Original Research

Automated and Continuous Monitoring of Animal Welfare through Digital Alerting

Johnny P Do, Erwin B Defensor, Christine V Ichim, Maria A Lim, Jordan A Mechanic, Mark D Rabe, and Laura R Schaevitz[*]

A primary goal in preclinical animal research is respectful and responsible care aimed toward minimizing stress and discomfort while enhancing collection of accurate and reproducible scientific data. Researchers use hands-on clinical observations and measurements as part of routine husbandry procedures or study protocols to monitor animal welfare. Although frequent assessments ensure the timely identification of animals with declining health, increased handling can result in additional stress on the animal and increased study variability. We investigated whether automated alerting regarding changes in behavior and physiology can complement existing welfare assessments to improve the identification of animals in pain or distress. Using historical data collected from a diverse range of therapeutic models, we developed algorithms that detect changes in motion and breathing rate frequently associated with sick animals but rare in healthy controls. To avoid introducing selection bias, we evaluated the performance of these algorithms by using retrospective analysis of all studies occurring over a 31-d period in our vivarium. Analyses revealed that the majority of the automated alerts occurred prior to or simultaneously with technicians' observations of declining health in animals. Additional analyses performed across the entire duration of 2 studies (animal models of rapid aging and lung metastasis) demonstrated the sensitivity, accuracy, and utility of automated alerting for detecting unhealthy subjects and those eligible for humane endpoints. The percentage of alerts per total subject days ranged between 0% and 24%, depending on the animal model. Automated alerting effectively complements standard clinical observations to enhance animal welfare and promote responsible scientific advancement.

Abbreviations: BCS, body condition score; FD, found dead; FN, false negative; FP, false positive; HE, humane endpoint; PPV, positive predictive value; TN, true negative; TP, true positive

Refining methods to minimize pain and distress is a fundamental principle for upholding animal welfare and maintaining scientific integrity in preclinical research. As such, there is a need to improve approaches for identifying clinical signs that warrant veterinary intervention or euthanasia to reduce both unpredicted (i.e., 'found dead') and predicted mortality of laboratory research animals. Rodents nearing endpoint show a number of clinical signs, including body weight loss, hypothermia, changes in physical appearance (e.g., squinted eyes, hunched posture), labored breathing, and decreased physical activity.[4,10] A number of these health indicators also are used to monitor aging mice[27,33,41] and various disease models, such as mouse models of infectious disease,[6,9,34] oncology,[25,39,42] and sepsis.[13,14] In the laboratory, researchers and staff constantly monitor animals for these clinical signs as part of routine husbandry procedures (e.g., daily cageside observation) or as part of a study protocol (e.g., clinical observation, body weight measurement, body condition scoring).

Monitoring of health indicators can be challenging to implement reliably and efficiently. Some of these challenges include: 1) daily, manual health checks or cageside observations that may unintentionally miss health indicators because assessments are commonly performed during the light cycle when rodents are least active; 2) manual scoring methods may require training for consistency; 3) methods may be time-consuming and labor-intensive for long-term studies or studies with large numbers of animals; and 4) handling of animals may induce inadvertent stress or increase inter- and intrasubject variability on-study.[6,10,17,18]

To help address these problems, we investigated whether automated alerting of changes in behavior and physiology can complement existing clinical observations and measurements to improve the identification of animals in pain or distress. We here describe the development of alerting algorithms using a digital platform that capitalizes on noninvasive, home cage monitoring and computer vision-derived biomarkers to continuously detect changes in motion and breathing rate.[15,16,26] We investigated the utility and accuracy of automated alerting by using 2 types of retrospective analysis. First, to determine the broad applicability of the alerts, we applied the algorithms to cross-sectional data collected from all studies in-progress in our vivarium over a 31-d time period; this time range was selected to maximize the variety of therapeutic models being tested. Second, because some studies in the first analysis were in the middle of execution, we next applied the algorithms to the entire duration of 2 specific studies: a genetic model of rapid aging and a lung metastasis model. These models were selected as case studies to provide a complete overview of algorithm performance where

unexpected mortality is common and disease onset is challenging to monitor. The validity of alerts was verified by evaluating them against data collected from standard clinical observations and measurements (e.g., manual cageside health checks, clinical observations, body weight measurements, body condition scoring) performed at the time of study execution.

