Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia in Vaccinology 5 (2011) 221 – 226
Procedia in Vaccinology
www.elsevier.com/locate/ procedia

NICEATM-ICCVAM[#] International Workshop on Alternative Methods to Reduce, Refine, and Replace the Use of Animals in Vaccine Potency and Safety Testing: State of the Science and Future Directions
Bethesda, Maryland, USA, 14-16 September 2010

Humane endpoints in vaccine potency testing

Coenraad F.M.Hendriksen *

Netherlands Vaccine Institute (NVI), Bilthoven, The Netherlands and Utrecht University, Faculty of Veterinary Medicine, Utrecht, The Netherlands

Abstract

Vaccine potency and safety testing is characterized by extensive use of laboratory animals and a relatively high percentage of test methods that involve severe pain and distress. This is particularly true for tests that are based on infection or challenge with a virulent microorganism. Traditionally, vaccine potency tests on inactivated vaccines require a vaccination–challenge procedure using severe clinical signs or even lethality as endpoints.

For several of these vaccines, 3R methods have been developed that include a nonclinical endpoint, ultimately resulting in reduction of animal numbers and a significant decrease in severity level. An example is the use of serology in potency testing of tetanus and diphtheria toxoid vaccines.

For some potency tests, however, replacement of the challenge procedure is not (yet) possible, and the implementation of humane endpoints might be an approach to limit the level and duration of pain and distress. The application of these endpoints is now allowed in most pharmacopoeias.

Establishing humane endpoints in vaccine potency testing requires the identification of parameters that are predictive of death, or severe clinical signs, in the animal during the observation period. As a case study, we present the results of work we performed on the identification of humane endpoints in whole cell pertussis (wP) vaccine potency testing (the mouse protection test or the Kendrick test). In this potency test, mice are challenged by intracerebral route 14 days after immunization with a lethal dose of virulent *B. pertussis* microorganisms. Animals are observed for 14 days, and the number of mice per dose group surviving this period is used for probit analysis and estimation of potency.

We have studied two types of humane endpoints: clinical signs and pathophysiological parameters (body weight and body temperature). Clinical signs in a wP potency test range from piloerection, hunched back posture, apathy, and convulsions to moribund condition. Also body temperature drops, and animals lose up to 50% of their body weight post-challenge.

Parameters were "validated" for relevance (prediction of death within the observation period) and reliability. Recommendations are given for implementation of humane endpoints in vaccine potency testing, also taking into account potential obstacles.

Selection and/or peer-review under responsibility of the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM).

Keywords: vaccine potency testing; humane endpoints

[#] The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods and the Interagency Coordinating Committee on the Validation of Alternative Methods

* Corresponding author *e-mail address*: Coenraad.Hendriksen@nvi-vaccin.nl

1877-282X
Selection and/or peer-review under responsibility of the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). doi:10.1016/j.provac.2011.10.022

1. Introduction

Refinement of animal procedures is one of the leading ethical principles in biomedical research when animal replacement alternatives are not available. From an animal welfare point of view, priority should be given to those procedures that involve severe pain and distress. According to the statistics on the use of laboratory animals in the Netherlands, one of the few countries that keep statistics on severity levels, 11.3% of all animal experiments result in severe clinical signs, severe distress, or even lethality [1].

An important strategy to limit the level and/or duration of severity is the implementation of humane endpoints. A humane endpoint can be defined as ‘‘The earliest indicator in an animal experiment of (potential) pain and/or distress that, within its scientific context and moral acceptability, can be used to avoid or limit such consequences by taking actions such as humane killing or terminating or alleviating the pain and distress” [2].

Existing regulations on the use of laboratory animals in biomedical research promote the application of humane endpoints. For instance, Article 13.3 of the recently adopted revised regulation on animal experimentation in the European Union states that “Death as an end-point shall be avoided as far as possible and replaced by early and humane end-points” [3].

