

## Original Research

# Predictive Observation-Based Endpoint Criteria for Mice Receiving Total Body Irradiation

Elizabeth A Nunamaker,<sup>1,\*</sup> Robert J Anderson,<sup>2</sup> James E Artwohl,<sup>1</sup> Alexander V Lyubimov,<sup>3,4</sup> and Jeffrey D Fortman<sup>1</sup>

Total body irradiation of mice is a commonly used research technique; however, humane endpoints have not been clearly identified. This situation has led to the inconsistent use of various endpoints, including death. To address this issue, we refined a cageside observation-based scoring system specifically for mice receiving total body irradiation. Male and female C57BL/6 mice (age, 8 wk) received 1 of 3 doses of radiation from 1 of 2 different radiation sources and were observed for progression of clinical signs. All mice were scored individually by using cageside observations of their body posture (score, 0 to 3), eye appearance (0 to 3), and activity level (0 to 3). Retrospective analysis of the observation score data indicated that death could be predicted accurately with total scores of 7 or greater, and observation scores were consistent between observers. This scoring system can be used to increase the consistent use of endpoint criteria in total body murine irradiation studies and ultimately to improve animal welfare.

**Abbreviations:** TBI, total body irradiation; ARS, acute radiation syndrome.

Total body irradiation (TBI) of mice is a widely used technique with numerous applications. Mice undergoing TBI followed by stem cell transplant are a model of engraftment kinetics<sup>13,14,17,18</sup> and graft-versus-host disease.<sup>10,28</sup> Mice may undergo multiple rounds of TBI followed by gene transfer to create chimeras to study various diseases.<sup>9,11,19,36</sup> More basically, the biologic effects of radiation in mice are a useful model for the various manifestations of human radiation sickness<sup>15,29,35,37</sup> and can be used to develop and evaluate treatments.<sup>5,23,37</sup>

Despite the widespread irradiation of mice, studies typically do not report criteria used for humane euthanasia or the mortality observed due to irradiation with failure of the graft, transplant, or treatment. This practice has resulted in the frequent and widespread use of death or the moribund state as the experimental endpoint.<sup>5,6,35</sup> The moribund condition, an unresponsive and immobile animal, is a commonly used endpoint for a variety of research protocols associated with high mortality or progressive and severe disease states.<sup>32,33</sup> Using the moribund criterion requires the animal to progress through all potential phases of pain and distress associated with the chosen model to a near-death state before the animal is euthanized.<sup>25</sup>

The ability to predict death with a high probability and high accuracy for mice receiving TBI would allow preemptive euthanasia. This action would ameliorate terminal pain and distress associated with acute radiation syndrome (ARS) and markedly improve animal welfare yet maintain scientific integrity in accordance with the recommendation of the *Guide for the Care and Use*

*of Laboratory Animals*.<sup>12</sup> Although objective endpoint criteria are generally preferable, this option is not always possible. For mice receiving TBI, neither weight loss<sup>26</sup> nor decreased food and water consumption<sup>29</sup> are good predictors of death, because many animals that survive demonstrate marked weight loss and reduced consumption. Furthermore, mice receiving TBI should be handled minimally to avoid increasing mortality.<sup>15,31</sup> As an alternative approach to objective endpoint criteria, subjective behavioral criteria may be used, providing that they are adequately described, monitored, and applied so that trends can be documented.<sup>24,34</sup>

The current study sought to refine an observation-based scoring system to assess the health status of irradiated mice and determine the broad applicability of endpoints across radiation doses, radiation sources, and animal sexes. It was performed in conjunction with an IACUC-approved study designed to establish survivability curves for mice undergoing TBI with no other experimental manipulation. Male and female C57BL/6 mice that received 1 of 3 TBI doses from 1 of 2 different radiation sources were evaluated by using daily cageside observational scoring of body posture, activity level, and eye appearance. The results were analyzed to verify consistency between observers and identify the predictive nature of these criteria of impending death, with the goal of establishing useful endpoint criteria for all mice receiving TBI.

## Materials and Methods

**Animals.** Mice used during this study were maintained in accordance with the *Guide for the Care and Use of Laboratory Animals*<sup>12</sup> at the University of Illinois at Chicago (Chicago, IL), an AAALAC-accredited institution. All procedures were reviewed and approved by the University of Illinois at Chicago Animal Care Committee. Male and female C57BL/6J (age, 6 wk;  $n = 120$

Received: 06 Jan 2013. Revision requested: 04 Feb 2013. Accepted: 16 Mar 2013.

<sup>1</sup>Biologic Resources Laboratory, <sup>2</sup>School of Public Health, Division of Epidemiology and Biostatistics, <sup>3</sup>Toxicology Research Laboratory, and <sup>4</sup>Department of Pharmacology, College of Medicine, University of Illinois at Chicago, Chicago, Illinois.

\*Corresponding author. Email: lizzynum@gmail.com

<sup>†</sup>Current affiliation: MED Institute, West Lafayette, Indiana.

for each sex) mice were purchased from The Jackson Laboratory (Bar Harbor, ME). Mice were maintained in facilities in which dirty-bedding contact-sentinel mice tested negative on a quarterly basis for Sendai virus, pneumonia virus of mice, mouse hepatitis virus, minute virus of mice, mouse parvovirus, Theiler murine encephalomyelitis virus, reovirus, rotavirus, mouse adenovirus, polyoma virus, K virus, mouse cytomegalovirus, mouse thymic virus, lymphocytic choriomeningitis virus, hantavirus, ectromelia virus, lactate dehydrogenase elevating virus, mouse norovirus, *Mycoplasma pulmonis*, and *Helicobacter* spp. In addition, sentinel mice were free of helminth and external parasites. On arrival, mice were housed 5 per cage in static autoclaved (sterilized) polysulfone microisolation cages (Ancare, Bellmore, NY) with irradiated diet (no. 7912, Harlan Teklad, Madison, WI), HCl-acidified municipal water in bottles, autoclaved hardwood bedding (Sani-chip, Harlan Teklad) and a nesting pad (Ancare) and on a 14:10-h light:dark cycle. The room temperatures and humidity were maintained at 20 to 25 °C and 30% to 70%, respectively. Mice were acclimated for 2 wk prior to irradiation. Once the mice were irradiated, they were housed individually with the same housing specifications to ensure that no clinical signs or deaths would be due to fighting. At the time of total body irradiation (TBI), mice were 8 wk old and weighed (mean  $\pm$  1 SD) 23.7  $\pm$  1.0 g (male mice) and 18.1  $\pm$  1.0 g (female mice).

Mice were weighed on days 1, 7, 11, 18, and 25 after TBI and immediately before euthanasia once they were identified as moribund. Body weights were converted to percentage of the baseline body weight for individual mice. The maximal percentage weight gains and losses were identified for each animal. Mice that received the same radiation dose, regardless of sex or radiation source, were combined for comparison.

Animals were irradiated with either a 6-MV LINAC photon source (model no. EX21, Varian Medical Systems, Palo Alto, CA) or Cs<sup>137</sup> irradiator (model no. 143-68, JL Shepherd and Associates, San Fernando, CA). Forty mice of each sex received a uniform, total-body, total midline tissue dose of 770, 820, or 870 cGy. For each sex, half of the number of mice was irradiated by using the LINAC source and the other half with the Cs<sup>137</sup> source. LINAC irradiation was administered at a dose rate of 80  $\pm$  2.5 cGy/min, and the Cs<sup>137</sup> irradiation used a dose rate of 267 cGy/min, until the targeted total-body midline tissue dose was reached.