Materials and Methods

The analyses described herein were performed retrospectively on available datasets from completed studies. Independent datasets were used for development of the automated alerting algorithm (part 1), performance evaluation of the alerting (part 2), and case study analysis (parts 3 and 4). The performance evaluation included all studies in-progress during a 31-d time in our vivarium. This time period was specifically selected to maximize the variety of therapeutic models being tested. In addition, we completed 2 retrospective case study analyses using data collected throughout a complete study: a genetic rodent model of rapid aging (part 3) and a lung metastasis model (part 4). All procedures were performed during the light cycle (0600 to 1800 PDT). Experiments were conducted in Vium's AAALAC-accredited Digital Vivarium in accordance with the current National Research Council *Guide for the Care and Use of Laboratory Animal*[11] and were IACUC-approved.

Animals and housing. All animals were maintained in SPF facility in which the following pathogens were screened for exclusion: viruses (parvovirus, murine norovirus, rodent coronavirus, murine rotavirus, theilovirus), bacteria (helicobacter, *Pasteurella pneumotropica*, *Corynebacterium bovis*), and parasites or protozoa and fungi (fur mites, pinworms, *Spironuoleus muris*, entamoeba, *Pneumocystis*). A complete list of pathogens screened is available on request. Animals were single-housed in instrumented IVC (Digital Smart House, Vium, San Mateo, CA, and Innovive, San Diego, CA) containing corncob, Alpha-Dri, or Alpha-Nest bedding (Shepherd Specialty Papers, Watertown, TN). Animals had unrestricted access to food (Pico Rodent Diet 5053, LabDiet, St Louis, MO) and acidified, sterile water (Innovive). Environmental enrichment, including running wheels, ladders, cotton squares (Ancare, Bellmore, NY), and foraging mixes (Veggie Relish, LabDiet), was provided in each cage.

The Vium digital platform. As described elsewhere,[15,16] Vium Digital Smart Houses consist of standard IVC slotted in Vium's rack system. Digital Smart Houses are outfitted with sensors and a high-definition camera that enables continuous, 24/7 monitoring of animals from each cage. Video collected by using this system is processed by using computer vision algorithms to produce a digital history of motion (mm/s) and breathing rate (breaths per minute).[15,16] The platform streams data (raw video, data analytics) to a secure cloud-based infrastructure that researchers can access remotely in near real-time.

Part 1: Development of alerting algorithm. To create the alerting algorithms, we used data from a cohort of male (*n* = 14) and female (*n* = 24) rodents with obvious behavioral and physiologic signatures of health decline from 13 completed studies. The 13 studies represented a broad range of biology and physiology comprising 4 aging, 3 immunology, 2 inflammation, one phenotyping, one oncology, one respiratory, and one metabolic model. Mouse strains included in these studies were C57BL/6J (*n* = 15), NOD.Cg-*Prkdc^scid^ Il2rg^tm1Sug^*/JicTac (*n* = 2), athymic nude mouse (*n* = 9), NZBWF1 (*n* = 3), and C3D2F1×CB6JF1 4-way crossed mice (*n* = 1), as well as genetically modified animals on the C57BL/6J background strain (*n* = 7) either bred inhouse or purchased from standard providers (Jackson Laboratories,

Bar Harbor ME; Taconic Biosciences, Rensselaer, NY; Envigo, Somerset, NJ). In addition, one rat strain was included in these studies (Lewis rat; Charles River Laboratories, Wilmington, MA). In these studies, moribundity was a criterion for humane endpoint euthanasia. Other criteria included (note that not all criteria were present in all studies): body condition score (BCS) ≤ 2, inability to access food and water, cardiorespiratory distress or agonal breathing, body weight loss ≥ 20% or 30% (depending on the study), bleeding from any orifice, severe rectal prolapse with necrosis, severe alopecia or skin lesions, diarrhea, convulsions, edema, and lymphadenopathy.