Humane endpoints are case specific and differ depending on the type of research and even on the type of experiment or animal model being used. This manuscript will focus on the identification of humane endpoints in vaccine potency testing, using the whole cell pertussis (wP) vaccine to illustrate possibilities and limitations.

2. Humane endpoints in vaccine potency testing

Significant numbers of laboratory animals are used for vaccine-related purposes. Animals are needed for vaccine research and development (R&D), nonclinical testing, and, particularly, for lot release testing required by statute. According to statistics in the Netherlands, the vaccines category accounted for about 15% of total laboratory animal use in 2009 [1].

Animal studies in vaccine R&D focus on pathogenicity of the virulent microorganism, antigen selection, antigen kinetics, and adjuvant. In nonclinical testing, the focus is on product safety, formulation, and application (dose, route of injection). In statutory lot release testing, animal tests are particularly needed to demonstrate potency and safety of each lot of vaccine produced. Of the total number of animals used for vaccine-related purposes, it is estimated that 30% are used for R&D and preclinical testing and 70% are used for lot release testing (Hoonakker, personal communication).

Due to the type of questions to be addressed in vaccine R&D and lot release testing, infection or challenge with the virulent microorganism is frequently performed. Consequently, clinical signs and lethality are common phenomena in these studies.

The animal model most relevant for implementation of humane endpoints is the potency test for lot release of inactivated vaccines. The model is generally based on a vaccination–challenge procedure. In short, animals are immunized with serial doses of the vaccine being studied and with a reference vaccine with known potency in International Units (IU), respectively. Several weeks after immunization, the animals are challenged with a lethal dose of virulent microorganisms. Deaths per dose group within a fixed observation period are recorded. The key parameter in the challenge test is the dose of vaccine protecting 50% of the animals against the lethal challenge (effective dose 50 = ED_{50}). Consequently, in a typical lethal challenge test, about 50% of all vaccinated animals are expected to die after challenge. Widely used vaccines incorporating lethal challenge potency testing include wP vaccine and rabies vaccines.

Specifications for test performance are given in the monographs of the European Pharmacopoeia, the Technical Report Series of the World Health Organization (WHO), and the U.S. Code of Federal Regulations (CFR) (USDA and FDA—rabies vaccine for veterinary use only). In the context of the lethal challenge procedure, it is worth mentioning that nonclinical endpoints have been developed for several products, such as tetanus and diphtheria toxoid [4, 5]. Assessing potency by serology instead of challenge is accepted for these products by the European Pharmacopoeia and the WHO.

Most guidelines on vaccine quality now allow for the application of humane endpoints. According to 9 CFR 117.4 (e) (USDA), “test animals that show clinical signs of illness that are due to the test may be treated or

humanely destroyed if the illness has progressed to a point … when death is certain to occur without therapeutic intervention" [6].

In order to identify humane endpoints for potency tests based on a challenge procedure, the working group Humane Endpoints for Lethal Parameters (HELP) was established in the late 1990s. The working group was a collaboration of the Netherlands Vaccine Institute (NVI, Netherlands), the Paul-Ehrlich Institute (PEI, Germany) and the University of Birmingham (UK), with financial support from the European Centre for the Validation of Alternative Methods (ECVAM). The focus of the NVI laboratory in the HELP activities was on wP vaccine potency testing.

3. Whole cell pertussis (wP) vaccine potency testing

In the past, wP vaccine was used worldwide in basic immunization programs. Due to the intrinsic side effects of the wP vaccine, an improved vaccine was developed. Nowadays, most Western countries use an improved vaccine, the acellular pertussis (aP) vaccine, based on a limited number of antigenic components of *Bordetella pertussis*. Nevertheless, the wP vaccine still is and will be the vaccine of choice in a large part of the world, particularly Third World countries, as aP vaccines are far more expensive than wP vaccines. As vaccine manufacturers in Western countries produce wP vaccine for export, the lethal challenge test continues to be performed routinely in these countries. It is estimated that annual use of animals for this test in Europe number over 100,000 mice [7].

