Refinement of Vaccine Potency Testing with the Use of Humane Endpoints

Coenraad F. M. Hendriksen and Björn Steen

Introduction

Public concern over the use of laboratory animals has increased in recent years. Laboratory animal procedures involving unrelieved pain and distress (e.g., the footpad injection of rodents with Freund's adjuvant), which were routine in the past, now require strong justification. Moral concern is reflected in the existing legislation on animal experimentation in the United States (CFR 1992) and in Europe (Council of Europe 1986) and in the publication of an increasing number of best practice codes (e.g., on immunization protocols). In many countries since the early 1990s, animal ethics committees have been established, housing conditions have been improved, and training curricula have been implemented for those working with laboratory animals.

Published results of a poll taken in the United Kingdom about public views on animal experimentation (Aldhous et al. 1999) reveal that people provided with impartial information appear to weigh the pros and cons of each experiment carefully before deciding whether they are willing to support the use of animals in research. The results also reveal that potential distress is an important factor in the public's support of specific uses of animals.

Death may be used as the experimental endpoint in cancer research, studies of infectious disease, carcinogenicity and toxicology (e.g., the LD50 test), drug comparisons, and vaccine potency testing (Hamm 1995). In this article, we discuss humane endpoints that can replace lethality as an endpoint in vaccine potency testing.

Vaccine Potency Testing

The production of most human and veterinary vaccines relies on the conventional technique of attenuation or inactivation of the virulent microorganism (e.g., rabies and polio vaccine) or on detoxification of the toxin (e.g., diphtheria and tetanus toxoid). As a consequence, various factors may affect the quality of the final product, necessitating extensive quality control testing of each batch of vaccine produced to ensure that its administration is both safe and sufficiently potent to induce protective immunity. The size of a batch varies but is usually between 100 and 1000 liters (about 2 x 10^3 and 2 x 10^6 human doses). Quality control is based on strict standard testing requirements determined by regulatory bodies such as the US Food and Drug Administration (FDA), the US Department of Agriculture (USDA), and the European Pharmacopoeia, or in guidelines such as those of the World Health Organization (WHO).

Many vaccine quality testing procedures use animal models, resulting in the need for large numbers of animals. Although no exact percentages are available, animal models for this purpose are estimated to comprise about 10% of the total number of laboratory animals used in biomedical research and testing (Hendriksen et al. 1994)—approximately 1 million mice, rats, and guinea pigs in European Union countries. In addition, the 10% estimate may even increase in view of the high priority now given to preventive health care policies. Reasons for this high priority include efforts to address the emergence of antibiotic-resistant strains of bacteria, increasing public health impact of viral infections, the high incidence of infections in livestock industries, the implementation of the WHO Expanded Programme on Immunization, and the establishment of vaccine production facilities in developing countries.

In vaccine quality control, animals are used for either safety or potency testing. The potency test on inactivated vaccines is particularly relevant to humane endpoints. The basic principle of the potency test is an immunization-challenge procedure, which dates back to the smallpox experiments of Jenner in 1798. Another historical example, and milestone in the early days of vaccine development, was Louis Pasteur's 1881 public demonstration of the effectiveness of an anthrax vaccine developed in his laboratory (Parish 1965). Three weeks after the immunization of 25 sheep, one goat, and six cows, these and an equal number of nonimmunized animals were challenged with the virulent anthrax organism. All of the immunized animals survived, and all the control animals died, thus demonstrating the efficacy of the vaccine.

Many of the potency tests for inactivated vaccines currently used are modifications of the immunization-challenge procedure. Frequently used models are quantitative multidilution tests, qualitative single-dilution tests, and indi-
rect protection tests whereby serum samples of immunized animals are titrated by an in vivo neutralization test (Hendriksen et al. 1994; Weisser and Hechler 1997). For human vaccines, animal models must be used, but veterinary vaccines can also be tested in the target animal species. Generally, large numbers of animals (even as many as 100) are needed for each test, as is the case for whole cell pertussis vaccine (described below) and tetanus and diphtheria toxoids.