Any mouse that became moribund (see description following) during the study was euthanized by CO<sub>2</sub> asphyxiation followed by cervical dislocation. All mice that survived to day 31 were euthanized in this manner also.

**Cageside observation scoring system and euthanasia criteria.** This scoring system was developed previously to assess TBI mice,<sup>26</sup> adapted from other scoring systems developed to assess pain and distress in rodents,<sup>22,24,27</sup> and based on behaviors exhibited in previous mouse TBI studies. Rodent cages were removed from the rack to improve visualization and to stimulate the mouse to move around the cage, but cages were not opened at any point during the scoring process. Mice received a score of 0 to 3 for each of the criteria of posture, eye appearance, and activity level. Posture was scored based on a hunched appearance (Figure 1). A score of 0 indicated normal body posture; 1 indicated a slightly hunched posture; 2 indicated a moderately hunched posture; and 3 indicated a severely hunched posture. Eye appearance was scored based on the level of openness (Figure 2). A score of 0 indicated eyes that were open more than 75%; 1 indicated

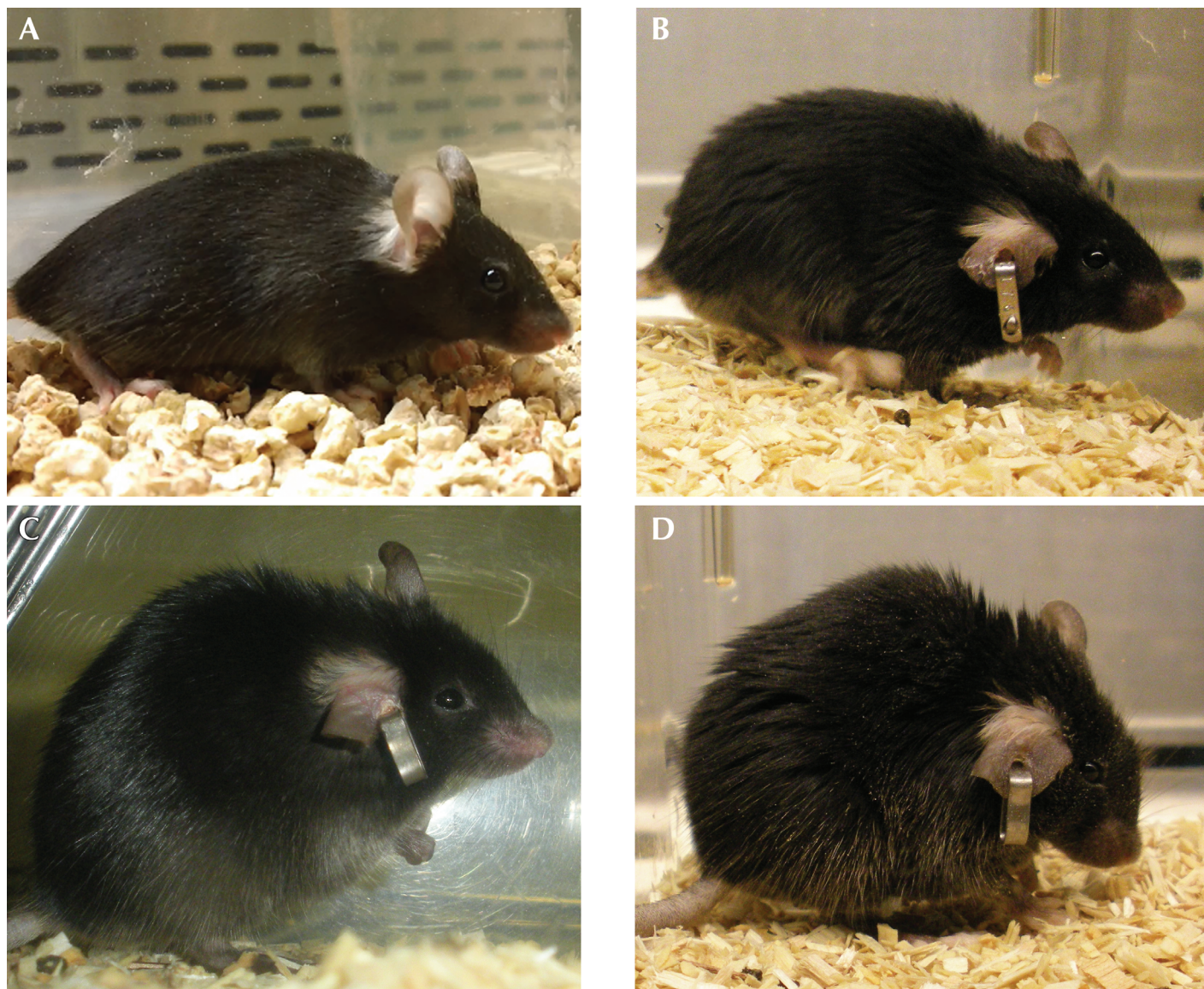
eyes that were 50% to 75% open; 2 indicated eyes 25% to 49% open; and 3 indicated eyes open less than 25%. It was important to prescreen C57BL/6 mice for anophthalmia and microphthalmia,<sup>16</sup> because these conditions had the potential to affect the score for eye appearance and artificially increase the mouse's score. Activity level was scored based on the amount the mouse moved in the cage. A score of 0 indicated an animal that moved around the cage normally and was very active. A score of 1 indicated a slightly reduced activity level or a mild gait abnormality. A score of 2 indicated a mouse that was moving very slowly or had a severely altered gait. A score of 3 indicated an animal that was reluctant to move, taking no more than 3 or 4 steps, or did not move at all.

Cageside observations by the veterinary staff and trainer were made in parallel with research staff. The mice were observed on days 1 through 30 after TBI to score the animals. There was a graded observation schedule based on the progression and subsequent remission of acute radiation syndrome clinical signs in all animals receiving TBI. All of the mice were scored by a veterinarian once daily (0700 to 0900) on study days 1 through 6 and 23 through 30 and twice daily on days 7 through 22 (0700 to 0900 and 1430 to 1630). A subset of 60 mice of the 240 total was scored by the trainer, who used the same graded observation schedule as did the veterinary staff. All of the mice were scored once daily (0630 to 0830) for the duration of the study by the research staff. All observers remained blinded to each other's scores at all time points for the duration of the study. During the critical radiation syndrome period, when the clinical signs were the most severe (days 7 through 22), all parties met in the evening (1600 to 1800) in the animal rooms to discuss mice with scores greater than 6 by any observer. This meeting provided an opportunity for all 3 groups to observe mice together at the same time and verbally verify agreement on observational scores.

The cageside observation score was used as euthanasia criteria. According to our previous experience with this scoring system,<sup>26</sup> initial euthanasia criteria was set as a score of 7 by any observer. Due to the unknown potential for score disparity between observers due to interobserver variability, the euthanasia criterion was increased to a score of 9 on study day 10; mice with a score of 9 had scores of 3 for all of the criteria and were considered moribund. Increasing the euthanasia criteria from a score of 7 to 9 early in the study was expected to decrease the potential for euthanasia of mice that might have otherwise gone on to survive. Mice meeting the euthanasia criteria were euthanized after all animals in the room had been scored.

