17
	K	21	/	17	/	12	/	60	/
	L	0	/	0	/	0	/	0	/
S3	Av	0	3	0	0	8	4	0	19
	Lo	0	1	0	0	0	0	0	3
	Up	13	5	16	1	39	11	5	35
	K	21	/	17	/	12	/	60	/
	L	0	/	0	/	1	/	0	/
S4	Av	0	0	0	0	0	0	3	11
	Lo	0	0	0	0	0	0	0	1
	Up	13	0	16	0	22	1	12	21
	K	21	/	17	/	12	/	60	/
	L	0	/	0	/	0	/	2	/
	Di1	/	0.50	/	0.40	/	0.57	/	0.61
	Di2	/	0.43	/	0.55	/	0.47	/	0.49
	Di3	/	0.15	/	0.12	/	0.52	/	0.54
	Di4	/	0.10	/	0.00	/	0.17	/	0.40

Same legend as Table 20

Hazard 120 (No check of unconsciousness after stunning) (Figure 23). For Severity 2 and 4 adverse effects the results were adhering. For severity 1 and 4 the variability between the experts' replies was high, indication a difficulty in interpreting the question in the questionnaire (section 5.2).

Figure 23. On-site and expert-opinion-based likelihood of the adverse effect (by severity levels) for the H120 "No check of unconsciousness after stunning"



Di1	Di2	Di3	Di4
0.61	0.49	0.54	0.40

On site observations:

- **Wide continuous horizontal line**: on-site estimate
- **Wide dashed horizontal lines**: confidence bands

Di1, Di2, Di3, and Di4: Dispersion coefficient of the expert judgement related to the 4 severity levels

Experts' replies:

- **Narrow intermediate horizontal bar**: experts' weighted average
- **Narrow horizontal bars at the extremes**: confidence bands
- **Circles**: experts' point estimates

Experts' uncertainty calculated in the weighted average

4.4 Comparison of risk characterisation based on data collected on-site and on expert opinion

Finally two risk assessments were performed as described in section 3.3. One risk assessment based on data collected in abattoir and one based on expert opinion.

The risk was estimated for the 24 hazards for which likelihood of occurrence of the adverse effects had been assessed.

The results of the comparison of the two risk estimates are illustrated in Table 26, where the 24 hazards analysed were ranked from the highest to the lowest *on-site estimated* risk. For each hazard the related risk ranked by the experts was also reported.

The results of the comparison of the risk estimates can be summarised as follows:

1. The experts generally estimated lower maximum values.
2. The experts ranked some hazards likewise in the empirical estimates (e.g. H115 “Wrong shot”; H116 “delayed second shot”; or H96 “Inadequate handling technique at stunning”; etc).
3. Other hazards were ranked considerably differently by the experts (Table 26, yellow cells), such as:
    a. H49 and H80 “Too many animals in the corridors and at the entrance of the stunning box”; H94 “Overall and continuous noise in the area close to the entrance of the stunning box”; H68, H2 and H37 “Energised prod in the three phases of pre-slaughter handling” (for these hazards the on-site estimated risk was higher).
    b. H18, 79 and 47 “Dirty floor” in all areas prior to stunning (i.e. unloading bay, corridors, area in front of the entrance of the stunning box); H120 “No check after stunning”; H3 “Inadequate handling technique at unloading”; and H15 “Steep truck ramp” (for these hazards the on-site estimated risk was lower).
4. For some hazards the on-site estimated risk was equal to 0% (e.g. H14”Slippery truck ramp”; H15 “Steep truck ramp”; and H18

"Dirty floor at unloading"), whereas this never happened in the experts' estimates. This was due to the fact that for these hazards no adverse effects were observed in the on-site data collections. For no hazards the experts indicated a 0% probability of occurrence of adverse effects.

All above aspects are discussed (section 5.3).

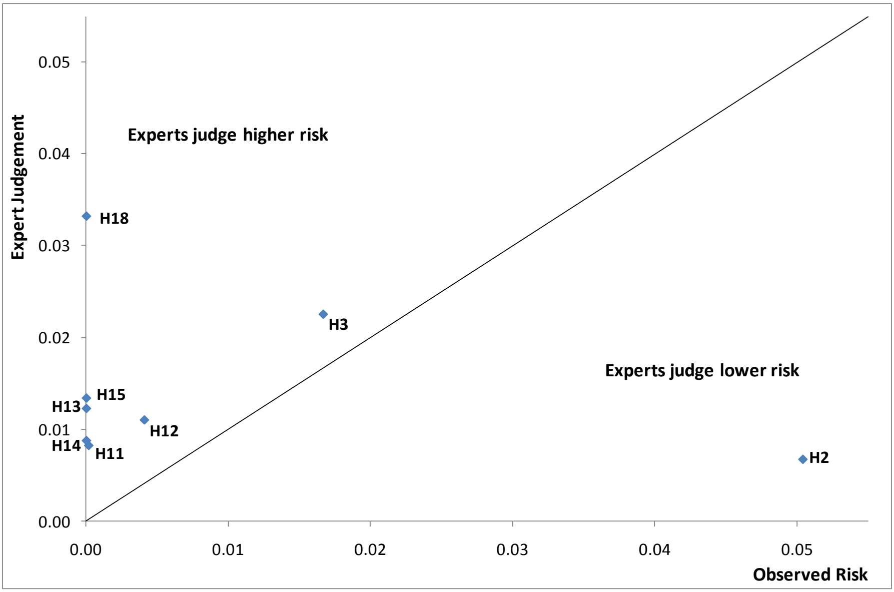
Details on the comparison of the risk estimates based on data collected on-site and on expert opinion are reported in the tables and figures below in this section.

Table 26. Comparison of empirical and expert opinion-based risk estimates for 24 hazards analysed

H	H name	On-site estimate		Experts' estimate	
		Overall risk (%)	Rank	Overall risk (%)	Rank
H68	Energised prod at the entrance of the stunning box	0.4776	1	0.0136	14
H94	Overall and continuous noise in the area close to the entrance of the stunning box	0.4278	2	0.0197	10
H115	Wrong shot	0.2174	3	0.0659	1
H49	Too many animals in the corridors	0.1417	4	0.0067	22
H116	Delayed second shot	0.1043	5	0.0457	3
H69	Inadequate handling technique at the entrance of the stunning box	0.1041	6	0.0201	9
H38	Inadequate handling technique in the corridors	0.0694	7	0.0143	12
H2	Energised prod at unloading	0.0504	8	0.0067	23
H37	Energised prod in the corridors	0.0492	9	0.0107	19
H114	Delayed stunning	0.0391	10	0.0143	13
H80	Too many animals in the area close to the entrance of the stunning box	0.0262	11	0.0052	24
H111	Too small stunning box	0.0174	12	0.0210	8
H120	No check after stunning	0.0174	13	0.0601	2
H3	Inadequate handling technique at unloading	0.0167	14	0.0225	6
H96	Inadequate handling technique at stunning	0.0065	15	0.0133	16
H12	Too rapid/uncontrolled funnelling at unloading	0.0041	16	0.0110	18
H9	Truck noise	0.0032	17	0.0155	11
H79	Dirty floor in the corridors in the area close to the entrance of the stunning box	0.0016	18	0.0258	5
H47	Dirty floor in the corridors	0.0008	19	0.0211	7
H11	Sudden noise due to other than the truck	0.0002	20	0.0082	21
H45	Sudden noise in the corridors	0.0001	21	0.0111	17
H14	Slippery truck ramp	0.0000	22	0.0087	20
H15	Steep truck ramp	0.0000	23	0.0134	15
H18	Dirty floor at unloading	0.0000	24	0.0333	4

Rank: Indicates the ranked risk related to the 24 hazards analysed. In the column “On-site estimate”, the hazards are ranked from the highest estimated risk (H68) to the lowest (H18). The related risk ranked by the experts is listed in the last column on the right hand side. For instance H68 presents the highest on-site estimated risk and is ranked at the 14th level in the experts’ estimates
Yellow cells: hazards for which the related risk was ranked considerably differently in the two risk assessments

Figure 24. Comparison of the risk estimates for the phase "unloading"



	H2	H3	H9	H11
Overall risk (%)	**Energised prod**	**Inadequate handling**	**Truck noise**	**Sudden noise**
Est	0.0504	0.0167	0.0032	0.0002
Exp	0.0067	0.0225	0.0155	0.0082
	H12	**H14**	**H15**	**H18**
Overall risk (%)	**Too rapid/uncontrolled funnelling**	**Slippery truck ramp**	**Steep truck ramp**	**Dirty floor**
Est	0.0041	0.0000	0.0000	0.0000
Exp	0.0110	0.0087	0.0134	0.0333

Est: on-site estimate
Exp: experts' weighted average
The data used for estimating the risk are illustrated in Table 27 and Table 28
For hazards 14, 15 and 18 no adverse effect was observed during the on-site data collections

Table 27. Data used for estimating the overall risk related to the hazards analysed at unloading

H number		H2		H3		H11		H12	
H name		Prod		Inadequate handling		Sudden noise		Too rapid/uncontrolled funnelling	
Estimates		Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	Av	7	11	47	17	8	19	8	19
	Low	5	5	28	7	4	6	4	13
	Up	9	16	66	27	13	31	15	25
S1	Av	0	43	7	34	2	22	15	27
	Low	na	24	0	19	0	12	4	12
	Up	na	63	34	48	13	32	34	42
	D	1	1	2	2	1	1	1	1
S2	Av	0	12	7	17	0	7	11	10
	Low	0	8	0	3	0	0	2	3
	Up	8	17	34	32	7	20	29	16
	D	1	1	2	2	2	2	2	2
S3	Av	100	1	0	2	0	0	0	1
	Low	92	0	0	0	0	0	0	0
	Up	100	2	19	5	7	1	11	1
	D	3	3	3	3	3	3	3	3
S4	Av	0	0	0	0	0	0	0	0
	Low	na	0	0	0	0	0	0	0
	Up	na	0	19	0	7	0	11	0
	D	3	3	3	3	3	3	3	3
S0		0	43	86	47	98	71	74	63

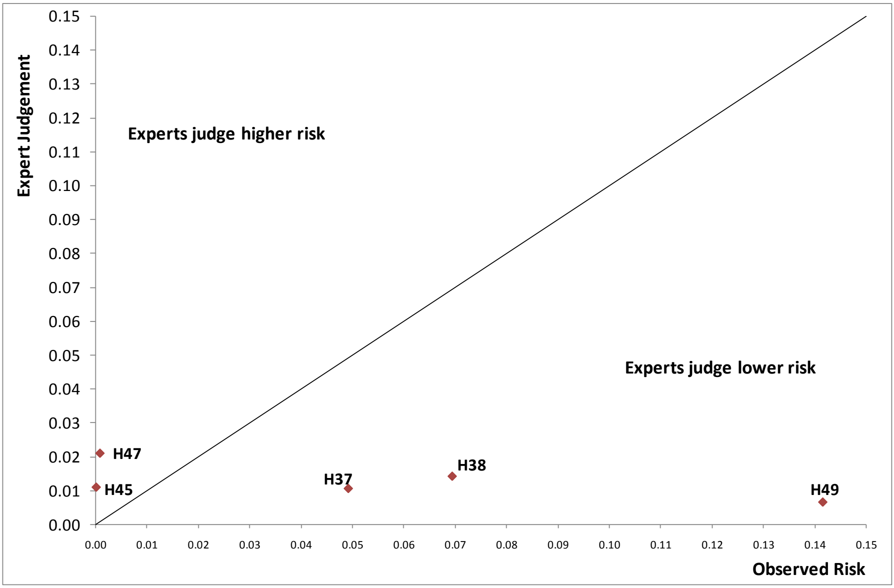
Yellow cells: 0% probability of adverse effect indicated by the experts and thus duration of the adverse effect estimated by the observer
Est: on-site estimate; **Exp**: experts' weighted average; **P**: probability of the hazard; **S1**, **S2**, **S3** and **S4**: probability of the adverse effect by severity level; **Av**: average; **Low** and **Up**: lower and upper limits; **D**: duration of the adverse effect; **S0**: residual category when the sum of the likelihoods of severity 1 to 4 does not reach 100%
In the on-site observations "Energised prod was always associated to severity 3 adverse effects