The third edition of the European Pharmacopoeia (Council of Europe 1997), the legally binding guidelines document for the Member States of the Council of Europe, includes 49 monographs on inactivated vaccines. Approximately 17 of the monographs specify lethality as the endpoint to be used in the potency test, including testing for pertussis, tetanus, diphtheria, rabies, leptospira, and clostridial vaccines. Some of the monographs, such as those related to tetanus and diphtheria vaccines, offer the possibility of using either lethality or clinical signs. The fact that lethality is specified as the endpoint reflects the primary scientific goal of a vaccine potency test: to demonstrate that the vaccine is able to protect against the worst possible case—a lethal infection dose with the pathogen. However, there are other pragmatic reasons. Death is an objective test parameter that is easy to standardize, and monitoring for death requires little time and only a minimal level of training.

Regulatory Bodies and Animal Welfare

Pharmacopoeias and other regulatory bodies are becoming more aware of animal welfare issues, partly because these bodies must apply the principles of animal welfare existing laws and regulations. For example, US government agencies that develop requirements for testing have to consider the federal government principles for the utilization and care of vertebrate animals used in testing, research, and training. Article IV of these Principles states, “Proper use of animals, including the avoidance or minimization of discomfort, distress and pain when consistent with sound scientific practices, is imperative” (OPRR 1996, p. i).

The WHO, the lead organization with regard to vaccine quality control guidelines in third world countries, has recently issued a guideline for the humane treatment and care of animals used in the production and quality control of vaccines (WHO 1997). Article II of the general recommendations states, “It is the legal and ethical obligation for all those involved in vaccine production and quality control to have a humane regard for their animal subjects, and to prevent as far as possible pain and distress…” (p. 211). The same guideline also explicitly refers to the introduction of humane endpoints: “Whenever possible, the lethality endpoint should be replaced with other parameters” (p. 212). Furthermore, the USDA has recently introduced a new paragraph in the Code of Federal Regulations, Title 9, requiring the introduction of humane endpoints: “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” (USDA 1998, par. 117.4, p. 673).

The European Pharmacopoeia reportedly will include a similar paragraph in its monographs (P. Castle, European Pharmacopoeia, personal communication, 1999). As can be seen from the wording in the USDA paragraph, validation of the predictive power of the humane endpoint will be needed before permission can be obtained from the regulatory authority. The European Pharmacopoeia already has indicated that it is willing to encourage work in this direction (Castle 1999).

Humanness and Humane Endpoints

The guiding principle in current regulations on the use of laboratory animals is the concept of replacement, reduction, and refinement (the “Three Rs”), introduced by Russell and Burch (1959) as a way of diminishing or removing “inhumanity” in animal experiments. The central problem, they said, is to determine what is and what is not humane, and how humanness can be promoted without prejudice to science and medical aims.

Although lethality has long been seen as an unavoidable consequence of the challenge procedure in most vaccine potency tests, there is now a growing interest in the Three Rs approach. Replacement is the best alternative from an animal welfare point of view. In vitro potency tests have been described for a large number of products, such as for rabies vaccine and leptospira vaccines (for an overview, see Hendriksen et al. 1994, Weisser and Hechler 1997). Some in vitro methods are based on the physicochemical and immunochemical characterization of vaccine batches. Another interesting in vitro approach is to focus on specific immunologic parameters, such as cytokine responses (McCullough et al. 1997), as an indication of the vaccine’s immunogenicity or to study its functional characteristics (Funnell et al. 1999). Although some progress can be expected in the near future, the validation and acceptance of in vitro methods has been a tedious process, and its success has been limited to date. Most of the inactivated vaccines still require an immunization-challenge procedure to determine the biologic activity of a new vaccine batch. For these reasons, reduction of numbers of animals and refinement, including the use of humane endpoints, appear to be the best avenues to improved animal welfare in vaccine potency testing.