**Training of observers.** Veterinary and research staff were trained by a single trainer, who developed the observation scoring system for mice receiving TBI. Training was completed by using a 30-min training session consisting of images and videos of mice displaying the full range of observable clinical signs and clinical scores (0 through 9). Colored copies of the scoring criteria images for posture (Figure 1) and eye openness (Figure 2) were made available to all observers during the scoring period. In addition, independent 30-min training sessions with the trainer and live animals were available on request.

**Statistics.** All statistical calculations were performed using commercially available software including SAS (version 9.2, SAS Institute, Cary, NC), STATA (version 11, StataCorp, College Station, TX), and MatLab (MathWorks, Natick, MA).



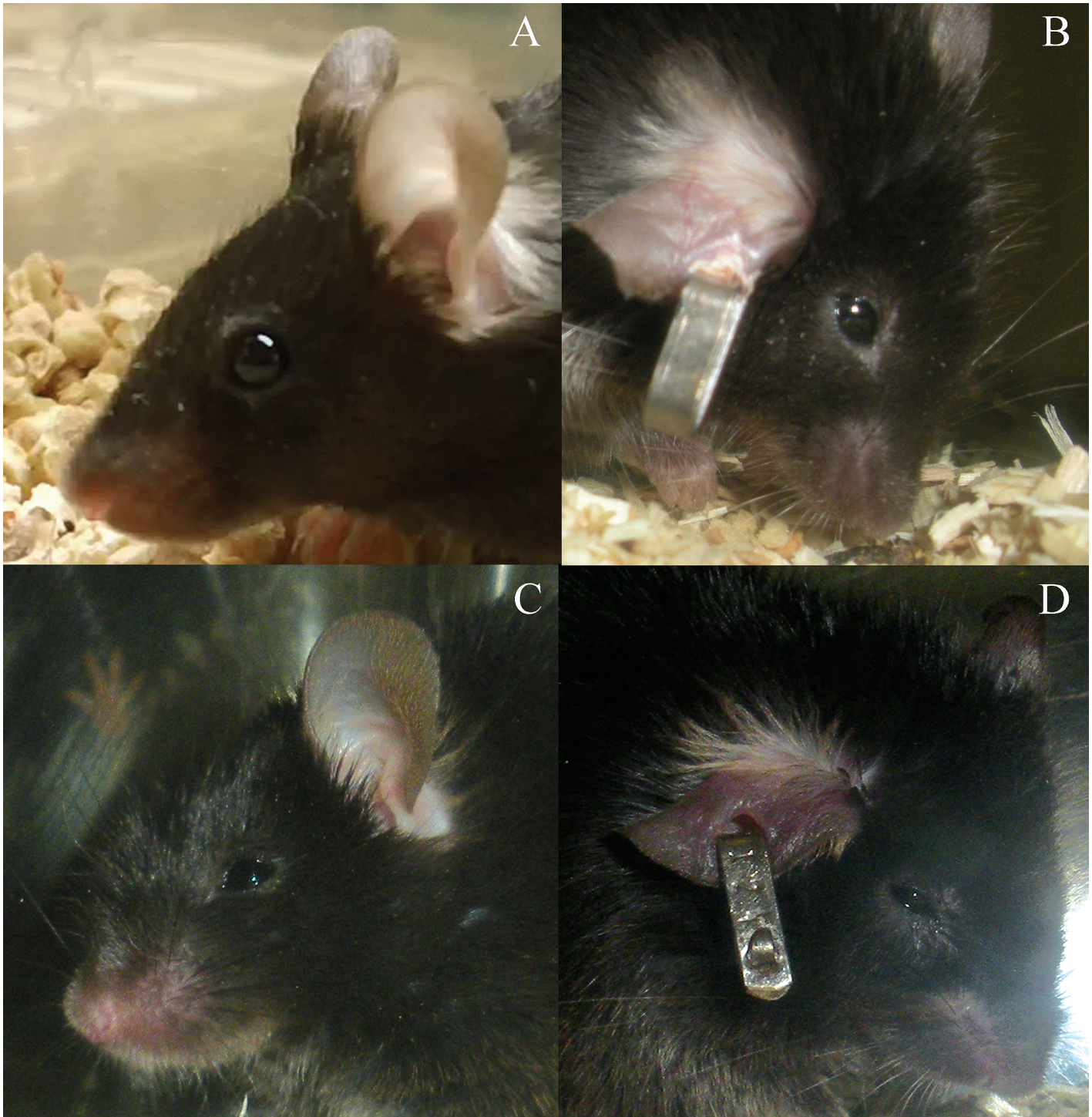
**Figure 1.** Scoring of posture according to cageside observations. (A) The normal mouse posture received a score of 0. (B) A slightly hunched posture yielded a score of 1. (C) A mouse that was moderately to severely hunched but was still able to rear up or stretch out to access food and water received a score of 2. (D) A mouse that was severely hunched and unable to rear up was scored as 3. This image is reprinted from reference 26 with permission.

From the cageside observation scoring data for each mouse, the first day on which each mouse scored a 5, 6, 7, 8, and 9 was identified. Subsequently, the number of days between the time that the mouse was first achieved a particular score and its death (whether found dead or euthanized) or censoring event (that is, end of the study) was calculated. Mice that skipped a particular score had the time of the observed higher score designated for both the observed score and the skipped score. For each of these starting score events, the median (that is, 50th percentile) survival times and the estimated 25th percentile and 75th percentile survival times for all mice, after reaching a specific score, were reported.

One independent morning observation was chosen for each of the 240 mice by using a random number generator, to evaluate interobserver variability. The scores of the research staff, veterinary staff, and trainer were all assessed against each other by creating 3 separate contingency tables. The generalized Stuart–Maxwell

marginal homogeneity test was applied to each contingency table to assess observer bias.<sup>8</sup> Significance was determined as a *P* value of less than 0.05.

The agreement between observer scores was evaluated by calculating the  $\kappa$  coefficient (interobserver variability), which estimates the proportion of concordant measurements and omits those that agree because of chance.<sup>20,21</sup> The unweighted Cohen, weighted Cicchetti–Allison,<sup>3</sup> and weighted Fleiss–Cohen<sup>7</sup>  $\kappa$  coefficients; asymptotic standard error; and 95% confidence intervals were reported. To further assess interobserver variability with respect to being within 1 or 2 units, additional weighted  $\kappa$  coefficients were calculated. Fade 2 assigned a maximal weight of 1 to the main diagonal with a symmetric, monotonic decrease over a 2-unit difference (0.67, 0.33). Similarly, Fade 3 assigned a maximal weight of 1 to the main diagonal with a symmetric, monotonic decrease over a 3-unit difference (0.75, 0.5, 0.25). Near



**Figure 2.** Scoring of eye appearance according to cageside observations. (A) Normal mouse eyes that were open more than 75% received a score of 0. (B) Mouse eyes that were open 50% to 75% yielded a score of 1. (C) Mice whose eyes were open 25% to 49% received the score of 2. (D) Mouse eyes that were open less than 25% were scored as 3. This image is reprinted from reference 26 with permission.