We collected 43 ethogram segments to create a training dataset for generating motion-based alerts (duration [mean ± 1 SD], 26.95 ± 24.30 d; range, 3.09 to 131.06 d; Figure 1 A). Of these, 28 segments were provided from 27 subjects with reported health events (e.g., sickness, humane endpoint euthanasia, unexpected death), whereas the remaining 15 segments were provided from 13 healthy subjects, thus representing both positive and negative examples, respectively. Scientists with experience using the digital platform manually identified and labeled segments with alert windows (e.g., periods of obvious decreased motion or changes in breathing rate that coincided with reported health events). Segments were categorized as true positive (TP) when an alert was generated within the predefined alert window corresponding to a reported health event; false positive (FP) when an alert was generated on a reportedly healthy animal, true negative (TN) for each day no alert was generated on reportedly healthy animals; and false negative (FN) for the absence of an alert within the predefined alert window corresponding to a reported health event (Figure 1 B).

To detect changes in motion, raw motion (mm/s) collected during the *test window* (the day or night being assessed) was normalized to a *rolling baseline window* (days or nights prior to the test window; Figure 1 C). A combination of parameters was optimized by using a grid search to create the most predictive algorithm, including the 1) minimal number of hours in the baseline window from which to calculate changes in motion; 2) minimal number of hours in the test window, and 3) motion threshold or the drop in normalized motion required to generate an alert.

We collected 43 ethogram segments to create a training dataset for generating breathing rate-related alerts (duration [mean ± 1 SD], 27.20 ± 24.41 d; range, 1.16 to 131.06 d; Figure 2 A). Of these, 30 segments were provided from 25 subjects with reported adverse health events, whereas the remaining 13 segments were provided from 13 healthy subjects. Similar to motion-related alerts, desired alert windows were manually identified by scientists, and the segments were categorized as TP, FP, TN, and FN (Figure 2 B). A combination of parameters was optimized using a grid search to create the most predictive algorithm, including the 1) minimal number of breathing rate observations and days in the baseline window; 2) minimal number of breathing rate observations and hours in the test window; and 3) breathing rate threshold or change in breathing rate required to generate an alert (calculated by using Z-scores; Figure 2 C).