Table 28. Data used for estimating the overall risk related to the hazards analysed at unloading (cont')

H number		H13		H14		H15		H18	
H name		Open/short truck barriers		Slippery truck ramp		Steep truck ramp		Dirty floor	
Estimates		Est	Exp	Est	Exp	Est	Exp	Est	Exp
P	Av	7	33	17	9	17	10	60	21
	Low	1	6	6	4	6	5	41	11
	Up	22	59	35	14	35	14	77	32
S1	Av	0	24	0	28	0	31	0	36
	Low	0	2	0	12	0	15	0	18
	Up	6	47	13	44	5	46	1	53
	D	1	1	1	1	1	1	2	2
S2	Av	0	5	0	15	0	19	0	18
	Low	0	0	0	4	0	2	0	8
	Up	6	12	13	26	5	36	1	29
	D	1	1	2	2	2	2	2	2
S3	Av	0	1	0	3	0	6	0	4
	Low	0	0	0	1	0	2	0	1
	Up	6	3	13	5	5	10	1	6
	D	3	3	3	3	3	3	3	3
S4	Av	0	0	0	0	0	0	0	1
	Low	0	0	0	0	0	0	0	0
	Up	6	0	13	0	5	1	1	2
	D	3	3	3	3	3	3	3	3
S0		100	69	100	54	100	44	100	42

Same legend as above table

Figure 25. Comparison of the risk estimates for the phase "corridors"



	H37	H38	H45	H47	H49
Overall risk (%)	**Energised prod in C**	**Inadequate handling in C**	**Sudden noise in C**	**Dirty floor in C**	**Too many animals in the flow in C**
Est	0.0492	0.0694	0.0001	0.0008	0.1417
Exp	0.0107	0.0143	0.0111	0.0211	0.0067

C: corridors
Est: on-site estimate
Exp: experts' weighted average
The data used for estimating the risk are illustrated in Table 29 and Table 30

Table 29. Data used for estimating the overall risk related to the hazards analysed in the corridors

H number		H37		H38		H45	
H name		Prod in C		Inadequate handling in C		Sudden noise in C	
Estimate		Est	Exp	Est	Exp	Est	Exp
P	Av	7	15	97	18	2	28
	Low	4	10	83	11	0	13
	Up	9	20	100	25	5	43
S1	Av	0	33	72	28	6	18
	Low	na	15	53	14	0	5
	Up	na	52	87	41	29	32
	D	1	1	1	1	1	1
S2	Av	0	13	3	12	0	7
	Low	0	5	0	2	0	0
	Up	10	20	18	22	16	19
	D	2	2	2	2	2	2
S3	Av	100	0	0	2	0	0
	Low	90	0	0	0	0	0
	Up	100	1	10	4	16	1
	D	3	3	3	3	3	3
S4	Av	0	0	0	0	0	0
	Low	na	0	0	0	0	0
	Up	na	0	10	0	16	0
	D	3	3	3	3	3	3
S0		0	54	24	58	94	75

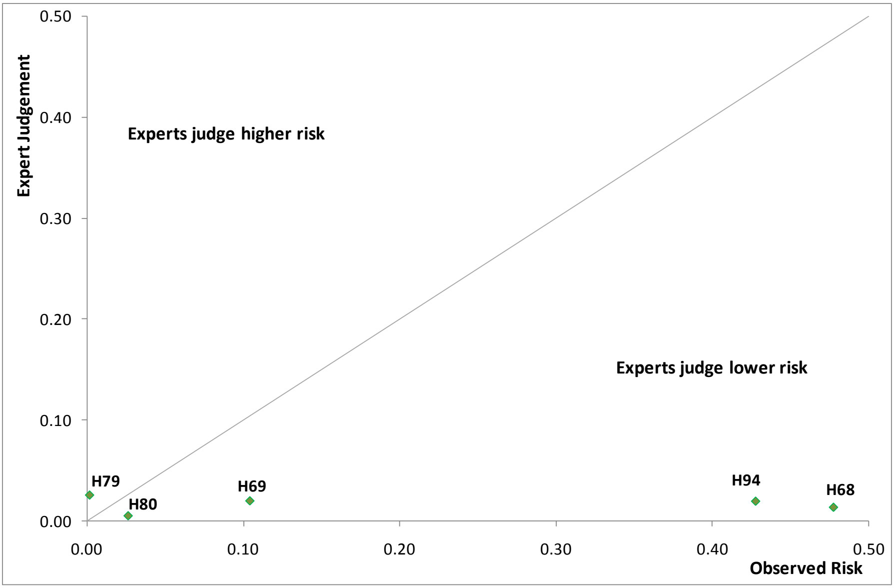
Yellow cells: 0% probability of adverse effect indicated by the experts and thus duration of the adverse effect estimated by the observer
Est: on-site estimate; **Exp**: experts' weighted average; **P**: probability of the hazard; **S1**, **S2**, **S3** and **S4**: probability of the adverse effect by severity level; **Av**: average; **Low** and **Up**: lower and upper limits; **D**: duration of the adverse effect; **S0**: residual category when the sum of the likelihoods of severity 1 to 4 does not reach 100%
In the on-site observations "Energised prod was always associated to severity 3 adverse effects

Table 30. Data used for estimating the overall risk related to the hazards analysed in the corridors (cont')

H number		H47		H49	
H name		Dirty floor in C		Too many animals in the flow in C	
Estimate		Est	Exp	Est	Exp
P	Av	73	20	80	14
	Low	54	9	61	8
	Up	88	31	92	19
S1	Av	0	30	0	26
	Low	0	12	0	7
	Up	1	49	11	45
	D	2	2	1	1
S2	Av	0	12	25	7
	Low	0	7	10	2
	Up	2	18	47	12
	D	2	2	2	2
S3	Av	0	1	13	0
	Low	0	0	3	0
	Up	1	3	32	1
	D	3	3	3	3
S4	Av	0	0	0	0
	Low	0	0	0	0
	Up	1	1	12	1
	D	3	3	3	3
S0		100	56	63	66

Same legend as above table

Figure 26. Comparison of the risk estimates for the phase "entrance of the stunning box"



	H68	H69	H79	H80	H94
Overall risk (%)	**Energised prod in E**	**Inadequate handling technique in E**	**Dirty floor in E**	**Too many animals in the flow in E**	**Overall and continuous noise in E**
Est	0.4776	0.1041	0.0016	0.0016	0.4278
Exp	0.0136	0.0201	0.0258	0.0258	0.0197

E: area close to the entrance of the stunning box
Est: on-site estimate
Exp: experts' weighted average
The data used for estimating the risk are illustrated in Table 31 and Table 32

Table 31. Data used for estimating the overall risk related to the hazards analysed at the entrance of the stunning box

H number		H68		H69		H79	
H name		Energised prod in E		Inadequate handling technique in E		Dirty floor in E	
Estimate		Est	Exp	Est	Exp	Est	Exp
P	Av	64	17	40	20	87	22
	Low	59	13	32	13	69	9
	Up	68	21	46	27	96	35
S1	Av	0	31	10	26	0	34
	Low	na	14	0	11	0	10
	Up	na	48	14	40	1	58
	D	1	1	1	1	2	2
S2	Av	0	13	48	18	0	13
	Low	0	4	40	5	0	7
	Up	1	22	56	32	1	19
	D	2	2	2	2	2	2
S3	Av	100	1	12	2	0	2
	Low	99	0	8	0	0	0
	Up	100	5	19	6	1	4
	D	3	3	3	3	3	3
S4	Av	0	0	0	0	0	0
	Low	na	0	0	0	0	0
	Up	na	0	2	0	1	1
	D	3	3	3	3	3	3
S0		0	54	30	54	100	51

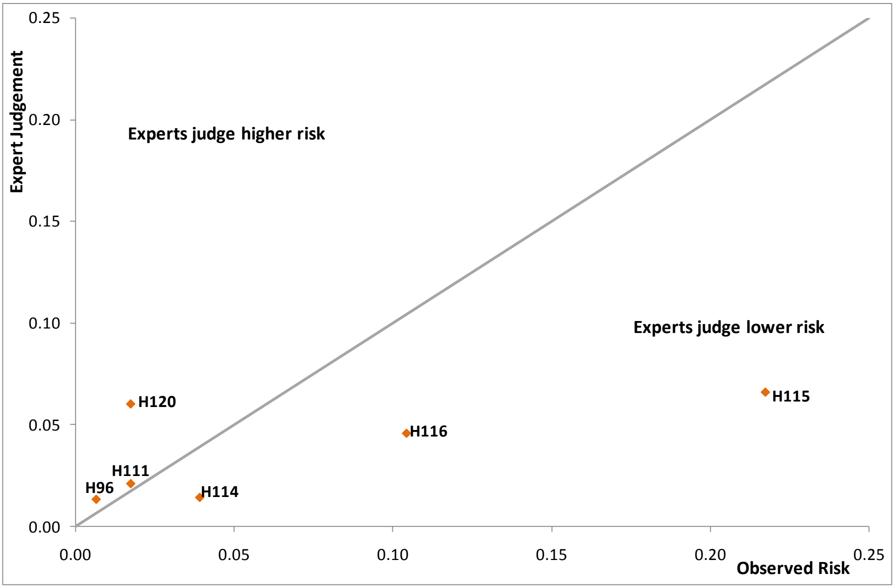
Yellow cells: 0% probability of adverse effect indicated by the experts and thus duration of the adverse effect estimated by the observer
Est: on-site estimate; **Exp**: experts' weighted average; **P**: probability of the hazard; **S1**, **S2**, **S3** and **S4**: probability of the adverse effect by severity level; **Av**: average; **Low** and **Up**: lower and upper limits; **D**: duration of the adverse effect; **S0**: residual category when the sum of the likelihoods of severity 1 to 4 does not reach 100%
In the on-site observations "Energised prod was always associated to severity 3 adverse effects

Table 32. Data used for estimating the overall risk related to the hazards analysed at the entrance of the stunning box (cont')

H number		H80		H94	
H name		Too many animals in the flow in E		Overall and continuous noise in E	
Estimate		Est	Exp	Est	Exp
P	Av	27	12	100	51
	Low	12	5	93	27
	Up	46	18	100	75
S1	Av	0	17	0	27
	Low	0	8	0	14
	Up	1	27	10	40
	D	1	1	1	1
S2	Av	26	7	43	5
	Low	20	0	26	0
	Up	32	14	63	12
	D	2	2	2	2
S3	Av	2	1	57	0
	Low	1	0	37	0
	Up	5	3	75	0
	D	3	3	2	2
S4	Av	0	0	0	0
	Low	0	0	0	0
	Up	1	0	10	0
	D	3	3	2	2
S0		73	75	0	68

Same legend as above table

Figure 27. Comparison of the risk estimates for the phase "stunning and killing"



	H96	H111	H114	H115	H116	H120
Overall risk (%)	**Inadequate handling at S/K**	**Too small stunning box**	**Delayed stunning**	**Wrong shot**	**Delayed 2nd shot**	**No check after stunni ng**
Est	0.0065	0.0174	0.0391	0.2174	0.1043	0.0174
Exp	0.0133	0.0210	0.0143	0.0659	0.0457	0.0601

S/K: stunning and killing
Est: on-site estimate
Exp: experts' weighted average
The data used for estimating the risk are illustrated in Table 33 and Table 34

Table 33. Data used for estimating the overall risk related to the hazards analysed at stunning and killing