1 assigned maximal weight to the main diagonal and within 1 unit, and Near 2 assigned maximal weight to the main diagonal and within 2 units. All other differences were weighted a 0, indicating disagreement. The  $\kappa$  coefficients and asymptotic standard errors for each weighted method were reported. The strength of the agreement was interpreted as follows: less than 0.2, poor;

0.21 to 0.40, fair; 0.41 to 0.6, moderate; 0.61 to 0.8, substantial; and greater than 0.81, almost perfect.<sup>21</sup>

Body weights were converted to percent of baseline (day 0) body weight to determine the percentage lost or gained during the 30-d study period and immediately preceding death. Values for percentage change from baseline weight were grouped

according to radiation dose, irrespective of animal sex and radiation source. Maximal percentage body weight losses and gains between mice that lived and died at each radiation dose were compared by using the Mann–Whitney *U* test. Differences between percentage body weight values were defined as significant at a *P* value of less than 0.05.

## Results

Male and female C57BL/6J mice received TBI at doses of 770, 820, or 870 cGy from either a 6 MV LINAC photon source or a Cs<sup>137</sup> irradiator and demonstrated survival rates (% alive after 30 d) that decreased with increasing radiation doses for both sexes and radiation sources. Of the 240 mice on study, 183 were either found dead (*n* = 16) or were euthanized (*n* = 167) over the 30-d study period (Table 1). Female mice irradiated with the Cs<sup>137</sup> irradiator had survival rates of 60%, 15%, and 5% at the doses of 770, 820, and 870 cGy, respectively (Figure 3 A). Female mice irradiated with the LINAC had survival rates of 70%, 10%, and 0% at total doses of 770, 820, and 870 cGy, respectively (Figure 3 B). Male mice irradiated with the Cs<sup>137</sup> irradiator had survival rates of 40%, 10%, and 0% at doses of 770, 820, and 870 cGy, respectively (Figure 3 C). Male mice irradiated with the LINAC had survival rates of 50%, 15%, and 10% at total doses of 770, 820, and 870 cGy, respectively (Figure 3 D).

Increasing total scores were positively correlated with animal death. Over the study period, 167 mice received a score of 5, 154 received a score of 6, 145 received a score of 7, and 144 received a score of 8 (Table 2). In addition, 13 mice had a maximal score of 5, 9 animals had a maximal score of 6, and 1 animal had a maximal score of 7; 67 mice never scored higher than 4 and were alive at the end of the study. Mice that were scored as 5, 6, 7, or 8 by the veterinarian had mortality rates of 86.2%, 93.5%, 99.3%, or 100%, respectively. Six mice died before reaching an observation score of 5, and 39 mice were not included in the median survival analysis because they were euthanized prior to reaching a score of 9 (Table 1). The overall median survival time associated with scores of 5, 6, 7, and 8 were 4.0, 2.5, 1.0, and 0.0 d, respectively, but survival times were variable due to both radiation dose and mouse sex (Table 2). A single mouse that scored a 9 died before it could be euthanized, and all 129 mice that were scored as 9 had a survival time of 0 d. A single mouse had unilateral microphthalmia, which was noted before the study started, and the normal eye was used for all observation scores.

Mice were evaluated by 3 observers to determine interobserver variability. Observations were not made simultaneously but were made within the same 2-h time frame. The agreement between observer scores was evaluated by calculating the  $\kappa$  coefficient, which estimates the proportion of concordant measurements and omits those that agree because of chance.<sup>21</sup> The unweighted Cohen  $\kappa$  was ‘fair’ between the researcher and veterinarian, researcher and trainer, and veterinarian and trainer (0.263, 0.373, and 0.336, respectively; Table 3). When a standard weighted  $\kappa$  coefficient was calculated for these same 3 comparisons, the strength of agreement increased to ‘substantial’ (0.752, 0.749, and 0.748; Cicchetti–Allison) or ‘almost perfect’ (0.933, 0.906, and 0.911; Fleiss–Cohen).

To more closely assess interobserver variability with regard to being within 1 or 2 units, additional weighted  $\kappa$  coefficients were calculated (Table 3). Fade 2 showed ‘substantial’ agreement for the 3 comparisons (0.603, 0.619, and 0.618), as did Fade 3 (0.670, 0.676, and 0.675). Near 1 exhibited ‘substantial’ agreement for the

**Table 1.** Highest last observation score by any observer for all mice that died or were euthanized

Score at last observation	No. of mice euthanized	No. of mice found dead
≤ 5	0	6
6	1	5
7	8	3
8	30	1
9	128	1
Total	167	16

3 comparisons (0.774, 0.731, and 0.755), whereas Near 2 assigned ‘almost perfect’ agreement (0.967, 0.877, and 0.907).

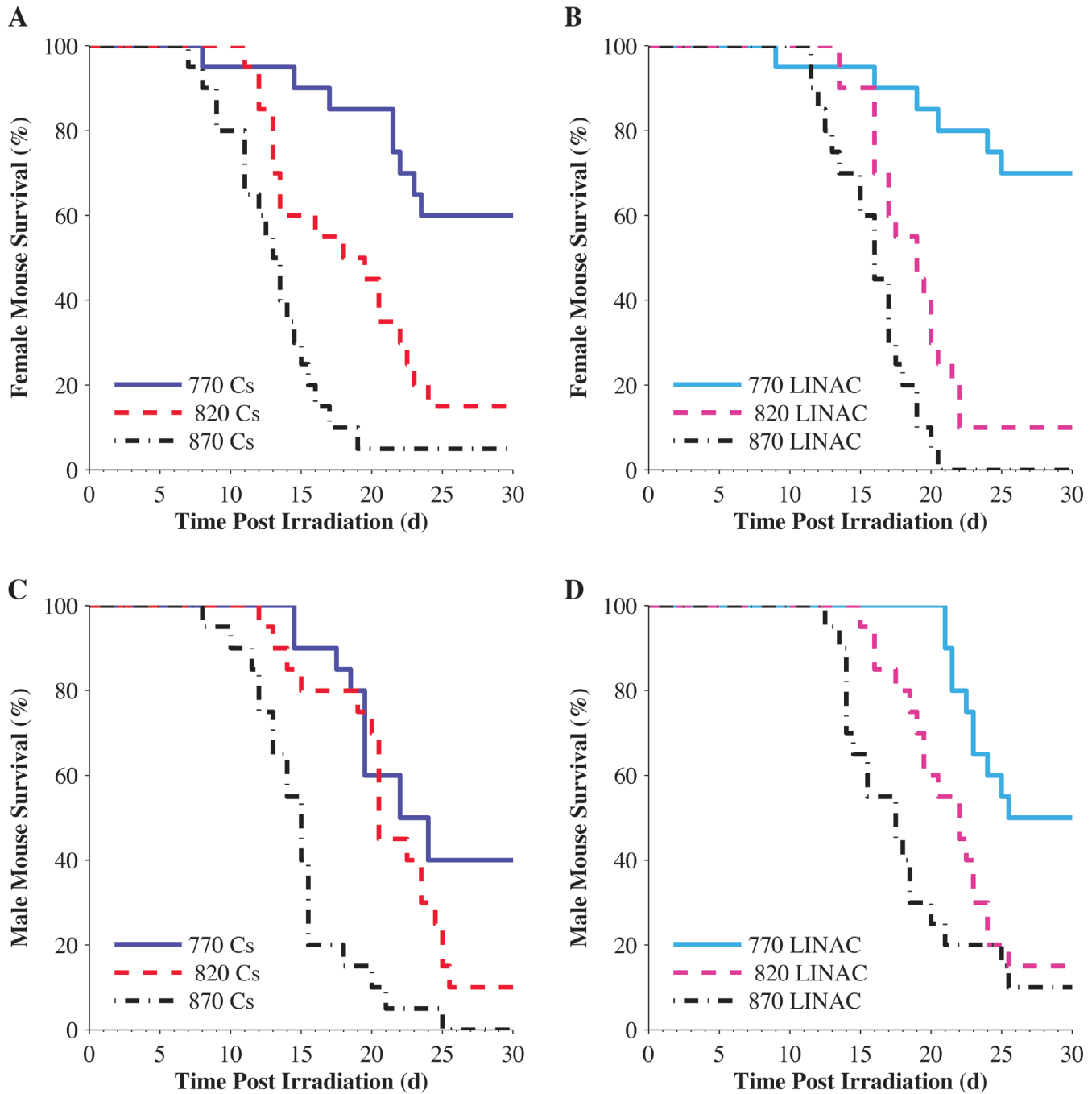
By using the generalized Stuart–Maxwell marginal homogeneity test for the 10 score categories, there was a significant (*P* < 0.0001, *n* = 240) difference between the principal investigator’s staff (that is, researchers) and veterinarians regarding what scores were used. However, the rate of use of the score categories was not significantly different either when comparing the researchers and trainer (*P* = 0.2478, *n* = 60) or veterinarians and trainer (*P* = 0.2361, *n* = 60).