The wP potency test, also called the mouse protection test (MPT) or Kendrick test, was developed around 1939 [8]. Groups of mice are immunized with three or four serial doses of the wP vaccine being studied and the reference preparation, respectively. After two weeks, animals are challenged by intracerebral route with 10 µl of 50 times the dose of virulent *B. pertussis* that is lethal for 50% of the animals injected with that dose ($50LD_{50}$). Intracerebral challenge is followed by an observation period of two weeks. The endpoint used is the number of deaths per dose group within the observation period. Each test requires at least 120 animals. Additionally, about 40 mice are needed for titration of the challenge culture.

3.1. Selection of humane endpoints

To identify parameters to be used as humane endpoints in wP potency testing, we divided the study into two parts: an evaluation phase and a validation phase [9]. None of the studies required additional animal use because they were run in parallel with routine lot release tests. Parameters monitored were clinical signs, body weight, and body temperature. These parameters were considered relevant in infectious disease models and are easily monitored in large numbers of animals. Clinical signs were recorded by observing the animals twice a day and by undisturbed videotaping. Videos were analyzed afterward. Body temperature was measured twice a day using both a temperature-sensitive transponder inserted subcutaneously and an external temperature monitor.

The evaluation phase of the study focused on the types of clinical signs seen after challenge and the possibility of grouping clinical signs in successive stages. Additionally, body temperature was monitored before and after challenge. **Table 1** shows the typical clinical signs of diseased animals and the order in which they occur. In addition to general clinical signs that occurred in all diseased animals, less common clinical signs were observed, such as circling, paralysis, and hypersensitivity. These clinical signs were not included in the stage categories. The first clinical signs related to virulent *B. pertussis* challenge were observed starting about three days after challenge.

Typically, body temperature of diseased animals decreases. This finding was confirmed in all studies. A slight increase in body temperature was seen only in the early days after challenge; therefore, we concluded that decrease in body temperature reflects disease progress. Also, decreases in body temperature and body weight were translated to ranges in order to allow easy evaluation. The ranges and clinical stages are given in **Table 2**.

In the validation phase of the study, predictivity and reliability of stages and ranges for death/survival in the observation period were assessed retrospectively for each individual animal in a series of routine batch release tests. A typical health monitoring sheet for two animals, one that died early in the observation period and one that died late in the observation period, is given in **Tables 2a** and **2b**.

Table 1. Grouping of clinical signs after intracerebral challenge of mice with virulent *B. pertussis*

Clinical Stages	Cardinal Clinical Signs[a]
0	Normal healthy animal
1	Reduced alertness Dull fur (piloerection, poor grooming) Reduced food intake Less social contact
2	Inactivity (lethargy) Hunched posture Social isolation No food and water intake Dehydration
3	Same as stage 2 + General weakness (e.g., disturbed locomotor activity) Dehydration Emaciation
4	Tonic and clonic convulsions (seizures) Comatose or recumbent Cachexia Moribund
	Death

[a] Clinical signs within two days after challenge are attributed to the challenge procedure itself, and those animals were excluded from further study.

Table 2a. Health monitoring sheet of clinical signs[a], body weight[b], and body temperature[c] of an animal (Mouse 2.1) that died early after challenge (Day 14)

Day	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Clinical Stage					1	2	4	[d]							
Weight (%)	100	95	88	79	72	65	60								
Body temp. (^{0}C)[d]	37.8	37.8	37.9	37.4	35.2	33.3	27.6								

[a] Clinical signs occurring at Days 14 and 15 were related to the challenge procedure.
[b] Body weight loss ranges: 20 - 30%; 30 – 40%; 40 – 50%; > 50%
[c] Body temperature drop ranges: 35.5 – 35.0^{0}C; 35.0 – 34.5^{0}C; 34.5 – 34.0^{0}C; <34.0^{0}C
[d] Day animal died