A humane endpoint can be defined as the point at which an experimental animal’s pain and/or distress can be terminated, minimized, or reduced by actions such as killing the animal humanely, terminating a painful procedure, or providing treatment to relieve pain and/or distress (CCAC 1998). Parameters for humane endpoints can be very diverse and include general clinical signs, nonspecific pathophysiologic effects (e.g., change in body weight or temperature), and behavioral effects (see Morton and Griffiths 1985) or they can be specific and study-dependent (e.g., paralysis in the
case of tetanus). Some of the specific variables that have been described in relation to vaccine potency tests or to infection models are discussed below.

One approach that might be considered as a humane endpoint and that has already proven successful is the replacement of the challenge procedure by an in vitro serologic test (e.g., enzyme-linked immunosorbent assay, hemagglutination, or VERO cell test) to assess antibody responses of animals after immunization. For the toxoid vaccines, in particular, good correlation between the titer of antibodies induced after immunization and the level of protection after challenge has been confirmed (Hauer 1997; Hendriksen et al. 1991; Kreeftenberg et al. 1985). Validation studies are now being finalized, and it is expected that some of the challenge tests (e.g., for tetanus toxoid and for some clostridial vaccines) will be replaced by serologic tests in the near future. An additional advantage of a serological potency test is that a precise value (antitoxin titer) is used instead of a qualitative outcome (death/survival), which increases the amount of data per animal and requires fewer animals without loss of test precision (Hendriksen et al. 1987).

If a challenge test is still unavoidable (e.g., if the immunity is [Th1] cell mediated), then an alternative approach might be to use the clearance rate of the microorganism after challenge, rather than lethality, as an outcome for evaluating the potency of a vaccine. Respiratory infection models have been described for the feline viral rhinotracheitis vaccine and for the whole cell and acellular pertussis vaccines (Mills et al. 1993). Hematologic variables such as neutrophil counts and plasma triglyceride concentrations have also been proposed (Toth et al. 1993).

The use of clinical signs as a criterion for euthanasia is emphasized in many papers and in some of the European Pharmacopoeia monographs. Such signs can be monitored, for example, in the development phases for new tetanus and diphtheria toxoids; however, there is a certain reluctance to include clinical signs in vaccine monographs for reasons of objectivity and cost effectiveness. In addition, information regarding the extent to which a clinical sign is a prognostic indicator for death is often not available.

When it is not possible to kill an animal humanely, supportive therapy such as fluid replacement and pain relief might be considered. Analgesic treatment was studied in the potency testing of the Clostridium chauvoei vaccine (Anonymous 1987) in which both partially protected and unprotected guinea pigs developed serious gangrenous lesions that eventually led to death. From the results of the study, it was concluded that the analgesic effect of buprenorphine is limited to the first 24 hr after challenge (the disease process takes several days). No deleterious effects were noted, only the decrease in the immunoglobulin A titer was statistically significant. It is possible that laboratory animals are denied analgesic treatment more often than is necessary.

**Strategy for Establishing Humane Endpoints**

A strategy to establish humane endpoints for animal testing should comprise the following steps: prioritization, test analysis, identification and evaluation of potential endpoints, validation of the endpoints selected, and, if tests are based on regulatory requirements, acceptance by regulatory authorities.

**Prioritization**

Aspects to be considered in prioritizing include aspects such as numbers of animals used (both locally and generally), animal distress, frequency of the test, and availability of readily transferable and applicable technologies. The possible existence of other more humane testing guidelines should also be investigated and considered. Introducing humane endpoints within specific guidelines will be arduous; however, acceptance will result in a substantial impact on animal welfare.

**Test Method Analysis**

The objective of test analysis is to fully consider opportunities for replacement and reduction alternatives. Consideration could include the use of in vitro or serologic methods, reduction of the number of animals per group, or reduction in the number of animal groups.

**Identification of Humane Endpoints**

Humane endpoints should be demonstrated to be appropriate for the type of test in a pilot study or in ongoing regulatory studies. We recommend using pragmatic as well as scientific arguments to tailor the endpoints to the specific type of test. Laborious and time-consuming techniques (e.g., monitoring biochemical parameters) will not be very helpful if the tests are routine, the number of animals used are extensive, and the time interval between finding an elevated value of a variable that indicates impending death and the "inhumane" endpoint is rather short. However, these values can be helpful when studying proposed endpoints.