Mice that died or were euthanized during the 30-d period after TBI lost more body weight than did those that lived (Figure 4). At the 770-cGy TBI dose, the 46 mice that lived had a median maximal weight loss of 7.0% during the study compared with 25.5% for the 34 mice that died (*P* < 0.001, Figure 4 A). At the 820-cGy TBI dose, the 11 mice that lived had a median maximal weight loss of 10.0% during the study compared with 24.9% for the 69 animals that died (*P* < 0.001, Figure 4 B). At the 870-cGy TBI dose, the 3 mice that lived had a median maximal weight loss of 3.9% during the study compared with 23.6% for the 77 mice that died (*P* < 0.05, Figure 4 C).

Mice that died during the 30-d period after TBI gained less body weight than did those that lived. At the 770-cGy TBI dose, the 46 mice that lived had a median maximal weight gain of 6.1% during the study compared with 0.4% for the 34 animals that died (*P* < 0.001). At the 820-cGy TBI dose, the 11 mice that lived had a median maximal weight gain of 2.7% during the study compared with 0.8% for the 69 mice that died (*P* < 0.05). At the 870-cGy TBI dose, the 3 mice that lived had a median maximal weight gain of 3.9% during the study compared with 0.5% for the 77 mice that died (*P* = 0.2144).

## Discussion

Once morbidity criteria have been identified as potential surrogates for mortality, they must be verified for applicability and consistency.<sup>4,34</sup> Although behavioral monitoring is an unobtrusive way to monitor pain and distress in rodents,<sup>25</sup> careful documentation through score assignment to clinical signs provides a helpful metric for tracking the deterioration or resolution of an animal’s clinical condition. Mice that receive TBI will either demonstrate no obvious behavioral changes, or they will respond behaviorally and subsequently either recover or die,<sup>15,26,30,35</sup> demonstrating the importance of tracking these animals’ clinical conditions. Although we previously demonstrated that scoring mouse posture, activity level, and eye appearance could be collectively used to assess the progression of ARS, the broad applicability was ques-



**Figure 3.** Kaplan–Meier survival curves for all mice ( $n = 20$  mice for each combination of sex, dose, and radiation source). (A) Survival of female mice after TBI doses of 770, 820, or 870 cGy from a  $Cs^{137}$  source. (B) Survival of female mice after TBI doses of 770, 820, or 870 cGy from a LINAC source. (C) Survival of male mice after TBI doses of 770, 820, or 870 cGy from a  $Cs^{137}$  source. (D) Survival of male mice after TBI doses of 770, 820, or 870 cGy from a LINAC source.

tionable due to the unknowns of interobserver variability or sex-, radiation-source-, or TBI-dose-associated differences.<sup>26</sup>

Assigning scores to a series of observations of clinical signs provides a mechanism for tracking the deterioration or resolution of the clinical condition of an animal. In addition, documenting the pattern of the clinical signs for an individual animal facilitates discussion between veterinary and research staff as a study pro-

gresses to assist in decisions about euthanasia of animals. The advantages of the observation-based score system have previously been acknowledged: there is closer observation of the animals by all involved, especially during the critical timeframe; subjective assessments of pain and distress are avoided; evidence-based opinion becomes possible based on the documented and scored clinical signs; limited scoring options lead to an increased consis-

**Table 2.** Survival times of mice after they reached a score of 5, 6, 7, or 8 according to the veterinarian

	Score	<i>n</i>	Died or eutha- nized	Censored	Median survival (d)	25% survival (d)	75% survival (d)
Overall	5	167	144	23	4.0	6.5	2.0
	6	154	144	10	2.5	4.0	1.0
	7	145	144	1	1.0	2.5	0.0
	8	144	144	0	0.0	0.75	0.0
Female 770 cGy	5	17	13	4	3.5	7.0	2.375
	6	14	13	1	2.5	3.5	1.5
	7	13	13	0	1.5	2.5	0.5
	8	13	13	0	0.0	0.5	0.0
Male 770 cGy	5	31	20	11	6.5	12.5	3.625
	6	26	20	6	4.0	6.5	2.5
	7	21	20	1	2.0	3.5	1.38
	8	20	20	0	0.75	1.25	0.5
Female 820 cGy	5	32	30	2	4.0	5.75	2.75
	6	30	30	0	2.75	4.0	1.0
	7	29	30	0	1.5	2.5	1.0
	8	29	30	0	0.25	1.0	0.0
Male 820 cGy	5	34	29	5	5.25	8.0	3.5
	6	31	29	2	3.0	4.0	1.125
	7	29	29	0	1.0	2.125	0.0
	8	29	29	0	0.0	0.625	0.0
Female 870 cGy	5	24	23	1	2.0	4.0	0.5
	6	24	23	1	1.5	2.25	0.0
	7	23	23	0	0.5	1.5	0.0
	8	23	23	0	0.0	0.5	0.0
Male 870 cGy	5	29	29	0	1.5	4.0	0.875
	6	29	29	0	1.0	2.125	0.5
	7	29	29	0	0.5	1.625	0.0
	8	29	29	0	0.0	0.5	0.0

Censored animals were alive at the end of the 30-d period. 25% survival indicates that 75% of the mice died within the reported number of days after attaining the given score. Similarly, 75% survival indicates that 25% of the mice died within the reported number of days after attaining the given score.