To determine the best combination of parameters for the motion and breathing rate algorithms, a grid search was performed on more than 20,000 combinations of parameters. Each combination was benchmarked by calculating the precision (TP / [TP + FP]), which measures the proportion of alerts related to reported adverse health events, and the recall (TP / [TP + FN]), which measures the proportion of total adverse health events captured by alerts. The final parameter combination was selected by filtering combinations with high
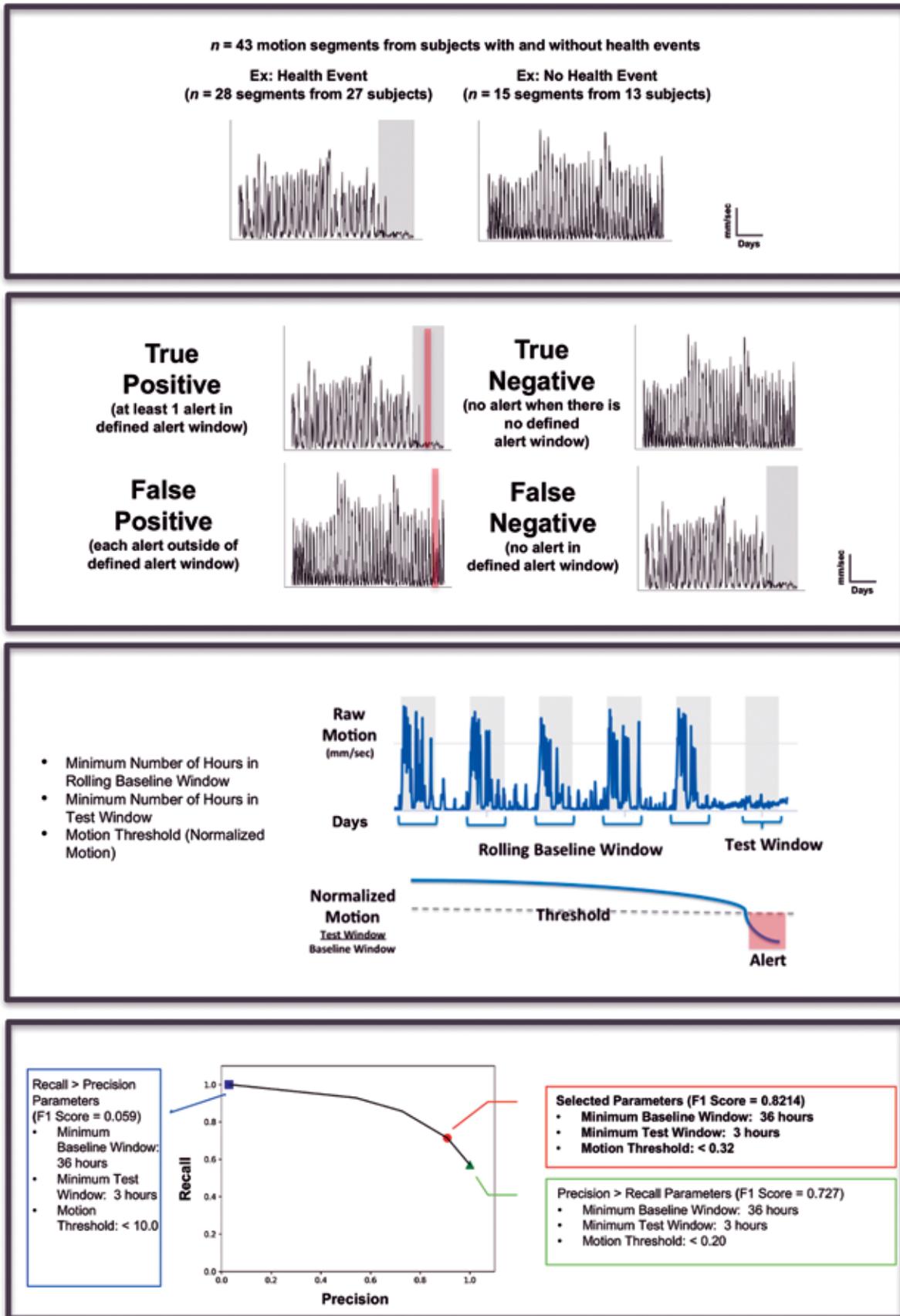
Figure 1. An automated alerting algorithm detects reductions in activity associated with health events. Flow diagram overviewing the development of a motion alerting algorithm. (A) Collect dataset. To create a training dataset, segments of motion ethograms were collected from a diverse range of studies ($n = 13$) comprised of subjects with health events ($n = 28$ segments from 27 subjects) or without health events ($n = 15$

F1 scores ([2 × precision × recall] / [precision + recall]), a metric that weights both precision and recall. Python 2.7.13 (Python Software Foundation, www.python.org) was used for statistical analyses.

Part 2: Retrospective performance evaluation of alerting algorithms. To evaluate the alerting algorithms, we used data from a cohort of male ($n = 216$) and female ($n = 115$) mice from 9 different studies that were simultaneously running in our vivarium during a 31-d time period. The 9 studies within this selected time period represented a broad range of biology and physiology and comprised 3 aging, 3 phenotyping, one oncology, one respiratory, and one kidney disease models. The mouse strains included were C57BL/6J ($n = 173$), CD1 ($n = 20$), athymic nude mouse ($n = 54$), C3D2F1×CB6JF1 4-way crossed mice ($n = 27$), as well as genetically modified animals on the following background strains: BALB/c, $n = 1$; C57BL/6N, $n = 24$; and C57BL/6, $n = 32$; mice were either bred inhouse or purchased from standard providers (Jackson Laboratories; Taconic Biosciences; Envigo). In all of these studies, a criterion for humane endpoint euthanasia was moribundity. Other criteria included (note that not all criteria were present in all studies): BCS ≤ 2; inability to access food and water; agonal breathing or cardiorespiratory distress; body weight loss ≥ 20%, bleeding from any orifice; and severe convulsions. Alerting was retrospectively applied to motion and breathing rate ethograms from a 31-d, consecutive time period and was set to provide alerts every 4 h. Animals were excluded from analysis when they were not currently enrolled in a study (e.g., sentinel or training mice) or were group-housed at the time of observation.