**Evaluation of Humane Endpoints**

The relevance of selected endpoints for an intended purpose should be assessed in the evaluation phase of the study. If the
alternative endpoints are sought to accurately predict imminent death, then the study should focus on evaluating the correlation between that endpoint and the death or survival of the animal (e.g., the relation between the level of paralysis and death in the case of tetanus). The reliability of the alternative endpoint should also be determined by evaluating the reproducibility of the humane endpoint within and between laboratories. In this regard, death has been (and unfortunately still is) a very attractive endpoint because it is a highly reproducible endpoint among different laboratories.

Validation

An endpoint selected on the basis of the evaluation study should be validated in a number of (e.g., five) routine tests to verify relevance and reproducibility. In the case of regulatory tests, other laboratories should be invited to participate in the validation study. Results of the evaluation and validation study should be published in peer-reviewed journals, and if the data are considered sufficiently sound by regulatory bodies, then the monograph should be revised accordingly.

We used this strategy to develop and validate humane endpoints for the whole cell pertussis vaccine potency test (Hendriksen et al. 1999). Some of the results of our study are summarized below.

Establishing Humane Endpoints for the Potency Test on the Whole Cell Pertussis Vaccine: A Case Study

Prioritization

Prioritizing was not difficult because one of the most frequently used products in medical health care programs is the whole cell pertussis vaccine. Although a new type of pertussis vaccine with fewer side effects (the acellular pertussis vaccine) was introduced to the market a few years ago, the less expensive whole cell vaccine is still widely used. Approximately 30% of all animals in our laboratory at the National Institute of Public Health and the Environment, Bilthoven, The Netherlands, are used for potency testing of the whole cell pertussis vaccine. The use of mice for this test is also extensive in western and third world countries. The design of this test is presented in the requirements of regulatory bodies such as the FDA, European Pharmacopoeia, and WHO, which specify an immunization-challenge test in mice—the Kendrick test (Kendrick et al. 1947).

According to this design, three or four groups of not fewer than 16 mice/group are each immunized intraperitoneally with a successive dilution of the test vaccine or a standard vaccine. A few unvaccinated groups are added for the control of the challenge culture. Two weeks after immunization, the animals are challenged intracerebrally with 10 µl of fully virulent pertussis organisms (Bordetella pertussis) while under anesthesia (it should be noted that in some laboratories, the animals are challenged without anesthesia). The animals are then observed for 14 days. Unprotected (animals immunized with a low vaccine dose that either does not protect or only partly protects against infection) and non-immunized (control) animals may develop a lethal infection and die, generally within 4 to 8 days after challenge. According to the CCAC (1991) scale of severity, this type of experiment would be categorized in the highest degree of severity (category 4). In the United Kingdom, this protocol is categorized as “substantial.” The parameter used in calculating the potency of the product is the ED50, the vaccine dose that protects 50% of the animals against a lethal challenge.

Test Analysis

Alternative, nonanimal approaches for replacing the challenge procedure in the Kendrick test (Canthaboo et al. 1999; Mills et al. 1993; Van der Ark et al. 1994) and possibilities for reduction alternatives have been described (Hendriksen et al. 1987); however, these approaches require extensive validation before they will be accepted. For this reason, we decided to identify humane endpoints that would allow the animals to be euthanized at an early stage.

Identification of Humane Endpoints

Some of the characteristics of the Kendrick test are described in the preceding paragraph. It is obvious that these characteristics limit the types of alternative endpoints that can be applied. Ideally, the endpoint should be easy to monitor, be reproducible, be a valid predictor of death, and result in the greatest reduction in distress, both in time and intensity (Hendriksen et al. 1999). For our study, we decided to focus on pathophysiologic variables (body weight and temperature) and on clinical signs because these variables are often described in relation to infection models (Acred et al. 1994; Cussler et al. 1998; Soothill et al. 1992) and are manageable in routine potency tests.