H number		H96		H111		H114	
H name		Inadequate handling at S/K		Too small stunning box		Delayed stunning	
Estimate		Est	Exp	Est	Exp	Est	Exp
P	Av	10	13	5	11	51	15
	Low	6	5	2	0	42	9
	Up	18	22	11	28	61	21
S1	Av	0	32	0	36	0	39
	Low	0	13	0	11	0	20
	Up	22	51	39	61	5	59
	D	1	1	2	2	1	1
S2	Av	0	12	100	25	0	17
	Low	0	1	61	11	0	10
	Up	22	23	100	39	5	25
	D	2	2	2	2	2	2
S3	Av	8	4	0	5	10	1
	Low	0	0	0	0	4	0
	Up	39	11	39	15	21	4
	D	3	3	3	3	3	2
S4	Av	0	0	0	0	0	0
	Low	0	0	0	0	0	0
	Up	22	1	39	0	5	0
	D	3	3	3	3	3	3
S0		92	52	0	34	90	42

Yellow cells: 0% probability of adverse effect indicated by the experts and thus duration of the adverse effect estimated by the observer
Est: on-site estimate; **Exp**: experts' weighted average; **P**: probability of the hazard; **S1**, **S2**, **S3** and **S4**: probability of the adverse effect by severity level; **Av**: average; **Low** and **Up**: lower and upper limits; **D**: duration of the adverse effect; **S0**: residual category when the sum of the likelihoods of severity 1 to 4 does not reach 100%

Table 34. Data used for estimating the overall risk related to the hazards analysed at stunning and killing (cont')

H number		H115		H116		H120	
H name		Wrong shot		Delayed second shot		No check after stunning	
Estimate		Est	Exp	Est	Exp	Est	Exp
P	Av	22	14	10	10	52	20
	Low	15	9	6	6	43	14
	Up	30	18	18	15	62	27
S1	Av	0	7	0	12	0	6
	Low	0	0	0	0	0	0
	Up	21	16	22	30	5	17
	D	2	2	1	1	2	2
S2	Av	0	16	0	10	0	7
	Low	0	5	0	0	0	0
	Up	21	27	22	21	5	17
	D	3	3	2	2	3	3
S3	Av	0	24	0	14	0	19
	Low	0	5	0	0	0	3
	Up	21	43	22	33	5	35
	D	3	3	3	3	3	3
S4	Av	100	22	100	29	3	11
	Low	79	3	78	0	0	1
	Up	100	41	100	58	12	21
	D	3	3	3	3	3	3
S0		0	31	0	34	97	57

Same legend as above table

5. PART V. Discussion

5.1 Results of exposure assessment

Consistent results on exposure assessment (14 out of 42 hazards)

In the comparison of the results of on-site and expert opinion-based exposure assessment, adhering results[35] occurred more rarely (14 out of 42 hazards).

Adhering results showed how the random choice of the experts gave a chance that the assumed agreement between the experts (i.e. the weighted average) was not in conflict with the on-site estimated values.

Adhering results occurred mainly for hazards that in the reality seem to have a rather low probability of occurrence (e.g. the use of the energised prod for unloading the animals and for driving them all along the unloading bay; too fast animal flow in the corridors; the presence of trucks with slippery or steep ramps; etc). The average probability of occurrence of these hazards was estimated as equal to 7% in the on-site observations and 13% by the experts.

Although presenting adhering results on exposure assessment, for the hazard "Sudden noise due to other than the truck at unloading" (H11) the variability between the experts was relatively higher (Di=0.41) than in all other cases in this category of adhering results[36]. This may show certain difficulty in observing and/or quantifying this hazard, which was actually one of the hazards "difficult to measure" also during the on-site observations.

For the hazards "Stunning box gates' opening size too small" and "Too big stunning box", not only the variability between the experts but also the

---

35 Adhering results corresponded to: confidence intervals of the on-site observations overlapping with the confidence intervals of the weighted averages of the experts; and homogeneous experts' replies (Di<0.6).

36 For the majority of the hazards in the category adhering results, variability between the experts was ≤0.27.

uncertainty was relatively higher than in all other cases. This may show a real lack of knowledge within the expert group on the occurrence of this specific hazard (which often depends on the size of the animals).

Inconsistent results on exposure assessment (24 out of 42 hazards)

For most of the hazards the results of on-site assessed exposure were inconsistent with the experts' estimates[37] and this indicated different situations.

Inconsistent results due to a disagreement between the observer and the experts in observing and/or quantifying hazards difficult to measure. Often inconsistent results occurred for hazards that were difficult to measure objectively.

For the risk assessment based on data collected in the abattoir these hazards were defined on the basis of the knowledge of the observer and the experience gained during the preliminary observations. Inconsistent results may indicate that the experts had a different perception of such hazards.

However, often for these hazards the variability between the experts was relatively higher compared to the other situations in this category of inconsistent results[38], showing a difficulty to measure the hazards in the experts' group. These hazards in the questionnaire to the experts were defined as in the risk assessments for animal welfare under appraisal.

This indicated the need for more precise definitions of the hazards in order to make them objectively measurable.

For instance the hazard "Dirty floor" in the three areas of pre-slaughter handling (H18, H47 and H79) was defined in the on-site observations as "Dirtiness (e.g. excess of manure that makes the floors slippery) such as it clearly induced in the preliminary observations at least one animal to show at least one of the indicators of poor welfare indentified in this study". In

[37] Inconsistent results were found when the confidence intervals overlapped and there was no variability between the experts, i.e. $Di < 0.6$.

[38] For the majority of the hazards in the category inconsistent results, variability between the experts was ≤ 0.29.

the on-site data collections this hazard was observed far more frequently than what was estimated by the experts, indicating that the observer may have had a different perception of dirty floors. However, the relatively high variability of the experts replies showed an overall difficulty in measuring this hazard (H18: Di=0.36; H47: Di=0.39; H79: Di=0.38).

Another example is the hazard “Sudden change of light intensity” during pre-slaughter handling (H10, H44, and H75). This hazard was never observed during the on-site observations whereas the experts assigned to it certain probability, but with relatively higher uncertainty and variability (H10: Di=0.41%; H44: Di=0.34; H75: Di=0.40).

The hazard “Too fast animal flow”[39] was never observed in the on-site observations, whereas the experts estimated for it certain probability of occurrence, with homogenous replies. This showed a different interpretation of the hazard between the observer and the experts. In fact, the observer tent to interpret any fast movements of the animals as a normal behaviour, as if the animals would speed up to fill-in the empty spaces in the corridors. This was also the interpretation of one of the experts who replied that “Too fast animal flow” should not be included in the list of hazards to beef cattle at slaughter.

Hazards due to the management. Often hazards due to the management were observed in the on-site observations far more frequently than what was perceived by the experts. This may be due to the fact that experts may have been biased as inspectors in the slaughter plants, whereas the observer was not recognised as an auditor or inspector assessing the work of the operators.

For example the results showed that during the on-site observations the slaughter men inadequately handled the animals considerably more frequently than what was estimated by the experts (i.e. hazard “Inadequate handling technique” during pre-slaughter handling).

The hazard “Energised prod at the entrance of the stunning box” during the on-site observations was observed far more frequently than how it was

[39] Both at unloading and in the area close to the entrance of the stunning box.

perceived by the experts. The entrance of the stunning box is a very delicate phase of the process, where normally the animals are reluctant to proceed and may be scared by the fact that they cannot see what happens inside the box. Although it is normally strongly recommended not to use the prod to force the animals into the box, often such tool may result useful to speed up the process.

For some hazards the results were simply inconsistent, showing that the on-site observations did not confirm the assumed agreement between the experts.

For instance the hazard "Faulty operating non-return gates" at unloading and in the corridors were never observed during the on-site observations and estimated with certain (although low) probability of occurrence by the experts (P=11%).

For the hazard "Gates' opening size too small" at unloading (H31) and in the corridors (H62) and for the hazard "Too narrow corridors" in the passageways (H59) and close to the entrance of the stunning box (H90), the results of the two exposure assessments were inconsistent (the hazards were never observed in the on-site observations whereas the experts estimated certain probability of occurrence). It may be argued that this was due to the fact that the hazards are slaughter plant-related and that were not present in the abattoir where the observations took place. However, during the preliminary observations of the three plants, it was noted that these hazards depend on the size of the animals and that particularly big (or long horned) animals could actually get squashed/jammed when passing the gates or in the corridors. In addition, although the results were inconsistent, the rather low P assigned by the experts (P=14% for H31; P=12% for H62; P=9% for H59; and P=12% for H90) was in line with the on-site observations.

High variability between the experts' replies on exposure assessment (4 out of 42 hazards)

For some hazards (4 out of 42) the variability between the experts' replies was much higher than in all other cases ($Di \geq 0.6$). This indicated different situations.

High variability between the experts due to a difficulty in observing some hazards and/or assessing their likelihood of occurrence
High variability between the experts occurred for hazards difficult to measure objectively, such as “Sudden noise due to the truck” (H9); “Sudden noise at the entrance of the stunning box” (H76); and “Overall and continuous noise originating in the stunning box” (H94).
Since the beginning of the on-site observations these hazards were recognised as difficult to measure objectively and had to be defined on the basis on the knowledge and experience of the observer.
The experts may have found it difficult to assess the actual occurrence of noises in the slaughter plants.
This again indicated the need for clear, measurable definitions of the hazards. Further, it outlined that in some cases it may be necessary to rely on on-site collected data for assessing hazard exposure.

High variability between the experts due to lack of expert knowledge
For the hazard “Open side barriers or lower than 1.5m on the truck ramp - Animal side vision not blocked” (H13) the high variability between the experts (Di=0.72) showed the actual lack of knowledge on the occurrence of this hazard (as the hazard was clearly described and not prone to misinterpretation).

5.2 Results of likelihood of the adverse effects by severity levels

High variability between the experts’ responses on likelihood of the adverse effects
The high variability between the experts was associated to different situations.
Difficulty in quantifying the severity of the adverse effects, especially for events of limited or moderate severity. For most of the hazards (23 out of 24 hazards) experts’ variability was relatively high (Di≥0.5) for probability of occurrence of severity 1 and/or severity 2 adverse effects (and low for severity 3 and 4 adverse effects).

In the survey the experts were asked to indicate the most likely probability of occurrence of the adverse effects by severity level (scale 1-4), when the animals are exposed to each hazard.

This question was based on the fact (supported by the available evidence and given as unquestionable in the questionnaire) that the adverse welfare effects occurring to beef cattle at slaughter are mainly represented by distress, pain, frustration, fear, anxiety, changes in adrenal or behavioural reactions, with severity levels varying from limited to critical (1-4 scale). To support the understanding of the 4 severity levels, some examples of indicators of poor welfare were given, without being prescriptive in linking the indicators to the various severity levels (section 2.7.3)[40].

In view of the above, the high variability between the experts' estimates associated to the probability of occurrence of Severity 1 and 2 adverse effects may be due to the fact that the experts had a thorough knowledge of what happens to the animals when exposed to the hazards listed in the questionnaire, but found it difficult to quantify the severity of such "events" (i.e. adverse effects).

For instance, it may have occurred that the experts knew that if exposed to inadequate handling the animals may move backwards, run or turn around; however, they did not know exactly whether to associate those events to the category "limited" severity (1) "moderate" severity (2) or maybe even "no severity" (i.e. the event was not interpreted as indicating a poor welfare status).

This indicated the need for a common understating of the welfare outcomes and how to quantify them.

An alternative to the above interpretation would be that the experts did not know what exactly happens to the animals when exposed to a determined hazard (i.e. they did not know how the animals would react to an operator handling them improperly). However, this seemed very unlikely

[40] On the contrary, in the on-site observations the welfare consequences of the exposure to the hazards were estimated using measurable indicators(s) of adverse effect(s) detected during the on-site observations (e.g. "Moving backwards, "Turning around", Slipping", "Falling", etc), which were associated to specific Severity levels (1-4 scale).

considering the expertise of the group.

Disagreement within the experts' group in quantifying the severity of the adverse effects. For 3 hazards out of 24, the experts gave non-homogenous replies on the probability of occurrence of severity 3 adverse effects (H96 "Inadequate handling technique at stunning" and H120 "No check of unconsciousness after stunning"), and severity 3 and 4 adverse effects (H116 "Delayed second shot"). This may be due to a disagreement within the experts' group in interpreting the adverse effects.