Part 3: Retrospective case study of genetic rodent model of rapid aging. We used a cohort of male and female WT and excision repair cross complementing gene (*Ercc1*)-deficient (*Ercc1Δ/−*) mice that were bred inhouse.[37,40] Briefly, *Ercc1+/−* (in the C57BL6J background) were crossed with *Ercc1Δ/+* mice (on the FVB background) to yield *Ercc1Δ/−* with an F1 C57BL6J/FVB hybrid background. WT littermates were used as controls. The group distribution of mice was as follows: female WT, $n = 14$; female *Ercc1Δ/−*, $n = 8$; male WT, $n = 19$; and male *Ercc1Δ/−*, $n = 7$. At weaning (approximately 25 d old), mice were single-housed and monitored weekly for body weight, clinical signs, and BCS. For this study, the criteria for humane endpoint euthanasia included: BCS ≤ 2, moribundity including unresponsiveness/lack of response to manipulation and/or inability to access food and water, and ulcerated tumors. Subjects that displayed weight loss of ≥20% were monitored for body weight and BCS twice weekly. Automated alerting was retrospectively applied to motion and breathing rate ethograms and was set to provide alerts at 4-h intervals. One subject was excluded from analysis due to

unexpected death on the first day of study observation. In this case, baseline motion or breathing rate values were not available for comparison or normalization during alerting analysis.

Part 4: Retrospective case study of induction rodent model of lung metastasis. We used a cohort of 7-wk-old, male BALB/C mice ($n = 29$; Charles River Lab, Wilmington, MA). As described in the literature,[12,43] animals were administered intravenous doses of either saline or CT26.CL25 (CT26) colon carcinoma cells (ATCC CRL-2639, American Type Culture Collection, Manassas, VA)[40] at 3 different titrations: $1 × 10^4$ cells (low dose), $5 × 10^4$ cells (intermediate dose), and $1 × 10^5$ cells (high dose; $n = 5$ or 6 per group). Additional cohorts of animals ($n = 3$ to 5 per group) received either saline or $2 × 10^6$ CT26 cells subcutaneously to control for the effects of tumor growth without the complication of metastasis. Animals were euthanized and tissues were collected at 28 d after inoculation. Criteria for humane endpoint euthanasia included: BCS = 1, moribundity, unresponsiveness or decreased response to manipulation, inability to self-right, cardiorespiratory distress or agonal breathing, body weight loss >30%, bleeding from any orifice, ulcerating tumors, and tumors that inhibited normal body functions (e.g., eating, drinking, defecation, urination). Automated alerting was retrospectively applied to motion and breathing rate ethograms and was set to provide alerts every 4 h. No animals were excluded from analysis.

Alerting and statistical analysis (parts 2 through 4). Total subject days were calculated as the number of subjects multiplied by the number of days alive on-study. Alerts were reviewed for any coinciding health event by using reports from standard clinical observations and measurements collected at the time of study execution, including manual cageside health checks, clinical examinations, body weight measurements, and BCS. Alerts were then labeled as TP when alerts coincided with the following observed health events: 1) early endpoint (e.g., found dead, humane endpoint); 2) adverse response to an invasive procedure (e.g., blood collection); or 3) expected declining health associated with disease induction that occurred within 5 d of the alert (e.g., lung injury in the bleomycin fibrosis model). Alerts that did not coincide with a reported health event were labeled as FP. Unalerted subjects with a reported adverse health event were labeled as FN, whereas unalerted healthy subjects were labeled as TN. Accuracy was calculated as (TP + TN) / ([TP + FP] + [FP + FN]). The positive predictive value (PPV) was calculated as TP / (TP + FP).