Table 2b. Health monitoring sheet of clinical signs[a], body weight, and body temperature of an animal (Mouse 1.3) that died late after challenge (Day 14)

Day	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Clinical Stage					1	1	2	2	3	3	3	3	3	[b]	
Weight (%)	100	92	88	76	70	65	64	66	57	58	55	53	51		
Body temp. (^{0}C)	37.0	37.1	37.4	36.9	36.3	35.1	34.7	34.5	33.5	33.2	32.7	30.2	26.2		

[a] Clinical signs occurring at Days 14 and 15 were related to the challenge procedure.
[b] Day animal died

Two conditions were considered crucial for the validity of identified humane endpoint stages and ranges. The first condition was low frequency of both false positive predictions and false negative predictions. *False positive* means that, based on the endpoint stage/range, the animal would have been euthanized, while, in fact, the animal would have survived the observation period. *False negative* means that the animal had already died before the

endpoint stage/range had been reached**. Table 3** shows the results of a typical study in which 76 animals were immunized with the reference wP preparation. In this experiment, 30 animals died during the observation period, while 46 animals survived. The number of false positive and false negative predictions is given, as well as the percentage of false positive and false negative predictions.

An endpoint with a high frequency of false positive results would invalidate the potency study; an endpoint with a high frequency of false negative results would compromise an effective animal welfare policy. Consequently, the ideal endpoint would be the endpoint with a low frequency of false positive and a low frequency of false negative results. As can be seen from **Table 3**, a drop in body temperature to 34.5 - 34.0^{0}C and a clinical stage of 3 meet the conditions for predictivity. It also can be noted that a loss of body weight is not a valid endpoint in wP potency testing: a 20 - 30% body weight loss results in a high frequency of false positives, a body weight loss > 40% results in a high frequency of false negatives, while a body weight loss of 30 - 40% results in a high frequency of both false positives and false negatives.

Table 3. False positive and false negative predictions when applying various endpoint criteria for humane euthanasia

Humane Endpoint	False positive[a]	%	False negative[b]	%
Temperature				
35.5 – 35.0^{0}C	9	19.5	0	0
35.0 – 34.5^{0}C	2	4.3	2	6.7
34.5 – 34.0^{0}C	0	0.0	2	6.7
< 34.0^{0}C	0	0.0	4	13.3
Body weight loss				
20 – 30%	21	45.6	0	0.0
30 – 40%	10	21.7	6	20.0
40 – 50%	1	2.2	25	83.3
> 50%	0	0.0	30	100.0
Clinical stage				
1	22	47.8	0	0.0
2	11	23.9	0	0.0
3	1	2.7	2	6.7
4	1	1.0	9	30.0

[a] Animals surviving that exhibited the listed endpoint but would have been erroneously killed if the endpoint had been used.

[b] Animals dying that did not exhibit the listed endpoint and thus would have died naturally and would not have been humanely killed if this endpoint had been used.

4. Discussion

In the presented study, we demonstrated that particular stages for body temperature and clinical signs are both relevant and reliable to replace death as the endpoint in a wP vaccine potency test. However, it should be kept in mind that endpoints cannot generally be transferred as such to other animal models. Thus, in our study, using the wP potency test as a case study, body weight appeared not to be a valid endpoint, while in other animal models, such as for rabies potency testing, body weight appeared to be predictive for death [10].

With regard to body weight, for example, using intervals of 2 grams in the 30% to 50% body weight loss range could further refine endpoint stages. Also, combining scoring for two endpoints could refine predictivity.