Evaluation of Humane Endpoints

We evaluated several variables in a pilot study of 76 animals. The animals were immunized with four (serial) dilutions (n = 16 per dilution) of the whole cell pertussis reference preparation and challenged 14 days thereafter. The study also included a positive control (challenge only, n = 8) and a negative control (challenge with the vehicle only, n = 4). Animals were monitored in the morning and in the afternoon for the required observation period of 14 days. Body weight was recorded every day, and body temperature twice a day. To exclude the effect of stress on temperature recording and because manipulation of diseased animals might lead to convulsions and death, we implanted temperature-sensitive mi-
crochip transponders subcutaneously 3 days before the challenge. As shown by Kort et al. (1999), body temperature generated by transponders (either subcutaneously or intraperitoneally) does not differ significantly from rectal temperature.

Before the evaluation, we videotaped and observed the animals in routine tests. We used score sheets (Morton 1995) to compile an inventory of the clinical signs commonly seen in a pertussis infection, identify the cardinal signs, and relate these signs to disease progress. Pertussis challenge in mice generally follows a fixed pattern: Animals are characterized by reduced activity, lack of grooming, ruffled fur, crouching, loss of appetite and reduced weight, dehydration, sunken eyes, apathy, loss of locomotor control, and finally convulsions that ultimately lead to death. The first signs of pertussis generally occur within a few days after challenge. Most of the animals died within 4 to 8 days after the challenge. Based on the score sheet and videotape data, we grouped cardinal signs according to their sequence into five clinical levels (Table 1) and coded the levels from stage 0 to 4. The results are summarized below.

Table 1 is a list of the predictability of death for each of three variables evaluated: body temperature endpoints, body weight loss, and severity of clinical score. "False positives" are animals that in fact survive when one of their indicated endpoint values of various variables (body temperature, body weight loss, and clinical score) is used to predict death. Thus, these animals would be erroneously killed if this value were used as the endpoint. "False negative" denotes that the animal died without the value of the variable either decreasing (body temperature and body weight) or increasing (clinical score) compared with the predicted level. In this case, use of the value as the endpoint would not result in humane killing of these animals; rather, they would die spontaneously. In most of the animals, loss of body weight was the first indication of a manifest pertussis infection. Body weight could decrease dramatically. Within a few days, animals could lose 20 to 30% of their initial weight because clinically infected animals refuse to eat or drink. Cachexia and dehydration are therefore prominent clinical signs, and it might be postulated that death is due to these phenomena rather than to the pertussis infection. Providing agar to facilitate liquid uptake did not improve the situation. As is evident in Table 2, a decrease in body weight is not a very useful indicator for a lethal outcome of a pertussis infection. A relatively large number of animals had already died before a substantial weight loss was seen, and some animals with a marked reduction of body weight survived the observation period. This finding is in accordance with the conclusion of Reim and Hulin (1998) and Morton (1999), that weight loss alone cannot usually be a criterion for euthanasia but rather should be considered along with other criteria. However, we do not recommend weighing inasmuch as the process of handling might be stressful to the clinically ill animals.

A straightforward and accurate criterion for early euthanasia appears to be the loss of muscular coordination (score 3), a clinical sign that can easily be distinguished from the other signs. Almost all animals showing this clinical sign died within 1 to 3 days after the loss of muscular coordination was noted. Consequently, killing these animals at the onset of this clinical sign would have reduced distress substantially. The total reduction in distress for the animals in this study would have been 23 to 53 days, using clinical score 3 instead of death (Table 3).