tency of scoring; score sheets reveal patterns of deterioration or recovery over time; and the score sheets encourage all involved to observe and recognize normal and abnormal behaviors in response to the experimental parameters.<sup>24</sup> As previously reported, scoring was limited to 3 specific observed clinical signs (body posture, eye appearance, and activity level) and provided guidance to assigning scores for those 3 parameters of 0 (normal) to 3 (severely abnormal).<sup>26</sup>

Observational scoring was predictive of death in all mice at all irradiation doses. Although there was some variation in survival between sex and radiation source at the 3 doses tested (Figure 3), the scoring system was used successfully to assess all irradiated mice (Table 2). All of the mice that received a score of 8 from the veterinarians died or progressed to humane euthanasia criteria (score of 9, moribund). Only a single male mouse in the lowest

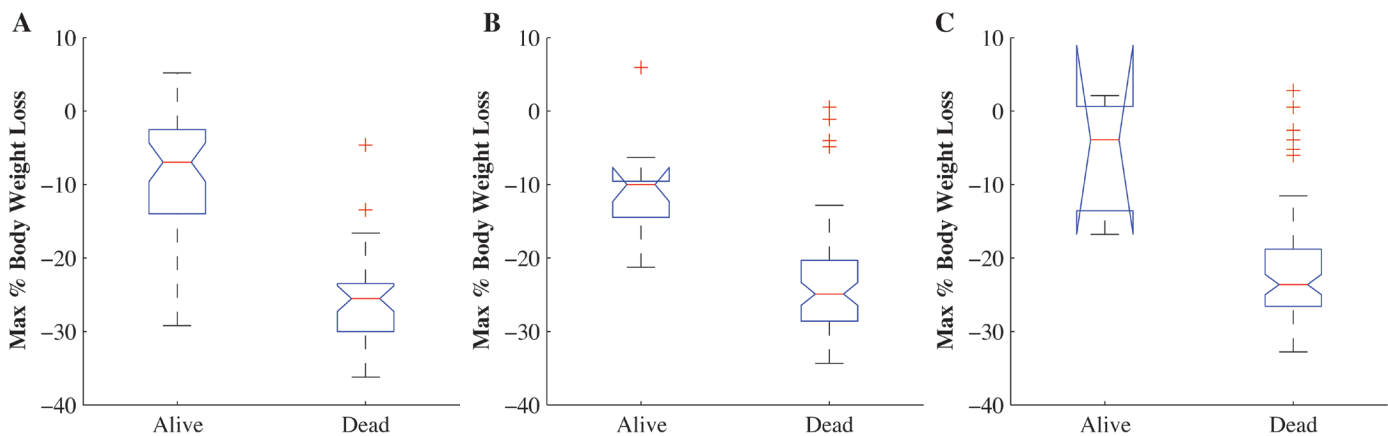
radiation dose group scored a 7 and survived; the remaining 144 animals that received a score of 7 from the veterinarians during the study died or were deemed moribund and euthanized. In the previous study using this scoring system with a radiation dose of 845 cGy delivered by a 6-MV LINAC irradiator, only 86.5% or 93.4% of mice that scored 7 or 8, respectively, died during the 30-d period.<sup>26</sup> This discrepancy in mortality rates associated with scores is most likely due to the antibiotics that the mice on the previous study received in their water; this practice has been demonstrated to prolong life and decrease mortality associated with ARS.<sup>29</sup> However, the difference alternatively may reflect the different vendors of the animals between the previous<sup>26</sup> and current study.

The median survival time of mice receiving TBI decreased as the radiation dose increased. Both male and female mice that re-

**Table 3.**  $\kappa$  coefficients of interobserver variability reported with the corresponding asymptotic standard error (ASE) and 95% confidence interval (CI), where applicable

$\kappa$	Researcher compared with veterinarian		Researcher compared with trainer		Veterinarian compared with trainer	
	$\kappa$ (ASE)	95% CI	$\kappa$ (ASE)	95% CI	$\kappa$ (ASE)	95% CI
Cohen	0.263 (0.034)	0.197, 0.329	0.373 (0.071)	0.235, 0.512	0.336 (0.071)	0.196, 0.475
Cicchetti–Allison	0.752 (0.017)	0.719, 0.785	0.749 (0.044)	0.664, 0.834	0.748 (0.039)	0.672, 0.823
Fleiss–Cohen	0.933 (0.008)	0.918, 0.947	0.906 (0.027)	0.854, 0.959	0.911 (0.024)	0.865, 0.958
Fade 2	0.603 (0.032)	—	0.619 (0.064)	—	0.618 (0.064)	—
Fade 3	0.670 (0.036)	—	0.676 (0.072)	—	0.675 (0.073)	—
Near 1	0.774 (0.043)	—	0.731 (0.084)	—	0.755 (0.084)	—
Near 2	0.967 (0.056)	—	0.877 (0.113)	—	0.907 (0.115)	—

$n = 240$  for researcher compared with veterinarian,  $n = 60$  for researcher compared with trainer, and  $n = 60$  for veterinarian compared with trainer. The strength of the agreement can be interpreted as follows:  $<0.2$ , poor;  $0.21\text{--}0.40$ , fair;  $0.41\text{--}0.6$ , moderate;  $0.61\text{--}0.8$ , substantial; and  $>0.81$ , almost perfect.<sup>21</sup>



**Figure 4.** Maximal body weight loss due to TBI. (A) The mice that died after TBI with 770 cGy ( $n = 34$ ) lost more ( $P < 0.001$ ) weight than did those that lived ( $n = 46$ ). (B) The mice that died after TBI with 820 cGy ( $n = 69$ ) lost more ( $P < 0.001$ ) weight than did those that lived ( $n = 11$ ) and 69 (dead). (C) The mice that died after TBI with 870 cGy ( $n = 77$ ) lost more ( $P < 0.05$ ) weight than did those that lived ( $n = 3$ ). +, outlier.

ceived 770 cGy of radiation lived longer after reaching scores of 6 and greater than did mice irradiated with higher doses (Table 2). This result is consistent with the literature describing the kinetics of ARS and differences between radiation doses.<sup>1,35</sup> Preemptively euthanizing a mouse once an observational score of 7 or 8 is reached has the potential to mitigate animal pain and distress associated with ARS. Euthanasia of animals at a set score of 7 would have minimized pain and distress for 171 mice that progressed through the syndrome and ultimately died or were euthanized; conversely, only a single mouse would have been euthanized prematurely. Furthermore, using a score of 7 to guide euthanasia of the mice would have only missed 11 animals that progressed to death quickly and were scored a 6 or less on their last veterinarian observation. Comparatively, euthanasia of mice at a set score of 8 would have minimized pain and distress for 160 mice that progressed through the syndrome and ultimately died or were euthanized and would have missed only 14 animals that progressed quickly and were scored a 7 or less on their last veterinarian observation. Depending on the type of study, however, the most appropriate score for euthanasia criteria may vary. In animals receiving antibiotics, a score of 8 may be a more appropriate set score for euthanasia, because these animals tend to live longer after TBI,<sup>29</sup> and more animals that score a 7 may recover.<sup>26</sup>

This practice would minimize animal pain and distress associated with ARS yet minimize the number of mice that were euthanized but that might otherwise survive. For each application of this scoring system, a pilot study could be performed to determine the number of animals that live or die after reaching scores of 7 or 8 to determine the best score for euthanasia criteria.