For example for the hazard "Delayed second shot" (H116) (defined as "when the first shot is unsuccessful, animal not immediately re-stunned") the variability between the experts' estimates was very high for all severity levels and in particular for severity 4 (Di=0.95).

In the on-site observations, every time this hazard was observed the consequent adverse effect was estimated as severity 4 (severity 4 P=100%), i.e. e*xtrem*e pain, distress, fear or anxiety, that if persisting would be life-threatening.

Although the definition of severity 4 was the most objective (due to the component "incompatible with life"), apparently *some experts agreed* on this estimate, whereas some others *fully disagreed*. Now, since it is difficult to argue that an animal exposed to an unsuccessful stun would *not* suffer from an adverse effect that "if persisting would be life-threatening", the disagreement of some experts (and the consequent variability in the replies) was interpreted as if some experts were reluctant to stick to the definition of severity provided in the questionnaire. In fact since the animals are close to be killed, some experts may have considered the situation as not extremely severe, independently of the definition given in the questionnaire.

For the hazard "Wrong shot" (H115), the experts' variability was high for probability of occurrence of severity 3 and 4 adverse effects (Di3=0.64 and Di4=0.67). Again these results showed an overall disagreement in assessing the consequences of this hazard likely to be due to the same reasons as for H116. In fact some experts agreed on the on-site estimates (i.e. a wrong shot has a very high probability of causing extremely severe consequences in the animals), whereas some were in complete disagreement.

Difficulty in interpreting the questions in the questionnaire. For the hazard H120 (No check of unconsciousness after stunning) the variability between the experts related to Severity 3 adverse effects was high. This may be due to a difficulty in assessing the situation, as the presence of acute consequences will depend upon the effectiveness of the stun (i.e. if the stun is adequate the absence of check would not hamper the welfare of the animal).

Inconsistent results on likelihood of the adverse effects

In some cases the experts' estimates showed a considerable difference from the empirical results, with no variability between the replies. The most likely reason for this was *a disagreement between the experts and the observer in quantifying the severity of the adverse effects*.

An example was the case of severity 3 adverse effects associated to the hazard energised prod (in all phases of pre-slaughter handling, i.e. at unloading, in the corridors and at the entrance of the stunning box). On the basis of the literature and in accordance within the research group, during the on-site observations the prod was always assigned to a severity level equal to 3 (severe pain). However, this was not agreed by the majority of the experts, who mainly associated to this hazard a severity level equal to 1, with rather low variability in their replies.

Another example was represented by H94 (Overall and continuous noise in the area close to the entrance of the stunning box), for which the probability of occurrence of severity 2 and 3 adverse effects was estimated as much higher in the on-site observations than by the experts.

5.3 Results of risk characterisation

The results of the risk estimates did not contradict the expected outcomes of exposure assessment combined with the estimated likelihood of the adverse effects by severity levels.

For instance, for the hazard Energised prod the results of likelihood of adverse effects showed a much higher estimated probability of occurrence of severity 3 adverse effects in the on-site observations. In addition, on-site exposure assessment for this hazard indicated a probability of occurrence at unloading and in the corridors comparable to the one assessed by the experts and a considerably higher probability at the entrance of the stunning box.

Thus unsurprisingly the overall risk due to Energised prod based on data collected in abattoir resulted much higher than the one assessed by the experts.

Another example is represented by the hazard Dirty floor, for which in the on-site observations adverse effects were almost never observed. As opposite, the experts estimated certain probability of occurrence for all 4 severity levels' adverse effects. So, although the hazard was detected in the on-site observations more frequently than what was perceived by the experts, the overall risk estimated by the experts was of course much higher.

Similar considerations may be done for all hazards assessed.

For some hazards the on-site estimated risk was equal to 0% (e.g. H14"Slippery truck ramp"; H15 "Steep truck ramp"; and H18 "Dirty floor at unloading"), whereas this never happened in the experts' estimates. This was due to the fact that for these hazards no adverse effects were observed in the on-site data collections. On the other hand, for no hazards the experts indicated a 0% probability of occurrence of severity 1 and 2 adverse effects.

6. PART VI. Conclusions

6.1 Introduction

While risk assessment is widely used in support of decision making in many areas of food and feed safety (e.g. veterinary epidemiology, toxicology, eco-toxicology), in the animal welfare field this method is relatively new. The first attempt to use formal risk assessment methodology for animal welfare was made in 2006 (EFSA, 2006a) and currently there are no standardised guidelines for animal welfare risk assessment. Furthermore, very little research has been conducted to assess the reliability of the existing methods, which mostly rely on qualitative data and are based on expert opinion.

This study aimed at assessing the scientific robustness of existing risk assessments for animal welfare and identifying methodological flaws of these methods.

A thorough analysis of the available animal welfare risk assessments was carried out and two risk assessments for beef cattle at slaughter (in northern Italy) were performed and compared, one based on empirical data and one based on expert opinion. The two new risk assessments were structured to be as similar as possible to the animal welfare risk assessments under appraisal.

For the risk assessment based on empirical data a novel method for performing on-site exposure assessment and adverse effect characterisation was developed.

The risk assessment based on expert opinion was carried out using information gathered via a tailored questionnaire distributed to a group of 11 experts.

A total of 56 hazards were identified via a literature review. Hazard identification in the two risk assessments was performed by asking the experts to assess the relevance of the reviewed hazards and by checking them via a series of preliminary observations in abattoirs. The outcomes of the questionnaire and the on-site observations showed that 14 hazards identified via the literature search were not relevant to the target population and the scenario of the assessment. Thus the results of exposure assessment were compared for 42 hazards. As some of them were never observed in the on-site data collections (or the number of animals exposed was < 5), adverse effect characterisation and final risk were estimated and compared for 24 hazards.

The analysis of the risk assessment method performed in this study, the variability between the experts' replies and the discordance of the results highlighted some inherent flaws and requirements of existing risk assessments for animal welfare.

6.2 Method for on-site data collection

To perform the risk assessment based on empirical data, a tailored method was developed for estimating hazard exposure and assessing the likelihood and magnitude of the adverse welfare effects, using on-site collected data.

For on-site hazard exposure assessment

While analysing the available risk assessments to develop the on-site data collection system, it was clear that the hazards qualitatively defined in the literature and in existing risk assessments had to be further defined (or re-defined) to make them measurable and quantifiable. A comprehensive list of clearly described and measurable hazards to beef cattle at slaughter was produced.

To reproduce the structure of the existing risk assessments for animal welfare under appraisal, interactions between hazards[41] and different hazards' intensities and durations were not considered. However, the on-site observations highlighted that interacting hazards often occur and that not including them may reduce the preciseness of adverse effect characterisation and the final risk estimate.

[41] Interacting hazards are hazards occurring at the same time and interacting among each others.

Indeed if two hazards that amplify each other impact are included separately in the analysis, this implies an impact that is too low than it could be in reality, which results in an underestimation of risk. Failure to take into account this extra adverse effect from interaction can lead to either underestimation or overestimation of the risk, depending on the direction of the extra effect. This is what actually happened in the risk assessment based on empirical data, where the adverse effects due to interacting hazards were randomly allocated to the hazards, leading to an unavoidable underestimation and overestimation of the effects of such interacting hazards.

Moreover if two interacting hazards are separately included in the analysis, but both depend on each other in a way that the occurrence of one hazard automatically calls for the occurrence of the other, the separate inclusion of both hazards could lead to an overestimation of risk. In contrast, if the presence of the first hazard makes the second hazard less common, it will lead to an underestimation of the risk.

The on-site observations also showed that the adverse effects depend on the intensity and duration of the hazards (i.e. hazard magnitude) on a case by case basis.

For on-site likelihood of the adverse effects

A novel, specific method for assessing the welfare of beef cattle at slaughter was developed and implemented.

The method was formulated integrating the scoring systems provided in the Welfare Quality® project (WQ®, 2009) and proposed by Grandin (2010) with the information gathered during the preliminary observations.

A comprehensive list of indicators of poor welfare of beef cattle at slaughter was produced. The indicators were further defined to make them objectively measurable and distinguishable from the others. Where necessary precise start and end times (or total duration) of the indicators were described.

Moreover, the indicators of poor welfare were classified by severity levels and this proved to be very useful to assess the severity of the various adverse effects under field conditions and in absence of video recording.

The list of indicators may be further validated and agreed to ensure inter-observer reliability and short- to long-term intra-plant variability (consistency) of the indicators. It could also be tested on (and integrated for) other subpopulations and systems (e.g. young calves, adult cows, or at lairage). More indicators may be identified to increase the sensitivity of the method especially for assessing the duration of the adverse effects and for enhancing the measurement of: low severity levels; “additional” variations in welfare (e.g. the stress, fear, etc “additional” to the indicators already identified); averse effects difficult to measure (e.g. the stress consequent to the prod); and cumulative-severity indicators[42]. The method could then be tested in other animal systems (e.g. transport and on farm).

6.3 Method for eliciting expert opinion

To perform the risk assessment based on expert opinion, a specific questionnaire was produced and distributed to a group of 11 experts.
Differently from the existing animal welfare risk assessments, where expert opinion is elicited within groups of experts who reach a *consensus* on the various estimates (with no explanation on how the consensus is reached), in this study the experts answered the questionnaire *independently*.
The consensus approach seems to have the advantage of providing a single value for each score and gives the impression of absolute homogeneity on the value within the group. However, as experts cannot be wholly certain of a value, there has to be some incorporation of variability to reflect the true level of variety of opinion within the group, and more importantly, in the animal species being assessed. In addition, the experts in a group cannot be said to be independent as their knowledge source is shared and there is the risk that a “dominant” expert imposes his/her opinion on the others.
Analysis of the variability between the experts replies
The approach proposed in this study permitted the analysis of the results

[42] The choice on whether to increase or not the level of severity was made on a case by case basis and in general considering the time occurring between the various events of the same indicator (section 2.7.2.2).

using an additional element, i.e. the coefficient of dispersion between the experts, which expresses the standard deviation in relation to the deviation of fully random answers and in turn showed the degree of homogeneity (coherent thinking/knowledge) or variability (indicating that an element would interfere with the replies) between the experts' responses.

This value proved to be essential to highlight aspects that in existing animal welfare risk assessments are often overlooked. Indeed high variability between the experts' replies indicated different situations:

a) Difficulty in observing and/or quantifying the risk assessment parameters (i.e. hazards and adverse effects). This highlighted the need for clearly defined and measurable hazards and adverse effects. In particular when the likelihood and the magnitude of the adverse effects are modelled using expert opinion, the description of the various severity classes should be very well defined and not open to interpretation. Thus indicators of poor welfare (classified by severity classes) should be used.
b) Lack of expert knowledge. This proved how the estimate of variability may support the risk assessment process by highlighting areas where empirical data may be more informative than expert opinion.
c) Disagreement on the definitions given in the questionnaire or difficulty in interpreting the questions. This indicated the need for an explanation of/agreement on the questionnaire, before answering it. While the results of the study showed that different experts may interpret the questions differently, it must be specified that a consensus opinion does not necessarily prevent from misinterpretations, which may simply be "embedded" in the agreed score.

Uncertainly analysis

Differently from the risk assessments methods under appraisal (where uncertainty is defined on the basis of the availability and reliability of scientific evidence only), in this study the definition of uncertainty was extended by including the concept of personal experience in the topic.

It may be argued to which extent the experts scored their uncertainty on the basis of the available evidence and to which extent using their personal experience. On the other hand, it must be noted that in existing risk

assessments for animal welfare the evidence underlying low or medium uncertainty levels is almost never referenced, which may lead to the same doubt.
In addition, few scientists world-wide are active in studying welfare aspects of stunning and killing and relatively few publications are available on this topic. Thus it is reasonable to conclude that in the risk assessment based on expert opinion performed in this study, whenever the experts indicated a low uncertainty, it was likely to be based on their knowledge and expertise. However, this may not be the case in other areas of animal welfare (e.g. on farm, where, depending on the species, more data may be available).
Thus it is concluded that it is paramount whenever low or medium uncertainties are indicated that the underlying literature is always reported.