Individual 2-way ANOVA was used to compare the effects of condition on normalized motion across time. Follow-up pairwise comparisons were made by using Dunnett tests for comparing among groups. *P* values less than 0.05 were considered significantly different. Prism 7.0 (GraphPad Software, La Jolla, CA) was used for statistical analysis.

---

segments from 13 subjects). Shaded gray area depicts the alert window (predefined by investigator), which shows changes in motion associated with a known health event (e.g., euthanasia, death). The inset legend indicates days as the unit for the *x* axis and raw motion (mm/s) as the unit for the *y* axis. (B) Categorize alerts. The following benchmarking criteria were used: true positive (TP) if an alert was generated within the predefined alert window corresponding to a reported health event; false positive (FP) if an alert was generated on a reportedly healthy animal, true negative (TN) for each day no alert was generated on reportedly healthy animals; and false negative (FN) for the absence of an alert within the predefined alert window corresponding to a reported health event. The shaded red area depicts an alert. Inset legend indicates days as the unit for the *x* axis and raw motion (mm/s) as the unit for the *y* axis. (C) Identify parameters. To detect changes in motion, we normalized motion (mm/s) collected during the current dark cycle test window, day or night, being surveyed for alerts, to a reference, rolling baseline window. To create the alerting algorithm, we identified a number of parameters, including the minimum or maximum number of hours in the baseline window used to calculate changes in motion, the minimum or maximum number of hours in the test window, and the normalized motion threshold. When normalized motion was below this defined threshold, an alert is generated (red shaded area). Gray shaded areas depict dark cycle period. (D) Optimize parameter combination by grid search. To identify the optimal combination of parameters for the algorithm, we performed a grid search. The selected parameters (red circle with corresponding text box) with an F1 score of 0.80 were the following: 1) minimum baseline window of 36 h with a rolling window that can extend for as long as the 10 previous nights (if data are available), 2) minimum test window of 3 h, and 3) motion threshold less than 0.32. The alerting frequency was set to every 4 h. Other examples of suboptimal parameters wherein Recall > Precision (blue square with corresponding text box) and Precision > Recall (green triangle with corresponding box) are illustrated also.