A body weight loss of 15% to 20% is a traditional endpoint criterion for a variety of studies; however, body weight is difficult to include in the evaluation of animals for TBI studies. The correlation between animal handling and increased mortality in TBI studies is well known, and therefore the frequency of weight measurement generally is limited to a maximum of every 2 to 3 d.<sup>15,31</sup> Weight was measured weekly and immediately prior to euthanasia in the current study; consequently, the positive predictive value of either a 15% or 20% weight loss for death was 93.0% or 97.1%, respectively. In our previous study in which animals were weighed weekly, 15% and 20% body weight loss had 80.6% and 84.9% positive predictive value for death.<sup>26</sup> The increased positive predictive value in the current study was expected because the measurement of body weight was prompted by the observational scores that deemed the animals to be moribund and thus already identified for euthanasia. Although weighing animals immediately prior to death increased the predictive value



of the percentage body weight, doing so is more stressful for the mouse than are cageside observations. In the current study, body weights also were less predictive than were observation scores of 7 or 8 (99.3% and 100%, respectively). This finding indicates that the use of body weight alone as an endpoint has less utility than do cageside observation scores; however, 20% weight loss could be considered as a secondary endpoint once a score of 7 has been reached.

Observers were not evaluating the mice at exactly the same time, but relatively good agreement between observers was noted. Identical scores or scores within 1 unit were ideal, but scores within 2 were considered acceptable. The documented trends were useful in monitoring the continued decline or resolution of the clinical condition of individual mice and therefore were valuable to prompt discussion between the research and veterinary staff. All of the observers were blinded to the scores that the others recorded at all time points, but there was daily discussion between the groups about specific mice that were identified as having scores of 7 or greater by any observer. Although not recorded on the score sheets, when the groups met to assess animals in the afternoon and they looked at the animals at the same time, a single score was agreed on, and the determination to wait or to euthanize was made.

The use of  $\kappa$  is a statistically robust metric for evaluating interobserver variability. The  $\kappa$  statistic is the proportion of agreement corrected for chance and has values of 0 for exactly chance agreement and +1 for perfect agreement beyond chance agreement. The Cohen  $\kappa$  is an unweighted statistic that requires perfect agreement to be classified as 'agreement' and is considered to be very conservative.<sup>2</sup> This statistic gives no weight to scores that are within 1 or 2 units of each other. In the current study, the Cohen  $\kappa$  indicated only fair agreement between each set of observers. This finding may be due to the combination of conservativeness of the statistic, to the fact that all mice were not scored at the same time, and to possible interobserver variability.

Alternatively, a weighted  $\kappa$  can be used to assign specific values to the levels of disagreement and can thus be adapted to each situation.<sup>2,7</sup> Two commonly used weighted  $\kappa$  are the Fleiss–Cohen<sup>7</sup> and Cicchetti–Allison<sup>3</sup> statistics. The Fleiss–Cohen  $\kappa$  is calculated based on weights assigned to the entire range of 10 possible scores (0 through 9), with the heaviest weights (that is, 1) assigned to those observations in exact agreement and with rapidly declining scores (in quadratic fashion) for those combinations further away from perfect agreement. The Cicchetti–Allison  $\kappa$  is similarly based on weights assigned to the entire range of 10 possible scores, with the heaviest weights assigned to those observations in exact agreement, but with the weight declining more gradually (in a linear fashion) for those observations further away from perfect agreement. Although both of these weighted  $\kappa$  statistics resulted in substantial or almost-perfect agreement between the different combinations of observers, neither is suited to this scoring system, because the investigators felt that, in this experimental situation, no weight should be given to a score that was more than 2 units away from exact agreement.

To address the shortcomings of these common weighted  $\kappa$  statistics, 2 weighting schemes were evaluated. The Fade  $\kappa$  assigned a weight of +1 to exact agreement and then linearly decreased to 0 over 3- or 4-unit differences in scores, corresponding to the labels Fade 2 and Fade 3. Specifically, Fade 2 assigned a weight of 0.667 to scores within 1 unit of agreement, a weight of 0.333 to scores

within 2 units, and a weight of 0 to all other score discrepancies. The Near  $\kappa$  scheme assigned a weight of +1 to scores of exact agreement and those within 1 or 2 units and a 0 to all other score discrepancies, corresponding to the labels Near 1 and Near 2. All 4 of these statistics resulted in either substantial or almost-perfect agreement between the different observers, but the most clinically relevant statistics seem to be the Fade 2 and Near 1  $\kappa$  statistics; both of these  $\kappa$  indicated that there was substantial agreement between the observers. This finding suggests that there is substantial agreement between observers and that this scoring system could be used by either research staff or veterinary staff independently, after appropriate training, to assess the clinical condition of the mice. Furthermore, the scoring system can be used to establish endpoint criteria for mice receiving TBI, such that animals could preemptively be euthanized prior to becoming moribund.

Despite inherent opposing biases for the research staff and veterinary staff, the data for the respective groups were not skewed toward higher or lower numbers. The research staff, who were not blinded to treatment assignment, might be expected to assign lower scores to avoid euthanizing animals. The tests of marginal symmetry indicated no significant difference in the distribution of scores between any of the 3 types of observers. Therefore there was no obvious bias toward higher or lower scores for the 2 main groups of observers, and the observer agreement information suggests that both the research and veterinary staffs were quite consistent with the scores assigned by the person who trained the participants in using the scoring mechanism. This result further indicates that training was effective for both groups of observers. The marginal homogeneity test did identify a difference between the veterinary and research staff. This difference, although statistically significant, is likely due to the difference in the numbers of scores of 0 and 1 between the 2 observers and is probably not clinically significant.

The results of the current study suggest that an observation-based scoring system can be an effective tool for harmonizing endpoint criteria between studies and for mitigating the pain and distress associated with ARS in mice. Using a cut-off score of 7 or 8 in studies involving TBI would accurately predict death as much as 2 d in advance, providing for euthanasia of mice before they become moribund and without affecting experimental outcomes. Furthermore, the use of consistent endpoint criteria will allow for improved comparison of results between studies. An investigator or IACUC could use this scoring system to establish behavior-based criteria for TBI studies based on the specific scientific objective of the protocol. Although these endpoint criteria may be useful for mice receiving bone marrow or stem cell transplants, establishing mortality curves, or even evaluating radiation therapies, there are limitations in using these same criteria for studies identifying mean survival times. Despite this limitation, the use of surrogate humane endpoint criteria on the basis of a well-defined observation-based scoring system that is applied consistently by research and veterinary personnel has the potential to positively affect animal welfare in studies using TBI.

## Acknowledgments

We thank Kassim Kabirov for allowing us to do this observational study in conjunction with ongoing Toxicology Research Laboratories' total body irradiation studies. We also thank Drs Julia Goldman, Cynthia Adams, Kelly Garcia, Lisa Halliday, and Jeanette Purcell for their assistance in scoring the animals.