6.4 Comparison of the two risk assessments

The risk assessment based on empirical data and the risk assessment based on expert opinion were compared and rarely produced comparable results.
Often the results were inconsistent (i.e. low variability between the experts and confidence intervals not overlapping) due to a disagreement between the observer and the experts in observing and/or quantifying the risk assessment parameters (i.e. the hazards and the adverse effects).
This occurred for example for hazards defined qualitatively in the literature and difficult to measure objectively (e.g. "Dirty floor" or "Too fast animal flow").
In likelihood of the adverse effects an example indicating a disagreement between the observer and the experts was the hazard energised prod, for which the experts clearly did not support the choice to always associate to this hazard to severity 3 adverse effects.
Inconsistent results often occurred for hazards due to the management in the slaughter plants, which were detected in the on-site observations far more frequently than what was perceived by the experts. This showed the importance of the observer effect (the experts were inspectors in the abattoirs) and the fact that in some cases empirical data may be more

informative than expert opinion or that the background and expertise of the expert group should be thoroughly considered when using expert opinion.

6.5 Data input in the risk assessment and documentation of the data input process

The detailed analysis of the risk assessment method performed in this study highlighted that, independently of the data input in the risk assessment (i.e. empirical data or expert opinion), attention should be paid to the method applied (i.e. for reviewing the literature or gathering expert opinion), which should be chosen in line with the best available methodologies.

A thorough documentation of the process is also paramount in both cases, to ensure greater transparency and reproducibility of the risk assessment.

6.5.1 Use of empirical data

When empirical data are available and used, the extent to which a risk assessment is useful in a specific situation is determined by the robustness of the method used to identify, select, appraise, and synthesise the available evidence. Ideally, when feasible and justified, the risk assessment parameters should be estimated using systematic reviews. A systematic review is an overview of existing evidence pertinent to a clearly formulated question, which uses pre-specified and standardised methods to identify and critically appraise relevant research, and to collect, report and analyse data from the studies that are included in the review (EFSA, 2010d).[43]

However, systematic reviews may not be always feasible or practical as resource-intensive and limited to addressing questions that are sufficiently well-structured to be answered in a primary study[44]. Thus it is suggested to consider systematic review in the phase when the conceptual model for the

[43] Statistical methods to synthesise the results of the included studies (meta-analysis) may or may not be used in the process.

[44] EFSA (2010d) illustrates a useful framework for assessing the suitability of questions to systematic review and some aspects which may serve as a check-list to assess whether a systematic review is needed when using risk assessment.

risk assessment is built and the required input information is identified.

Independently of the method applied for reviewing the literature (i.e. systematic review or classic review), the review process should be clearly documented. This implies illustrating the search strategies used (i.e. combination of search terms and Boolean[45] operators); the sources of literature searched (e.g. bibliographic databases, scientific journals tables of contents, specialised websites, etc); the criteria (if any) applied to select the studies for inclusion in the reviews; the method (if any) used for assessing the reliability of the studies; and the approach to synthesising the findings of the included studies.

This study noted that these aspects are rarely if ever reported in existing risk assessments for animal welfare, which present low transparency and reproducibility.

6.5.2 Use of expert opinion

If empirical data lack and expert opinion is used, attention should be paid to the method applied.

Ideally gathering a large group of experts together and asking each expert to provide scores independently of the others would solve some problems of variability (as they will tend to converge on an average estimate, with a narrow range, as the number of experts increases due to the classic principles of the central limit theorem).

However, an introductory phase should be foreseen where all experts reach a common understanding and agreement on the hazards and adverse effects to be assessed as well as on how they are measured.

Accordingly a clear definition of the hazards and animal welfare adverse effects is paramount.

The Delphi Method may be considered as a further alternative way to elicit expert opinion. Experts have to score in two or more rounds. At the end of

[45] Boolean operators are words used to combine terms or concepts when conducting electronic bibliographic searches. Examples include “AND” (used to narrow a search), “OR” (used to broaden a search) and “NOT” (used to exclude terms from a search).

each round a summary of the experts scoring is proved and the individual reasoning. Experts have to check their answers in the light of the answers of other experts. It is expected that the group will converge towards the right answer. The process is stopped after a predefined stop criterion (Spoolder, 2010; Rowe and Wright, 1999).

When expert opinion is used, criteria for choosing the experts should also be considered. For instance in this study some hazards due to the management were observed in the on-site observations far more frequently than what was perceived by the experts and this may be due to the fact that experts may have been biased as inspectors in the slaughter plants.

Transparency is also paramount. It needs to be clearly documented how the experts were chosen, whether there was a conflict of interest, how many experts were asked, how many scored in the end, which technique was used to obtain expert opinion and how expert opinion was scored.

This study outlined that the method for reaching the consensus opinion between the experts and for handling potential disagreements is never reported in existing animal welfare risk assessments.

6.6 Conclusions

This research project reported a unique validation study of existing methods for animal welfare risk assessment.

The detailed analysis of the risk assessment method performed in this study highlighted that, independently of the data input in the risk model (i.e. empirical data or expert opinion), risk assessment for animal welfare requires a more detailed and measurable description of the hazards to animal welfare as well as a thorough consideration of interactions between hazards and of different hazards' intensities and durations.

Accordingly, an approach to defining the hazards to animal welfare in a measurable and quantifiable way was proposed, which demonstrated to be very useful for assessing on-site hazard exposure. The most suitable way to deal with interacting hazards and different hazards' magnitudes seemed to consider them at the beginning of the assessment, when indentifying and

describing the hazards related to the scenario of the assessment and the target population.
Moreover, the study showed that one of the most crucial parts of risk assessment for animal welfare (both based on empirical data and expert opinion) is the description of the adverse effects consequent to the hazards and the quantification of their severity and duration.
Thus, a novel method for assessing animal welfare was developed which proved very valuable to produce reliable data on adverse effect likelihood and magnitude. The method was based on a precise definition of the welfare outcomes and a differentiation between adverse welfare effects and measurable indicators of adverse effects. The latter were associated to different severity classes defined qualitatively on the basis of the intensity of the behavioural responses (demonstrating pain, frustration, fear, anxiety, changes in adrenal or behavioural reactions) of the animals.
This approach to hazard description and welfare outcome definition and assessment is recommended for enhancing empirical research on animal welfare, which will improve the reliability of the data used in risk assessment models.
When expert opinion is used, this approach is recommended to assist a common level of understanding of the various parameters necessary for the risk assessment and produce a more robust estimate of hazard exposure and adverse effect likelihood and magnitude. Indeed to employ advance modelling techniques, common understanding and agreement by welfare scientists on the type of parameters are paramount.
This study also proposed an alternative method for eliciting expert opinion based on independent scoring of the risk assessment parameters. The approach showed very useful implications for identifying sources of uncertainties that are normally overlooked in existing risk assessments for animal welfare, such as difficulties in assessing the risk assessment parameters, disagreements between the experts or lack of expert knowledge.
Finally this study highlighted that, independently of the data input in the risk assessment (i.e. empirical data or expert opinion), the method for either reviewing the literature or gathering expert opinion should be chosen in

light of the best available practices. Suggestions for improvement included the use of systematic reviews for reviewing empirical data and the Delphi method for gathering expert opinion. The process and any decisions taken should be documented to ensure greater transparency and reproducibility.

References

Algers B, Anil H, Blokhuis H, Fuchs K, Hultgren J, Lambooij B, Nunes T, Paulsen P, Smulders F, 2009. Project to develop Animal Welfare Risk Assessment Guidelines on Stunning and Killing. Technical report submitted to EFSA (European Food Safety Authority), developed on the proposal CFP/EFSA/AHAW/2007/01. Accepted for publication on 15/10/2009. Available at: http://www.efsa.europa.eu/cs/BlobServer/External_Rep/cfp_ahaw_2007_01_final.pdf?ssbinary=true.

Asante-Duah K, 2002. Public health risk assessment for human exposure to chemicals. In: B.J. Alloway and J.T. Trevors, Editors, Environmental Pollution vol. 6, Kluwer Academic Publishers, Dordrecht, Netherland K.

Bracke MBM, Spruit BM, Metz JHM, 1999. Overall welfare assessment reviewed, Part 1: Is it possible, Netherland Journal of Agricultural Science 47, pp. 279-91.

Bracke MBM, Edwards SA, Engel B, Buist WG, Algers B, 2008. Expert opinion as 'validation' of risk assessment applied to calf welfare. Acta Veterinaria Scandinavica, 50:29. doi:10.1186/1751-0147-50-29.

CAC (Codex Alimentarius Commission), 1999. Principles and Guidelines for the Conduct of Microbiological Risk Assessment, CAC/GL-30, 7pp.

Dalla Villa P, Marahrens M, Velarde Calvo A, Di Nardo A, Kleinschmidt N, Fuentes Alvarez C, Truar A, Di Fede E, Luis Otero J, Müller-Graf C, 2009. Project to develop Animal Welfare Risk Assessment Guidelines on Transport. (Question No EFSA-Q-2008-04997) Project ID (CFP/EFSA/AHAW/2008/02). Accepted for Publication on 23 November 2009. Available at: http://www.efsa.europa.eu/en/supporting/pub/21e.htm.

European Food Safety Authority (EFSA), 2004. Opinion of the Scientific Panel on Animal Health and Welfare on a request from the Commission related to welfare aspects of the main systems of stunning and killing the main commercial species of animals, The EFSA Journal (2004), 45, 1-29. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2006a. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on The risks of poor welfare in intensive calf farming systems. The EFSA Journal (2006) 366, 1-36. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2006b. Scientific Opinion on "Animal health and welfare risks associated with the import of wild birds other than poultry into the European Union". The EFSA Journal (2006) 410, 1-55. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2006c. The welfare aspects of the main systems of stunning and killing applied to commercially farmed deer, goats, rabbits, ostriches, ducks, geese and quail (Question N° EFSA-Q-2005-005). Annex to the EFSA Journal (2006) 326, 1-18. Report, 71 pp. 79. Available at: www.efsa.europa.eu

EFSA (European Food Safety Authority), 2007a. Scientific Opinion of the Panel on Animal Welfare on a request from the Commission on animal health and welfare in fattening pigs in relation to housing and husbandry. The EFSA Journal (2007) 564, 1-14. Available at: www.efsa.europa.eu.

EFSA (European Food Safety Authority), 2007b. Scientific Opinion of the Panel on Animal Health and Welfare on a request from Commission on the risks associated with tail biting in pigs and possible means to reduce the need for tail docking considering the different housing and husbandry systems. The EFSA Journal (2007) 611, 1-13. Available at: www.efsa.europa.eu.

EFSA (European Food Safety Authority), 2007c. Scientific Opinion of the Panel on Animal Welfare on a request from the Commission on animal health and welfare aspects of different housing and husbandry systems for adult breeding boars, pregnant, farrowing sows and unweaned piglets. The EFSA Journal (2007) 572, 1-13. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2007d. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the Animal Welfare aspects of the killing and skinning of seals. The EFSA Journal 610, 1-123. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2008. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on Animal welfare aspects of husbandry systems for farmed Atlantic salmon. The EFSA Journal (2008)736, 1-31. Available at: www.efsa.europa.eu.