A manifest pertussis infection resulted in a decreased body temperature. The decrease generally started about 4 days after challenge. Body temperature levels less than 34°C were observed even in animals not in a moribund state. For analyzing the data, four cut-off points for body temperature were used: <35.5°C, <35.0°C, <34.5°C, and <34.0°C. As is shown in Table 2, a body temperature less than 34.5°C was a very accurate predictor of the lethal progress of the disease. The decrease, not an increase, in body temperature is in accordance with previous reports on infection models (Kort et al. 1999; Soothill et al. 1992; Thurting and Schonbaum 1994). This conclusion might seem contradictory to the observation that animals prefer to lie separately, a behavior that increases

### Table 1 Clinical signs of pertussis infection in laboratory animals

<table>
<thead>
<tr>
<th>Stage</th>
<th>Clinical signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No abnormalities</td>
</tr>
<tr>
<td>1</td>
<td>Cardinal clinical signs</td>
</tr>
<tr>
<td></td>
<td>Less alert</td>
</tr>
<tr>
<td></td>
<td>Dull fur</td>
</tr>
<tr>
<td></td>
<td>Piloerection, especially around the neck</td>
</tr>
<tr>
<td></td>
<td>Clinical signs that might occasionally occur</td>
</tr>
<tr>
<td></td>
<td>Vocalization</td>
</tr>
<tr>
<td></td>
<td>Poor grooming</td>
</tr>
<tr>
<td></td>
<td>Nose and eye discharge</td>
</tr>
<tr>
<td>2</td>
<td>Inactivity, hunched back posture, piloerection over the entire body, social isolation, no food and water intake, dehydration, signs of emaciation</td>
</tr>
<tr>
<td>3</td>
<td>Cardinal clinical signs</td>
</tr>
<tr>
<td></td>
<td>The same as for stage 2, but animals also show disturbed locomotor activity. Animals fall easily after being pushed but are still able to get up again.</td>
</tr>
<tr>
<td></td>
<td>Clinical sign that might occur</td>
</tr>
<tr>
<td></td>
<td>Hind legs paralyzed</td>
</tr>
<tr>
<td>4</td>
<td>Tonic-clonic convulsions, comatose</td>
</tr>
</tbody>
</table>

---

Table 2 Calculated false-positive and false-negative results resulting from various endpoint criteria used for humane killing

<table>
<thead>
<tr>
<th>Humane endpoint</th>
<th>False-positive a.m.</th>
<th>(n=46)c</th>
<th>% a.m.</th>
<th>% p.m.</th>
<th>False-negative d a.m.</th>
<th>(n=30)e</th>
<th>% a.m.</th>
<th>% p.m.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.5-35.0°C</td>
<td>9</td>
<td>4</td>
<td>19.5</td>
<td>8.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>35.0-34.5°C</td>
<td>2</td>
<td>3</td>
<td>4.3</td>
<td>6.5</td>
<td>2</td>
<td>0</td>
<td>6.7</td>
<td>0</td>
</tr>
<tr>
<td>34.5-34.0°C</td>
<td>0</td>
<td>1</td>
<td>2.2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>6.7</td>
<td>0</td>
</tr>
<tr>
<td>≤34.0°C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>13.3</td>
<td>0</td>
</tr>
<tr>
<td>Body weight loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-30%</td>
<td>21</td>
<td>17</td>
<td>45.6</td>
<td>37.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30-40%</td>
<td>10</td>
<td>7</td>
<td>21.7</td>
<td>15.2</td>
<td>6</td>
<td>2</td>
<td>20</td>
<td>6.7</td>
</tr>
<tr>
<td>40-50%</td>
<td>1</td>
<td>0</td>
<td>2.2</td>
<td>0</td>
<td>25</td>
<td>15</td>
<td>83.3</td>
<td>50</td>
</tr>
<tr>
<td>&gt;50%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>28</td>
<td>100</td>
<td>93.3</td>
</tr>
<tr>
<td>Clinical score</td>
<td>1</td>
<td>22</td>
<td>37</td>
<td>47.8</td>
<td>80.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>13</td>
<td>23.9</td>
<td>28.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2.17</td>
<td>4.3</td>
<td>2</td>
<td>1</td>
<td>6.7</td>
<td>2.2</td>
</tr>
<tr>
<td>4</td>
<td>1*</td>
<td>1</td>
<td>0</td>
<td>2.2</td>
<td>9</td>
<td>7</td>
<td>30</td>
<td>15.2</td>
</tr>
</tbody>
</table>


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

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.