Regarding predictivity, most pharmacopoeias now accept the use of humane endpoints. However, an informal inquiry of vaccine manufacturers, performed by HELP several years after pharmacopoeias allowed the use of humane endpoints, showed that few manufacturers actually had implemented humane endpoint strategies (unpublished data). Although the reasons for reluctance may be diverse, several possibilities are suggested below:

- Clinical signs are generally subjective in nature, and interpretation differences might occur between observers.
- Using humane endpoints implies that animals have to be observed more frequently than only once a day, which will increase labor costs of experiments.
- Humane endpoints are model specific. Consequently, for each product, studies must be performed to identify humane endpoints and to validate predictivity.
- Although use of humane endpoints has been implemented in most pharmacopoeias (without being a strict requirement), regulatory authorities in individual countries might still adhere to results obtained by lethal challenge. Manufacturers will react to the diversity in specifications by providing challenge results.

Not all of the listed obstacles are easy to tackle. International acceptance of humane endpoints will require harmonization or mutual recognition of test requirements. Although this has been a key topic of discussion for quite some time, little progress has been made. Arguments can successfully be made in favor of implementing humane endpoints, including the facts that additional costs are limited and the subjectivity of monitoring clinical signs can be avoided with additional on-site training. Information on humane endpoints and clinical signs will be provided shortly on the website: www.humane-endpoints.info, which is a website of the Department of Animals in Science and Society of the Faculty of Veterinary Medicine, Utrecht University.

Implementing humane endpoints in vaccine quality control does not result in scientific benefits or economic profits but is a matter of humane care and, consequently, of ethics. Although not responsible, regulatory authorities could play a role in encouraging implementation of humane endpoints when reviewing manufacturer's submissions for lot release based on lethal challenge end-points.

References

[1] Zo doende 2009. Jaaroverzicht van de Voedsel en Waren Autoriteit over dierproeven en proefdieren (Annual report of the Food Safety Authority about animal experimentation and laboratory animals). VWA, 2010 The Netherlands (http://www.vwa.nl/)

[2] Hendriksen C, Morton D, Cussler K. Use of humane endpoints to minimise suffering. In: The COST Manual of Laboratory AnimalCare and Use. Refinement, Reduction and Research. Eds. Howard B, Nevalainen T and Perreta G. CRC Press, Boca Raton, US. ISBN 978-1-4398-2492-4, 2011:333-355.

[3] European Commission. Directive 2010/63/EU on the Protection of Animals Use for Scientific Purposes. Official Journal of the European Union, 20.10.2010, L276/33

[4] Winsnes R, Hendriksen C. Collaborative study for the validation of serological methods for potency testing of tetanus toxoid vaccines for human use. Pharmeuropa Bio 2000;1:83-124.

[5] Winsnes, R, Sesardic D, Daas A, Behr-Gross ME. Collaborative study for the validation of serological methods for potency testing of diphtheria toxoid vaccine (part 2). Pharmeuropa Bio 2006;1:73-88.

[6] USDA, Animal Welfare. Final Rules: CFR, Title 9, Animals and Animal Products, Part 117 Animals at Licensed Establishments, 117.4 (e) - Test animals Animals and Animal Products, 1998 page 813

[7] Von Hunolstein C, Gomez Miguel MJ, Pezzella C, Scopetti F, Behr-Gross ME, Halder M, Hoffmann S, Levels L, van der Gun J, Hendriksen C. Evaluation of two serological methods for potency testing of whole cell pertussis vaccines. Pharmeuropa Bio 2008;1:7-18.

[8] Kendrick PL, Eldering G. A study in active immunization against pertussis. Am J Hyg 1939;29:133-53.

[9] Hendriksen CFM, Steen B, Visser J, Cussler K, Morton D & Streijger F. The evaluation of humane endpoints in pertussis vaccine potency testing. In: Humane Endpoints in Animal Experiments for Biomedical Research. Hendriksen CFM & Morton DB (eds). The Royal Society Medicine Press, London (UK). 1999:106-113.

[10] Cussler K, Morton DB & Hendriksen CFM. Possibilities for the use of humane endpoints in vaccine potency tests. In: Progress in the Reduction, Refinement and Replacement of Animal Experimentation. Balls M, Van Zeller A-M & Halder M. (eds.). Elsevier, Amsterdam, 2000:915-927.