EFSA (European Food Safety Authority), 2009a. Panel on Animal Health and Welfare (AHAW); Guidance on Good Practice in Conducting Scientific Assessments in Animal Health using Modelling. EFSA Journal 2009; 7(12):1419. [38 pp.]. doi:10.2093/j.efsa.2009.1419. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009b. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on welfare aspect of the main systems of stunning and killing of farmed Atlantic salmon. The EFSA Journal (2009) 1012, 1-77. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009c. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on Species-specific welfare aspects of the main systems of stunning and killing of farmed rainbow trout. The EFSA Journal (2009) 1013, 1-55. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009d. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on welfare aspect of the main systems of stunning and killing of farmed eel (Anguilla anguilla). The EFSA Journal (2009) 1014, 1-42. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009e. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on welfare aspect of the main systems of stunning and killing of farmed seabass and seabream. The EFSA Journal (2009) 1010, 1-52. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009f . Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on Species-specific welfare aspects of the main systems of stunning and killing of farmed carp. The EFSA Journal (2009) 1013, 1-37. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009g. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on welfare aspect of the main systems of stunning and killing of farmed turbot. The EFSA Journal (2009) 1073, 1-34. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009h. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on the species-specific welfare aspects of the main systems of stunning and killing of farmed tuna. The EFSA Journal (2009) 1072, 1-53. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009j. Scientific Opinion of the Panel on Animal Health and Welfare on a request from European Commission on the overall effects of farming systems on dairy cow welfare and disease. The EFSA Journal (2009) 1143, 1-38. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009k. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on leg and locomotion problems in dairy cows. The EFSA Journal (2009) 1142, 1-57. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009l. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on udder problems in dairy cows. The EFSA Journal (2009) 1141, 1-60. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009m. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on metabolic and reproductive problems in dairy cows. The EFSA Journal (2009) 1140, 1-75. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2009n. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on behaviour, fear and pain problems in dairy cows. The EFSA Journal (2009) 1139, 1-68. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2010a. EFSA Panel on Animal Health and Welfare (AHAW): Scientific Opinion on welfare aspects of the management and housing of the grand-parent and parent stocks raised and kept for breeding purposes. EFSA Journal 2010; 8(7):1667. [81 pp.]. doi:10.2903/j.efsa.2010.1667. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2010b. EFSA Panel on Animal Health and Welfare (AHAW): Scientific Opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. EFSA Journal 2010; 8 (7):1666. [82 pp.]. doi:10.2903/j.efsa.2010.1666. Available at: www.efsa.europa.eu.

European Food Safety Authority (EFSA), 2010c. EFSA Panel on Animal Health and Welfare (AHAW): Scientific Opinion on the welfare aspects of the practice of harvesting feathers from live geese for down production. EFSA Journal 2010; 8(11):1886. [57 pp.]. doi:10.2903/j.efsa.2010.1886. Available at: www.efsa.europa.eu.

EFSA (European Food Safety Authority), 2010d. Application of systematic review methodology to food and feed safety assessments to support decision making. The EFSA Journal (2010), 8(5):1637, pp. 1-90. Available at: www.efsa.europa.eu.

Grandin T, 2010. Recommended Animal Handling Guidelines and Audit

Guide for Cattle, Pigs, and Sheep (2005 Edition, with 2007 and 2010 Updates). AMI Foundation. Published by American Meat Institute Foundation. Available at: http://www.grandin.com/RecAnimalHandling Guidelines.html (accessed 21/02/2011).

IPCS (International Programme on Chemical Safety), 2004. Harmonization Project Document No.1 - IPCS risk assessment terminology. WHO, Geneva

Johnson NL, Kemp AW, Kotz S, 2002. Univariate Discrete Distributions, 3rd Edition. New York: Wiley Series in Probability and Statistics, 2005. J Hartung: Statistik, 13. Auflage. München: Oldenbourg.

OIE (Office International des Épizooties - World Organisation for Animal Health) 2004a. Handbook on Import Risk Analysis for Animals and Animal Products, Volume1. Introduction and qualitative risk analysis pp.57.

OIE (Office International des Épizooties - World Organisation for Animal Health) 2004b. Handbook on Import Risk Analysis for Animals and Animal Products, Volume 2. Quantitative risk assessment, pp.126.

OIE (Office International des Épizooties - World Organisation for Animal Health), 2005a. Guidelines for the Killing of Animals for Disease Control Purposes. Available at: http://www.oie.int/eng/ normes/mcode/en_ chapitre_3.7.6.htm.

OIE (Office International des Épizooties - World Organisation for Animal Health), 2005b. Terrestrial Animal Health Standards Commission, March 2005. Available at: http://www.oie.int/downld/ SC/2006/A_TAHSC_ MARCH2006.pdf#page=85.

Rowe G, Wright G, 1999. The Delphi technique as a forecasting tool: issues and analysis. International Journal of Forecasting, 15, 4, 353-375.

Spoolder H, van Niekerk TF, Velarde A, Fuentes C, Muller-Graf C, Mounier L, Ilieski V, Collins RA, 2010. Animal welfare risk assessment guidelines on housing and management. Project developed on the procurement project CFP/EFSA/AHAW/2009/01 (Question No EFSA-Q-2009-00844). Accepted for Publication on 15 December 2010.

Available at: http://www.efsa.europa.eu/en/supporting/pub/87e.htm.

WQ® (Welfare Quality®), 2009. Welfare Quality® assessment protocol for cattle. Welfare Quality® Consortium, Lelystad, Netherlands.

APPENDIX 1. DATA COLLECTIONS FORMS

Tailored forms were developed to collect data in abattoir on hazard exposure and likelihood of the adverse effects (by severity levels).

Each form was composed by two easy-to-handle worksheets:

1. Worksheet 1, containing the list of hazards potentially occurring to beef cattle at slaughter and analysed in this study, ordered by number (Figure 28)[46]. This worksheet was placed on the hand-held notebook to provide easy access to the list of hazards to be recorded in the second worksheet.
2. Worksheet 2, for recording the indicators of adverse effects and the hazards detected. The worksheet is reported in Figure 29 and Figure 30 (respectively for group and individual observations of pre-slaughter handling) and in Figure 31 (for stunning and killing phase).

   The method to fill in the forms is illustrated immediately following Figure 29 and applies to all forms.

[46] The hazards are described in Appendix 2. List of hazards

Figure 28. 1. Worksheet containing the list of hazards to beef cattle at slaughter used during the on-site observations

N	HAZARD	N	HAZARD	N	HAZARD	N	HAZARD
1	Energised prod	9	Animal side vision not blocked on the truck ramp (<1.5m)	17	Inadequate humidity	25	Wrong shot
2	Inadequate handling technique	10	Slippery truck ramp	18	Inadequate temperature	26	No check after stunning
3	Shouting	11	Steep truck ramp	19	Sharp barriers/driveway edges	27	Second shot delayed
4	Hitting	12	Narrow truck ramp	20	Too wide corridors	28	Too small stunning box
5	Finished noise due to truck	13	Sharp protrusions on the truck ramp	21	Too narrow corridors	29	Delayed stunning
6	Sudden change of light intensity	14	Dirty floor surface	22	Gates' opening size too small	30	Too big stunning box
7	Finished noise due to other than the truck	15	Too many animals in the flow	23	Faulty operating non-return gates (in the corridors or in the	31	Stunning box gate opening size too small
8	Too rapid/ uncontrolled funnelling into single file	16	Too fast animal flow	24	Overall and continuous noise at the entrance of the stunning box	32	Sharp barriers/driveway edges in the stunning box

Figure 29. Data collection form for *group* observations of pre-slaughter handling phases (by phase)

Date and time: ________________**Plant:**_____________ **Phase of the slaughter process:** ***(unloading, corridors, or stunning box entrance)***

Group observation N:___________**Group size:**_____________ **Time of the observation (beginning of slaughtering or after a while):**___________

Esc		Balk		Bru / Wou		Fall		Fall no rec		Lame		Hes		Jump		Mount		Mov back		Prod	Run		Slip		Turn around	
I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	I	H	I	H	I	H
HAZARDS occurring without causing an adverse effect																										

Notes:________________________________

The form above was used for the *group* observations of the phases unloading, corridors and entrance of the stunning box.

The codes “Esc”, “Balk” etc refer to the indicators of poor welfare identified in this study and illustrated in section 2.7.2.1, Table 6.

The form was filled in as follows:

a. When an indicator of poor welfare was spotted, the observer crossed the appropriate cell (I) and inserted next to it (cell H) the number corresponding to the hazard causing the adverse effect (provided in the worksheet in Figure 28). As the categories of severity were pre-assigned to the various welfare indicators, severity of the adverse effects was calculated simultaneously (section 2.7.2.2).
b. For hazards occurring without any consequent adverse effect, the hazard numbers were inserted in the space at the bottom of the worksheet.

If the hazard did not vary during the same observation (e.g. “dirty floor”), the hazard was ticked only once in the space at the bottom (to indicate “hazard present”) and every time it caused an adverse effect in the related column “H”.

If in the group observations a hazard affected the entire group, the number of animals in the group was noted in the part “notes”.

For the indicator “Prod” the related column “H” for the hazard was not included as already implicit in the code “Prod”, i.e. “stress, pain, etc due to prod” (section 2.7.2.2).

The same method was applied to fill in all data collection forms.

In the *individual* observations of the pre-slaughter handling phases (Figure 30) each animal was observed from unloading to the entrance of the stunning box. Thus each form contained 1 individual observation only.

As the stunning and killing process was very fast, the related data collection (Figure 31) was designed for containing 15 individual observations and avoiding changing the form for every new animal entering the stunning box.

Figure 30. Data collection form for *individual* observations of pre-slaughter handling phases (all phases)

Date and time: ________________Plant:____________

Individual observation N:___________ Time of the observation (beginning of slaughtering or after a while):____________

	Esc		Balk		Bru / Wou		Fall		Fall no rec		Lame		Hes		Jump		Mount		Mov back		Prod	Run		Slip		Turn around	
	I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	I	H	I	H	I	H
Unloading																											
Corridors																											
Entrance																											

HAZARDS occurring without causing an adverse effect

Unloading	Corridors	Entrance

Notes:________________________________

Figure 31. **Data collection form for *individual* observations of stunning and killing phase**

Date and time: ________________Plant:____________ Phase of the slaughter process: Stunning and killing

N of animals	Esc		Bru / Wou		Fall		Fall no rec		Bad stunned		Slip		Turn around		Back to C		Push		Voc alone		Hazards occurring without causing and advesre effect							
	I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	H	I	H								
1																												
2																												
3																												
4																												
5																												
6																												
7																												
8																												
9																												
10																												
11																												
12																												
13																												
14																												
15																												

Notes on the animals					
1		6		11	
2		7		12	
3		8		13	
4		9		14	
5		10		15	

APPENDIX 2. LIST OF HAZARDS

All hazards identified via the literature review and analysed for the two risk assessments performed in this study are illustrated in Table 35.

The table reports:

1. 1st column: N (i.e. number identifying he hazard). If the hazard occurs in more than one phase of the slaughter process, the hazard is repeated for each phase.
2. 2nd column: hazard name and place of occurrence (unloading (U); corridors (C); entrance of the stunning box (E); and stunning and killing (S/K)).
3. 3rd column: how the hazard was defined for the on-site observations. In this column "*indicators of poor welfare identified in this study*" refers to the indicators described in section 2.7.2.1.
4. 4th column: how the hazard was counted in the group observations of the phases unloading, corridors and entrance of the stunning box. At stunning and killing the observations were individual only.
5. 5th column: how the hazard was described (asked) in the questionnaire to the experts.

The method for counting the hazards is illustrated in section 2.6.2.

The hazards excluded from the analysis and the reasons for not considering them are reported in Table 36.

Table 35. List of hazards potentially occurring to beef calves at slaughter and analysed in this study[47]

N	Hazard	Definition for the on-site observations	Count in group obs	Description in the questionnaire
H2 **H37** **H68**	Energised prod (U, C, E)	Use of an energised prod to: • facilitate unloading and the move in the unloading area • drive the animals along the corridors • force the animals into the stunning box	Individual	Same
H3 **H38** **H69** **H96**	Inadequate handling technique (directed to the animals) (U, C, E, S/K)	At unloading, in the corridors, or at the entrance of the stunning box: the operator proceeds straight towards the animal invading its flight zone[48] and changes direction OR touches the animal in the wrong place (forepart of the body such as shoulder, neck, thorax, etc) *to make it move* and stops when the animal moves on again In the stunning box: The operator approaches the head from the front	Pair	Same, but merged with “Shouting” and “Hitting” (as in the literature)

[47] The legend to this table is illustrated in the text at the beginning of this Appendix.