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.

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.

Probably invalid data.

Table 3 Reduction of distress (days) using clinical score 3 as the endpoint. The number of days that the animal would have survived without euthanasia is given.

<table>
<thead>
<tr>
<th>Reduction (days)</th>
<th>0/1</th>
<th>1/2</th>
<th>2/3</th>
<th>3/4</th>
<th>4/5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day of euthanasia after immunization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

heat loss. Watkinson and Gordon (1993) suggest that hypoesthesia as induced by autonomic and behavioral thermoregulatory mechanisms might be beneficial for survival. Indeed, it was shown that greater proportions of these hypothermic mice die if they are moved to warmer areas (Klein et al. 1992).

Validation

Although body temperature was considered to be an attractive endpoint because of its objectivity, we decided to validate first only the selected clinical sign (loss of locomotor coordination). Our reason was that transponders, although reusable (Kort et al. 1999), are expensive (at the time of the study approximately US$10 per transponder), which might be an argument for laboratories not to use this endpoint. For an experiment of at least 140 animals, the cost will be US$1400, which is about 80% of the total costs (about 1700 US$ without the use of transponders) of a whole cell pertussis vaccine potency test. Our validation study is now complete, and other laboratories will soon be invited to confirm the use of clinical signs as an alternative to the lethal endpoint.
Discussion

Death as an endpoint in immunization-challenge procedures for the potency testing of vaccines is attractive to many because it is objective, is not time-consuming, and can be monitored by nonskilled staff. This attractiveness was confirmed by an informal survey of vaccine manufacturers and national control laboratories on the use of humane endpoints, which revealed that almost all laboratories adhered to death as the endpoint indicated by the test guidelines (Hendriksen et al. 1999).

As a result of the evolving process of the public's concern over the use of laboratory animals, protocols based on lethal endpoints, for example, are now being challenged. Some people believe that based on moral and legal considerations, alternatives for lethal endpoints must be found. In addition, regulatory authorities are now more aware of animal welfare issues, as is apparent from the increased number of meetings and workshops organized on this issue (e.g., Pharmacopoeia European 1998). Many studies have been initiated to search for alternative methods, with emphasis on the introduction of in vitro techniques. Until now, progress has been limited, and it is likely that a number of vaccine potency tests based on a challenge procedure will continue to be specified in the existing regulations. Therefore, refinement, including humane endpoints, will remain as important and as urgent as replacement (Burch 1995).

We believe that regulatory authorities are likely to accept the use of humane endpoints, on the condition that these endpoints accurately predict imminent death. For example, the European Pharmacopoeia first considered the incorporation of humane endpoints when the monograph on pertussis vaccine was last revised; however, it later decided not to do so because, as they said (Castle 1999), "it was really a question of training of technicians who carry out the test to recognize moribund animals." They continued, "this does not exclude the use of humane endpoints but it puts the onus on the analyst of demonstrating its equivalence."

This policy challenges some conditions that are essential in the acceptance of humane endpoints. First, their acceptance must be based on the same principles of validation as those underlying the validation of in vitro methods, that is, the demonstration of their relevance (equivalence) and reliability (intra- and interlaboratory variation) (Balls et al. 1995). Validation is a tedious and expensive process, which might explain why vaccine manufacturers are not very eager to evaluate humane endpoints. Second, vaccine potency tests require extensive standardization. The introduction of humane endpoints such as clinical signs might conflict with this principle. Therefore, if possible, preference should be given to objective endpoints such as body weight or body temperature. In addition, training and education of staff might be needed, for example, in observing the animals and making accurate assessments of their conditions (Oifert 1995). The production of videos or interactive CD-ROMs for each specific vaccine potency test could be helpful in revealing the disease progress and its cardinal signs and in transferring the alternative endpoints to other laboratories. Training and education also involve attitude. Results of the survey described above confirm that most laboratories observe the animals only once a day. This frequency might be adequate when lethality is used as an endpoint, but it certainly is not sufficient when humane endpoints are used. Montgomery (1990) suggests observing animals twice daily, and more during critical times. However, daily observations are often separated by as much as 8 hr, leaving a relatively long period during the night. We therefore suggest introducing "night biotechnicians" in the laboratory (see Hendriksen et al. 1997).