## References

1. Anno GH, Baum SJ, Withers HR, Young RW. 1989. Symptomatology of acute radiation effects in humans after exposure to doses of 0.5–30 Gy. *Health Phys* 56:821–838.
2. Banerjee M, Capozzoli M, McSweeney L, Sinha D. 1999. Beyond  $\kappa$ : a review of interrater agreement measures. *Can J Stat* 27:3–23.
3. Cicchetti DV, Feinstein AR. 1990. High agreement but low  $\kappa$ : II. Resolving the paradoxes. *J Clin Epidemiol* 43:551–558.
4. Clarke R. 1997. Issues in experimental design and endpoint analysis in the study of experimental cytotoxic agents in vivo in breast cancer and other models. *Breast Cancer Res Treat* 46:255–278.
5. Davis TA, Landauer MR, Mog SR, Barshishat-Kupper M, Zins SR, Amare MF, Day RM. 2010. Timing of captopril administration determines radiation protection or radiation sensitization in a murine model of total body irradiation. *Exp Hematol* 38:270–281.
6. DiCarlo AL, Maher C, Hick JL, Hanfling D, Dainiak N, Chao N, Bader JL, Coleman CN, Weinstock DM. 2011. Radiation injury after a nuclear detonation: medical consequences and the need for scarce resources allocation. *Disaster Med Public Health Prep* 5 Suppl 1:S32–S44.
7. Fleiss JL, Cohen J. 1973. The equivalence of weighted  $\kappa$  and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 33:613–619.
8. Fleiss JL, Levin B, Paik MC. 2003. *Statistical methods for rates and proportions*, 3rd ed. New York (NY): John Wiley and Sons.
9. Galun E, Burakova T, Ketzinel M, Lubin I, Shezen E, Kahana Y, Eid A, Ilan Y, Rivkind A, Pizov G, Shouval D, Reisner Y. 1995. Hepatitis C virus viremia in SCID→BNX mouse chimera. *J Infect Dis* 172:25–30.
10. Hess AD, Bright EC, Thoburn C, Vogelsang GB, Jones RJ, Kennedy MJ. 1997. Specificity of effector T lymphocytes in autologous graft-versus-host disease: role of the major histocompatibility complex class II invariant chain peptide. *Blood* 89:2203–2209.
11. Hidalgo A, Chang J, Jang JE, Peired AJ, Chiang EY, Frenette PS. 2009. Heterotypic interactions enabled by polarized neutrophil microdomains mediate thromboinflammatory injury. *Nat Med* 15:384–391.
12. Institute of Laboratory Animal Research. 2011. *Guide for the care and use of laboratory animals*, 8th ed. Washington (DC): National Academies Press.
13. Jackson KA, Mi T, Goodell MA. 1999. Hematopoietic potential of stem cells isolated from murine skeletal muscle. *Proc Natl Acad Sci USA* 96:14482–14486.
14. Jones RJ, Celano P, Sharkis SJ, Sensenbrenner LL. 1989. Two phases of engraftment established by serial bone marrow transplantation in mice. *Blood* 73:397–401.
15. Kallman RF, Silini G. 1964. Recuperation from lethal injury by whole-body irradiation. I. Kinetic aspects and the relationship with conditioning dose in C57Bl mice. *Radiat Res* 22:622–642.
16. Kalter H. 1968. Sporadic congenital malformations of newborn inbred mice. *Teratology* 1:193–199.
17. Kennedy DW, Abkowitz JL. 1997. Kinetics of central nervous system microglial and macrophage engraftment: analysis using a transgenic bone marrow transplantation model. *Blood* 90:986–993.
18. Krause DS, Theise ND, Collector MI, Henegariu O, Hwang S, Gardner R, Neutzel S, Sharkis SJ. 2001. Multiorgan, multilineage engraftment by a single bone marrow-derived stem cell. *Cell* 105:369–377.
19. Lambertsen KL, Deierborg T, Gregersen R, Clausen BH, Wirenfeldt M, Nielsen HH, Dalmau I, Diemer NH, Dagnaes-Hansen F, Johansen FF, Keating A, Finsen B. 2011. Differences in origin of reactive microglia in bone marrow chimeric mouse and rat after transient global ischemia. *J Neuropathol Exp Neurol* 70:481–494.
20. Landis JR, Koch GG. 1977. An application of hierarchical  $\kappa$ -type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 33:363–374.
21. Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174.
22. Lloyd M, Wolfensohn S. 1999. Practical use of distress scoring systems in the application of humane endpoints, p 48–53. *Proceedings of the International Conference on Humane Endpoints in Animal Experiments for Biomedical Research*. Zeist (The Netherlands): Royal Society of Medicine Press.
23. Moccia KD, Olsen CH, Mitchell JM, Landauer MR. 2010. Evaluation of hydration and nutritional gels as supportive care after total-body irradiation in mice (*Mus musculus*). *J Am Assoc Lab Anim Sci* 49:323–328.
24. Morton DB. 2000. A systematic approach for establishing humane endpoints. *ILAR J* 41:80–86.
25. Morton DB, Griffiths PH. 1985. Guidelines on the recognition of pain, distress, and discomfort in experimental animals and a hypothesis for assessment. *Vet Rec* 116:431–436.
26. Nunamaker EA, Artwohl JE, Anderson RJ, Fortman JD. 2013. Endpoint refinement for total body irradiation of C57BL/6 mice. *Comp Med* 63:22–28.
27. Paster EV, Villines KA, Hickman DL. 2009. Endpoints for mouse abdominal tumor models: refinement of current criteria. *Comp Med* 59:234–241.
28. Penack O, Holler E, van den Brink MR. 2010. Graft-versus-host disease: regulation by microbe-associated molecules and innate immune receptors. *Blood* 115:1865–1872.
29. Plett PA, Sampson CH, Chua HL, Joshi M, Booth C, Gough A, Johnson CS, Katz BP, Farese AM, Parker J, Macvittie TJ, Orschell CM. 2012. Establishing a murine model of the hematopoietic syndrome of the acute radiation syndrome. *Health Phys* 103:343–355.
30. Roderick TH. 1963. The response of 27 inbred strains of mice to daily doses of whole-body X-irradiation. *Radiat Res* 20:631–639.
31. Rugh R, Castro V, Balter S, Kennelly EV, Marsden DS, Warmund J, Wollin M. 1963. X-rays: are there cyclic variations in radiosensitivity? *Science* 142:53–56.
32. Stokes WS. 2002. Humane endpoints for laboratory animals used in regulatory testing. *ILAR J* 43 Suppl:S31–S38.
33. Toth LA. 1997. The moribund state as an experimental endpoint. *Contemp Top Lab Anim Sci* 36:44–48.
34. Toth LA. 2000. Defining the moribund condition as an experimental endpoint for animal research. *ILAR J* 41:72–79.
35. Travis EL, Peters LJ, McNeill J, Thames HD Jr, Karolis C. 1985. Effect of dose rate on total body irradiation: lethality and pathologic findings. *Radiation Oncol* 4:341–351.
36. Turhan A, Weiss LA, Mohandas N, Coller BS, Frenette PS. 2002. Primary role for adherent leukocytes in sickle cell vascular occlusion: a new paradigm. *Proc Natl Acad Sci USA* 99:3047–3051.
37. Williams JP, Brown SL, Georges GE, Hauer-Jensen M, Hill RP, Huser AK, Kirsch DG, Macvittie TJ, Mason KA, Medhora MM, Moulder JE, Okunieff P, Otterson MF, Robbins ME, Smathers JB, McBride WH. 2010. Animal models for medical countermeasures to radiation exposure. *Radiat Res* 173:557–578.