[48] The flight zone of an animal is the area surrounding the animal that will cause alarm and escape behaviour when encroached upon. If a person enters the flight zone of an animal, the animal will move away. The size of the flight zone depends upon the tameness of the animal. Completely tame animals have no flight zone; that is, they will allow a person to approach and touch them.

N	Hazard	Definition for the on-site observations	Count in group obs	Description in the questionnaire
		or touches or hits the animal to make it place its head, without using e.g. water flow or other appropriate means		
H4 **H39** **H70** **H97**	Shouting (U, C, E, S/K)	The operator shouts improperly at the animal to make it move or, in the stunning box, to make it place the head in the right position, and stops	N of shouts heard by the observer	Merged with "Inadequate handling technique" (as in the literature)
H5 **H40** **H71** **H98**	Hitting (U, C, E, S/K)	The operator hits the animal to make the animal move or, in the stunning box, to make it place the head in the right position, using the handling tools inappropriately or the hands, and stops	Individual	Merged with "Inadequate handling technique" (as in the literature)
H11 **H45** **H76**	Finished, sudden noise (U, C, E)	Any noise (e.g. hissing/clanging/banging) originating e.g. from the gates, or any other part of the plant, etc, that starts and stops and that in the preliminary observations clearly induced at least one animal to show at least one of the indicators of poor welfare identified in this study	Triple	Any "non-continuous" noise (e.g. hissing/clanging/banging) originating e.g. from the gates, or any other part of the plant
H94	Overall and continuous noise in E	Global continuous noise originating in the stunning box that at the preliminary observations clearly induced at least one animal to show at least one of the indicators of poor welfare identified in this study	Equal to the number of animals observed	Continuous noise originating in the stunning box

N	Hazard	Definition for the on-site observations	Count in group obs	Description in the questionnaire
H9	Truck noise (U)	Any noise (e.g. hissing noise or noise due to the animals moving on the truck or ramp) produced by the decks, gates, or any other part of the truck during the unloading of the animals, that starts and stops and that in the preliminary observations clearly induced at least one animal to show at least one of the indicators of poor welfare identified in this study	Triple	Any noise (e.g. hissing noise) produced by the decks, gate, or any other part of the truck
H13	Open/short side barriers on the truck ramp (U)	Animal side vision not blocked (barriers <1.5m)	Group	Same
H14	Slippery truck ramp (U)	Slipperiness due to e.g. dirtiness (e.g. excess of manure that makes the ramp slippery) inter cleat distances > 20-30cm, etc) such as in the preliminary observations it clearly induced at least one animal to show at least one of the indicators of poor welfare identified in this study	Group	Same
H15	Steep truck ramp (U)	Steepness of the truck ramp: • ≥20% - 11 ° • or maximum slope of 20-25° if it is a non-	Group	Same

N	Hazard	Definition for the on-site observations	Count in group obs	Description in the questionnaire
		slippery floor with inter cleat distances at 20-30 cm • step between ramp and ground > 15cm height		
H16	Narrow truck ramp (U)	<2m	Group	Same
H17	Sharp protrusions on the truck ramp (U)	Sharp edges on the truck ramp such as the animals can get injured	Group	Same
H18 **H47** **H79**	Dirty floor surface (poor maintenance) (U, C, E)	Dirtiness (e.g. excess of manure that makes the floors slippery) such as in the preliminary observations it clearly induced at least one animal to show at least one of the indicators of poor welfare indentified in this study	Group at U and in C Individual in E	Dirty floor (e.g. excess of manure that makes the floors slippery)
H12	Inadequate funnelling into single file at unloading	When entering the corridors from the unloading bay the animals are squashed and the flow is not continuous This hazard was not considered in other phases as plant related (Table 36)	Pair	Inadequate funnelling into single file

N	Hazard	Definition for the on-site observations	Count in group obs	Description in the questionnaire
	(U)			
H22 **H51** **H82**	Too fast animal flow (U, C, E)	Too rapid animal flow, that in the preliminary observations clearly induced at least one animal to show at least one of the indicators of poor welfare indentified in this study	Group (Individual in E)	Too fast animal flow
H20 **H49** **H80**	Too many animals in the flow (U, C, E)	Too many animals in the flow, in such a way that the space allowance is inadequate (the animals are squashed)	Group U and C Individual in E[49]	U
H25 **H54** **H85**	Inadequate temperatur e (U, C, E)	Inadequate temperature (too hot, too cold)	Equal to the number of animals observed	Same
H26 **H55** **H86**	Inadequate humidity (U, C, E)	Humidity > 80% (due to inadequate ventilation)	Equal to the number of animals observed	Same

[49] For some hazards the unit of observation varied depending on the place of occurrence. For instance “Too many animals in the flow” in E was counted at individual level as the flow was continuous (no groups counted). Thus every time the hazard was detected, the number of animals exposed to it was counted.

N	Hazard	Definition for the on-site observations	Count in group obs	Description in the questionnaire
H33 H64 H108	Faulty operating non-return gates U, C, (E), S/K	The gate doesn't close once the animal has passed through it and allows the animal to move back to the previous phase of the process (this hazard was not included in the phase E as it may be confused with the stunning box gate, which is a specific hazard of the S/K phase)	Individual	Same
H28 H57 H88	Sharp barriers/driveway edges (U, C, E, S/K)	Sharp barriers/driveway edges such as in the preliminary observations at least one animal showed at least one of the indicators of poor welfare identified in this study (e.g. bruises/wounds)	Individual	Sharp barriers/driveway edges
H58	Inadequate driveway size and design (C, E)	Too wide corridors (the animals easily turn around)	Individual	Same
H59		Too narrow corridors (the animals easily get squashed)	Individual	Same
H31 H62 H106	Inadequate gates' size (U, C, (E) S/K)	Gates in the unloading bay or in the corridors or gate of the stunning box: opening size too small, so the animal is squashed (this hazard is not included in the list related to the phase E as it may be confused with the stunning box gate, which is a specific hazard of	Individual	Same

N	Hazard	Definition for the on-site observations	Count in group obs	Description in the questionnaire
		the S/K phase)		
H10 H44 H75	Sudden change of light intensity (U, C, E)	Any sudden change of light, too dark environment, glaring objects etc, that in the preliminary observations clearly induced at least one animal to show at least one of the indicators of poor welfare identified in this study	Individual	Any sudden change of light, too dark environment, glaring objects etc
H111	Stunning box size	Too small stunning box compared to the size of the animals (the animal is squashed - no adjustable sides in the stunning box – or is sort of “jammed” by the horns)	na	Too small stunning box
H112		Too big stunning box compared to the size of the animals (the animals can turn around OR two animals can fit in the stunning box at the same time)	na	Too big stunning box
H114	Delayed stunning once the animal has entered the stunning box	Whenever the stunning operation (placing of the head, etc) does not start immediately after the animal enters the stunning box	na	Delayed stunning once the animal has entered the stunning box

N	Hazard	Definition for the on-site observations	Count in group obs	Description in the questionnaire
H115	Wrong shot	Inappropriate stun caused by: • Wrong shooting position (not right distance from the head AND/OR inaccurate target area - not directed to brain centre) AND/OR • Wrong angle captive bolts not fired perpendicular AND/OR • air pressure too low	na	Same
H120	No check (after shot)	The operator does not check at all the appropriateness of the stun	na	Same
H116	Second shot delayed	When the first shot is unsuccessful, animal not immediately re-stunned	na	Same

Table 36. Hazards excluded from the analysis

Hazard	Description	Reason for not including the hazard
Inappropriate driving tools (sharp or pointed tools) (U, C, E)	Use of sharp or pointed tools, to: • unload the animals and drive them along the unloading bay • drive the animals along the corridors • force the animals into the stunning box (at the entrance of the stunning box)	Slaughter plant-related (as forbidden in the slaughter plant where the definitive observations took place)
Open barriers (U, C, E, S/K)	Animal side vision not blocked (barriers not concrete)	Slaughter plant-related
Inadequate floor surface (structural) (U, C, E, S/K)	In the unloading bay, in the corridors, in front of the stunning box entrance: on the basis of the preliminary observations, this hazard was defined as: inadequate floor surface (structural) due to slipperiness (gaps, potholes, etc), steepness, up/down, too rough, etc such as it clearly induced at least one animal to show least one of the identified indicators of poor welfare In the stunning box: improper design of the stunning box floor that caused at least one animal to slip OR to experience a loss of balance that alters its steadiness in the stunning box, with visible fear and frustration	Slaughter plant-related
Delayed shackling, hoisting and sticking (S/K)	Delay before exsanguination	Slaughter plant-related (dependent on the layout of the area)
Incorrect sticking (S/K)	Incorrect sticking, delayed bleed out	Not possible to assess in the slaughter plant (due to the distance of the

Hazard	Description	Reason for not including the hazard
		observer)
Gun wrong model (S/K)	/	Slaughter plant-related
Gun poorly maintained (S/K)	/	Not possible to assess
Inadequate cartridge design (S/K)	Design of the cartridge (mass, diameter and length) inappropriate to an effective stun	Not possible to assess
Stunning box design (S/K)	Improper design head restrainer, belly plate, rump push	Slaughter plant-related
Stunning box with dark entrance (S/K)	Entrance of the stunning box not properly lighted (entrance darker than the passageways)	Slaughter plant-related
Light in the stunning box (S/K)	Overhead light too intense	Slaughter plant-related
Dead ends (U, C)	In the corridors or unloading bay	Slaughter plant-related
Curvature < 120° (U, C)	In the corridors or unloading bay	Slaughter plant-related
Inadequate	When passing from a multiple-files corridor to a single-file corridor or into the	Slaughter plant-related

Hazard	Description	Reason for not including the hazard
funnelling into single file (C, E)	stunning box. The animals are squashed and the flow is not continuous	(depending on the layout of the corridors etc)
Overall and continuous noise originating in the stunning box (U, C)	On the basis of the preliminary observations, this hazard was defined as: global continuous noise originated in the stunning box (e.g. due to the shots, the stunned animals falling down, the chain and the hook to hoist the animals, etc)	Slaughter plant-related

APPENDIX 3. RESULTS OF THE ANALYSIS OF THE GROUP AND INDIVIDUAL OBSERVATIONS

In this section the results of the analysis of the group and individual observations of the pre-slaughter handling phases (i.e. unloading, corridors and entrance in the stunning box) are reported.