One of the first studies on the use of humane endpoints in vaccine potency testing was that of Mussett and Sheffield (1976), who presented the results of a collaborative study on paralysis in tetanus vaccine potency testing as an alternative to lethality. This clinical sign was accepted by the regulatory authorities and included in the monographs on human and veterinary tetanus toxoid vaccines. Since then, there have been no reported studies until recently, when results were published on the evaluation of humane endpoints (clinical signs, body temperature, and body weight) in rabies vaccine (Cussler et al. 1999), erysipelas vaccine (Johannes et al. 1999), and whole cell pertussis vaccine potency testing (Hendriksen et al. 1999).

Future studies that include other vaccine potency tests are needed. We recommend the allocation of budgeted funds for these future studies and suggest that regulatory authorities and/or vaccine manufacturers might fund such studies at least in part. However, results of the survey described above lead us to believe that even after validation and acceptance, vaccine quality laboratories will be rather reluctant to implement humane endpoints in their test protocols. Even in the case of the tetanus potency test, for which the European Pharmacopoeia offers the choice between lethality and paralysis, most laboratories have chosen lethality. Therefore, we believe that a revised regulatory policy that allows implementation of humane endpoints is needed. This policy should incorporate validated humane endpoints in the official testing requirements, not as alternatives to lethality but as the only endpoints allowed. We further believe the introduction of humane endpoints contributes to the refinement of animal tests that involve severe distress and constitutes not an ultimate goal but instead, an intermediate step in the process of replacement with in vitro alternatives.

New and sophisticated vaccine production strategies such as the use of recombinant DNA techniques are now being studied. Introduction of these techniques will enable the production of well-defined and consistent products. These new techniques will allow for a shift from the use of animals for routine quality control purposes to research and development and, consequently, a reduction in the number of animals needed.
References


Parish HJ. 1965 History of Immunization. London: Livingstone Ltd.
Pharmacopoeia European. 1998. Alternatives to animal tests in the develop-
ment and control of Erysipelas vaccines. Pharmeuropa, BIO 98:1-100.
Russell WMS, Burch RL. 1959. The Principles of Humane Experimental
Technique. London: Methuen & Co. Ltd.
Soothill JS, Morton DB, Ahma A. 1992. The HIDS0 (hypothermia-inducing
dose 50): An alternative to the LD50 for measurement of bacterial viru-
Thuring JWGM, Schonbaum E. 1994. Hypothermia and social behaviour in
Toth LA, Tolley EA, Krueger JM. 1993. Sleep as a prognostic indicator
Code of Federal Regulations. Title 9, Animals and Animal Products.
Beltsville, MD: USDA.
Van der Ark A, Van Straaten-van de Kapelle I, Akkermans A, Hendriksen
CFM, Van der Donk. 1994. Development of pertussis serological po-
tency test: Serological assessment of antibody response induced by
whole cell vaccine as an alternative to the mouse protection in an intrac-
erebral challenge model. Biologicals 22:233-244.
Van Loveren H, Gianotten N, Hendriksen CFM, Schuurman HJ, Van der
Watkinson WP, Gordon CJ. 1993. Caveats regarding the use of the labora-
tory rat as a model for acute toxicological studies: Modulation of toxic
response via physiological and behavioral mechanisms. Toxicology
81:15-31.
Weisser K, Hechler U. 1997. Animal Welfare Aspects in the Quality Con-
trol of Immunobiologics. A Critical Evaluation of the Animal Tests in
Pharmacopoeial Monographs. Nottingham: FRAME.
for Testing of Vaccines used in the WHO Expanded Programme on