Table 37. Hazards occurring during pre-slaughter handling, for which on-site exposure assessment was estimated by *merging* the results of the group and individual observations

Reason for merging the results	Hazards
• the 2 CI overlapped • the unit of observation was the same in the group and individual observations • the number of hazards detected corresponded to the number of animals exposed to those hazards in both observations	1. H 2. Prod at unloading 2. H 10. Sudden change of light intensity at unloading 3. H25. Temperature (not adequate) at unloading 4. H26. Humidity (not adequate) at unloading 5. H28. Sharp barriers/driveway edges at unloading 6. H31. Gates' opening size too small at unloading 7. H33. Faulty operating non-return gates (at the exit of the unloading bay) 8. H 37. Prod in the corridors 9. H 38. Sudden change of light intensity in the corridors 10. H54. Inadequate temperature in the corridors 11. H55. Inadequate humidity in the corridors 12. H57. Sharp barriers/driveway edges in the corridors 13. H58. Too wide corridors 14. H59. Too narrow corridors 15. H62. Gates' opening size too small 16. H64. Faulty operating non-return gates (at the exit of the unloading bay) 17. H68. Prod at the entrance of the stunning box 18. H75. Sudden change of light intensity at the

Reason for merging the results	Hazards
	entrance of the stunning box 19. H82. Too fast animal flow at the entrance of the stunning box 20. H85. Inadequate temperature at the entrance of the stunning box 21. H86. Inadequate humidity at the entrance of the stunning box 22. H88. Sharp barriers/driveway edges at the entrance of the stunning box 23. H89. Too wide corridors at the entrance of the stunning box 24. H90. Too narrow corridors at the entrance of the stunning box 25. H94. Overall and continuous noise at the entrance of the stunning box

Table 38. Hazards occurring during pre-slaughter handling, for which on-site exposure assessment was estimated by using the *group* observations only

Reason for not merging the results	Hazards*	Reason for using the results of the group observations
• The 2 CI overlapped • Hazards not counted in the same way in the group and individual observations	1. H9. Noise due to truck (unit of observation in the group observations: triples) 2. H11. Noise due to other than truck (unit of observation in the group observations: triples) 3. H12. Too rapid/uncontrolled funnelling into single file (unit of observation in the group observations: pairs) 4. H45. Sudden noise in the corridors (unit of observation in the group observations: triples) 5. H76. Sudden noise at the entrance of the stunning box 6. H69. Inadequate handling technique at the entrance of the stunning box	• Group observations CI smaller than individual observations CI

*For the units of observation in the group observations, see also 2.6.2

Table 39. Hazards occurring during pre-slaughter handling, for which on-site exposure assessment was estimated by using the *individual* observations only

Reason for not merging the results	Hazards	Reason for using the results of the individual observations
• The 2 CI overlapped • Hazards not counted in same way in the group and individual observations	Unit of observation: group* 1. H13. Open side barriers or lower than 1.5 m on the truck ramp 2. H14. Slippery truck ramp 3. H15. Steep truck ramp 4. H16. Narrow truck ramp 5. H17. Sharp protrusions on the truck ramp 6. H18. Dirty floor in the unloading bay 7. H20. Too many animals in the flow at unloading 8. H22. Too fast animal flow at unloading 9. H47. Dirty floor in the corridors 10. H49. Too many animals in the flow in the corridors 11. H51. Too fast animal flow in the corridors	• The individual observations were more precise (smaller CI)
• The 2 CI did not overlap • Hazard counted in the same way in the group and individual observations	1. H79. Dirty floor at the entrance of the stunning box 2. H80. Too many animals in the flow at the entrance of the stunning box	The estimate at group level > estimate at individual level: the individual observations were used, as more precise (as the observer could focus

Reason for not merging the results	Hazards	Reason for using the results of the individual observations
		on the animals one by one)
The hazard was split into 3 sub-hazards in the on-site observations (i.e. "shouting", "hitting" and "inadequate handling technique") and in the group observations the 3 sub-hazards were counted in a different way	1. H3. Inadequate handling technique at unloading 2. H38. Inadequate handling technique in the corridors	In the individual observations the 3 sub-hazards "shouting", "hitting" and "inadequate handling technique" were counted in the same way (i.e. equal to the number of animals exposed to the hazard) and thus the results of the 3 individual observations for the 3 sub-hazards were merged and used in the final analysis

*For the units of observation in the group observations, see also 2.6.2

Table 40. Hazards occurring during pre-slaughter handling, for which on-site likelihood of the adverse effects (by severity levels) was estimated *merging* the results of the group and individual observations

Reason for merging the results	Hazards
• the 4 CI overlapped; • the number of indicators detected corresponded to the number of animals showing those indicators, both in the group and individual observations OR no adverse effect observed in one or both observations	1. H2. Prod at unloading 2. H9. Noise due to truck at unloading 3. H11. Noise due to other than truck at unloading 4. H12. Too rapid/uncontrolled funnelling into single file at unloading 5. H13. Open side barriers or lower than 1.5 m on the truck ramp at unloading 6. H14. Slippery truck ramp at unloading 7. H15. Steep truck ramp at unloading 8. H18. Dirty floor in the unloading bay 9. H20. Too many animals in the flow at unloading 10. H37. Prod in the corridors 11. H45. Sudden noise in the corridors 12. H47. Dirty floor in the corridors 13. H68. Prod at the entrance of the stunning box 14. H79 Dirty floor at the entrance of the stunning box

Table 41. Hazards occurring during pre-slaughter handling, for which on-site likelihood of the adverse effects (by severity levels) was estimated using the *group* observations only

Reason for not merging the results	Hazards	Reason for using the results of the group observations
• The 4 CI overlapped • In the group observations, the number of indicators detected did not correspond to the number of animals showing them	1. H69. Inadequate handling technique at the entrance of the stunning box 2. H80. Too many animals in the flow at the entrance of the stunning box	The group observations had a smaller CI

Table 42. Hazards occurring during pre-slaughter handling for which on-site likelihood of the adverse effects (by severity levels) was estimated using the *individual* observations only

Reason for not merging the results	Hazards	Reason for using the results of the individual observations
• The 4 CI did not overlap • In the group observations, the number of indicators detected did not correspond to the number of animals showing them	1. H49. Too many animals in the flow in the corridors 2. H94. Overall and continuous noise at the entrance of the stunning box	The individual observations were used, as more precise (e.g. adverse effects of cumulative-severity could be detected)
Not possible to compare the results of the group and individual observations	1. H3. Inadequate handling technique at unloading* 2. H38. Inadequate handling technique in the corridors*	*

*In the on-site observations these hazards were split into 3 sub-hazards (i.e. “shouting”, “hitting” and “inadequate handling technique”), which in the group observations were counted using different units of observations and in a different way (i.e. “shouting” and “inadequate handling technique” were equal to the number of shouts multiplied by 2; and “hitting” was equal to the number of animals hit).

In the individual observations the 3 sub-hazards were counted in the same way (i.e. equal to the number of animals exposed to the hazard) and thus the results of the 3 individual observations for the 3 sub-hazards were merged and used in the final analysis

At the entrance of the stunning box, the 3 sub-hazards “shouting”, “hitting” and “inadequate handling technique” were counted in the same way also in the group observations (due to the peculiarity of the situation) and thus they were merged into a single hazard, for which it was possible to use the results of the group observations (Table 41)

GLOSSARY

Adverse effect	The degree to which the animals fail to cope with the specific hazards
Adverse effect characterisation	Identification of welfare consequences and likelihood of these consequences that arise when the individual animals are exposed to the hazards
Anxiety	A feeling of uneasiness, apprehension or dread that occurs when a serious future problem is predicted
Boolean operator	Boolean operators are words used to combine terms or concepts when conducting electronic bibliographic searches. Examples include "AND" (used to narrow a search), "OR" (used to broaden a search) and "NOT" (used to exclude terms from a search)
Decision making using risk assessment	The process of weighing policy alternatives in the light of the results of risk assessment and, if required, selecting and implementing appropriate control options (i.e. prevention, elimination, or reduction of hazards or minimisation of risks), including regulatory measures (This term incorporates risk management).
Dose-response assessment	The determination of the relationship between the magnitude of exposure of animals to a certain hazards and the severity and frequency of associated effects on animal welfare
Empirical	The word empirical denotes information gained by means of observation, experience, or experiment. Empirical data are data produced by an experiment or observation
Exposure	For a given population of individuals, the extent to which a hazard that affects welfare is likely to be encountered
Exposure assessment	Assessment of the probability of exposure to the specific scenario in the target population. This is defined e.g. by the proportion of the population exposed, duration and frequency in time
Fear	A feeling that occurs when an individual perceives that there is danger or a high risk of danger
Flight zone	The flight zone of an animal is the area surrounding the animal that will cause alarm and escape behaviour when

	encroached upon. If a person enters the flight zone of an animal, the animal will move away. The size of the flight zone depends upon the tameness of the animal. Completely tame animals have no flight zone; that is, they will allow a person to approach and touch them
Frustration	The occurrence of high levels of causal factors promoting the occurrence of a behaviour that is prevented by the absence of a key stimulus or presence of a social or physical barrier
Hazard	Any aspect of the environment or the animal, alterations in which may have the potential to impair the welfare of animals, for example housing, management procedures, transport, stunning, killing, genotype
Hazard identification	Hazard identification is the process to identify, describe and clearly specify changes in features (i.e. hazards) that can impair the welfare of the animals under consideration and in the given scenario, causing adverse effect(s)
Lairage	A place where livestock are kept temporarily, e.g. in a market or in the slaughter plant
Magnitude of an adverse effect	The score resulting from the product of the severity and the duration of a negative effect on welfare due to the hazard under consideration
Model	A (simplifying) representation of the essentials (parameters, relations, processes, or mechanisms) of an existing system (or a system to be constructed) which incorporates existing knowledge and/or assumptions about the relationship between all system components in an explicit form that can be investigated by systematic or manipulative experiments
Pain	An aversive sensation and feeling associated with actual or potential tissue damage
Qualitative Risk Assessment	An assessment that generates an estimate of categorical nature or based on an ordinal scoring system. The outcome of such an assessment is a classification of output into descriptive categories
Quantitative Risk Assessment	An assessment that generates an estimate of a numerical nature directly tied to a measurement or calculation. Depending on the type of model tool used,

	an indication of the associated uncertainties - expressed either as extreme values, confidence intervals or prediction intervals are needed
Risk	A risk is a function of the probability of an adverse effect and the duration and severity of that effect, consequent to a hazard
Risk analysis	A process consisting of three components: risk assessment, risk management and risk communication
Risk assessment	The process of qualitative estimation of the probability of occurrence, severity and duration of adverse effects in a population
Risk communication	The interactive exchange of information and opinions concerning the risk and risk management among risk assessors, risk managers, consumers and other interested parties
Risk management	Decision making using risk and benefit analysis outcomes
Semi-quantitative risk assessment	Within risk assessment, probabilities of an event are assessed and described textually on a scale from negligible, indicating that the probability of an event or the estimated risk cannot be differentiated from zero (and in practical terms can be ignored) to extremely high
Severity	The extent of the negative effect on welfare as judged from scientific indicators of welfare at a particular time
Stress	An environmental effect on an individual which over-taxes its control systems and reduces its fitness or appears likely to do so
Stun	The process that renders insensible to pain or other sensory inputs
Systematic review	An overview of existing evidence pertinent to a clearly formulated question, which uses pre-specified and standardised methods to identify and critically appraise relevant research, and to collect, report and analyse data from the studies that are included in the review
Transparency	Characteristics of a process where the rationale, the logic of development, constraints, assumptions, value judgements, decisions, limitations and uncertainties of the expressed determination are fully and systematically stated, documented, and accessible for review

Uncertainty analysis	A method used to estimate the uncertainty associated with model inputs, assumptions and structure/form
Welfare	The state of an individual as regards its attempts to cope with its environment. Welfare can be measured and varies from very good to very poor. It includes the health and feelings of the animal as well as the results of other coping attempts
Welfare indicators	Measures used to assess the welfare of an animal

ABBREVIATIONS

C	Phase of the slaughter process: animals moving long the corridors
CAC	Codex Alimentarius Commission
CI	Confidence interval for probability of the hazards
CI1	Confidence interval for probability of severity 1 adverse effect(s)
CI2	Confidence interval for probability of severity 2 adverse effect(s)
CI3	Confidence interval for probability of severity 3 adverse effect(s)
CI4	Confidence interval for probability of severity 4 adverse effect(s)
cm	Centimeter
D	Duration
Di	Dispersion coefficient of the expert judgement on probability of the hazards
Di1	Dispersion coefficient of the expert judgement on probability of severity 1 adverse effect(s)
Di2	Dispersion coefficient of the expert judgement on probability of severity 2 adverse effect(s)
Di3	Dispersion coefficient of the expert judgement on probability of severity 3 adverse effect(s)
Di4	Dispersion coefficient of the expert judgement on probability of severity 4 adverse effect(s)
E	Phase of the slaughter process: passage from the corridors into the stunning box
EFSA	European Food Safety Authority
FAO	Food and Agriculture Organization of the United Nations
H	Hazard
K	Number of animals exposed to the hazard
L	Number of animals showing the adverse effect(s)
m	Meter
N	Number (and number of observed animals)

na	Not available
OIE	Office International des Épizooties (World Organisation for Animal Health)
P	Probability of hazard
P1	Probability of severity 1 adverse effect(s)
P2	Probability of severity 2 adverse effect(s)
P3	Probability of severity 3 adverse effect(s)
P4	Probability of severity 4 adverse effect(s)
S/K	Phase of the slaughter process: stunning and killing
T	Temperature
U	Phase of the slaughter process: unloading
WHO	World Health Organisation
WQ	Welfare Quality® project