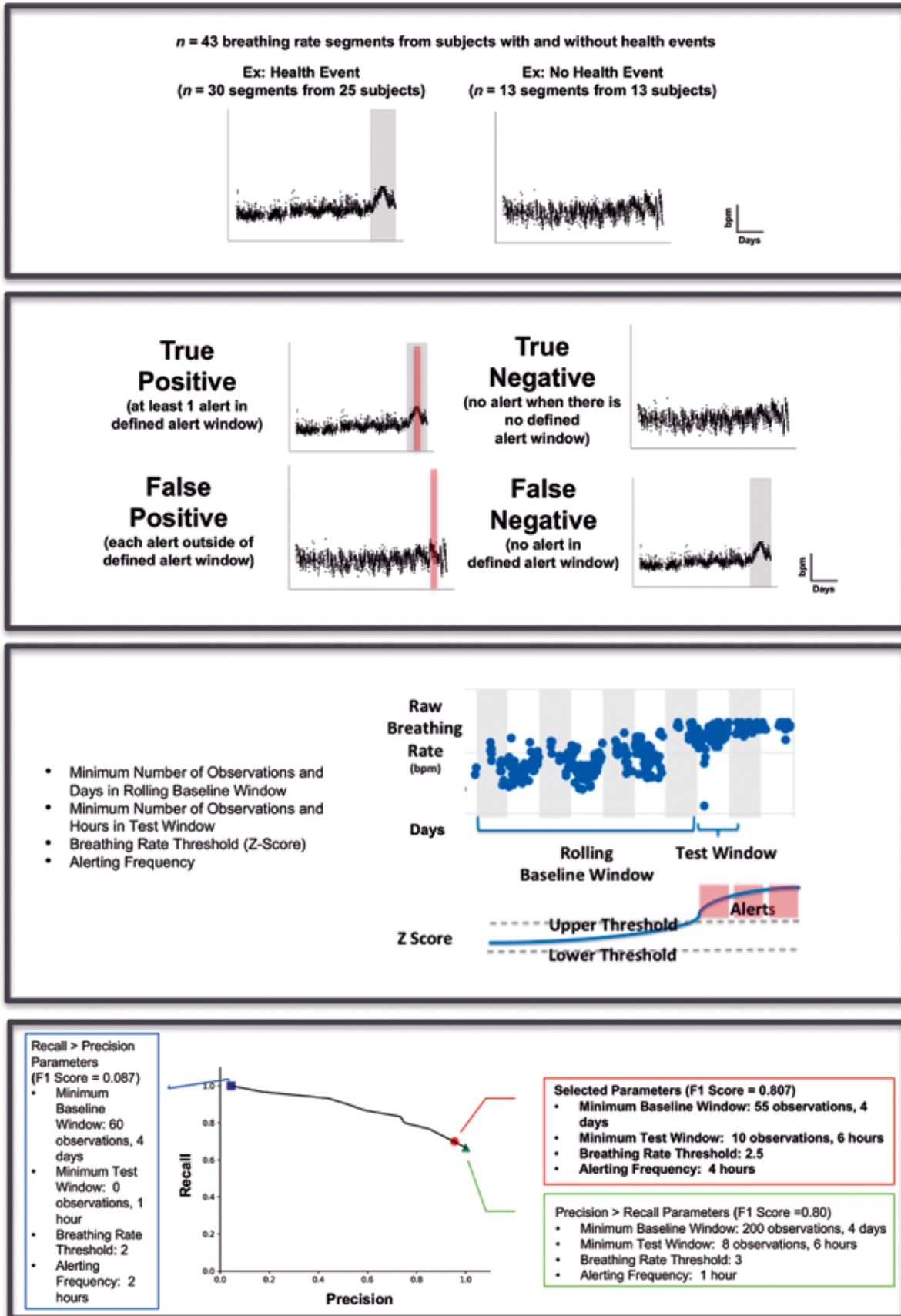
Figure 2. An automated alerting algorithm detects changes in breathing rate associated with health events. Flow diagram overviewing the development of a breathing rate alerting algorithm: (A) Collect dataset. To create a training dataset, segments of breathing rate ethograms were collected from a diverse range of studies (*n* = 13) comprised of subjects with health events (*n* = 30 segments from 25 subjects) or without health events (*n* = 13 segments from 13 subjects). Shaded gray area depicts the alert window (predefined by investigator), which shows changes in

Results

Alerting algorithm for detecting changes in motion. Animals in distress or nearing death often demonstrate lethargy and decreased physical activity.[4,10] To create an algorithm that automatically detects reductions in motion associated with adverse health events, we used a training dataset consisting of ethogram segments from various subjects with and without health events (Figure 1 A) and categorized these segments as TP, FP, TN, or FN according to the reports from clinical observations collected at the time of study execution (Figure 1 B). To detect changes in motion, raw motion data (mm/s) from the dark cycle test window were normalized to a rolling baseline window (Figure 1 C). A grid search was performed by varying the magnitude of the motion decrease and the sizes of the test and baseline windows to identify the highest performing algorithm. The following parameters were found to optimize performance: 1) a motion threshold of 0.32 mm/s (<68% reduction in normalized motion); 2) a baseline window with a minimum of 36 h (i.e., 3 nights) that may extend to include the 10 previous nights when these data are available; and 3) a test window comprising a minimum of 3 h (Figure 1 D). In the training dataset, this algorithm correctly identified 71.4% of all motion segments with reported health events (recall), with 90.9% of alerts representing true health events (precision; F1 score = 0.80).

Alerting algorithm for detecting changes in breathing rate. Animals in distress or nearing death may demonstrate either slowed or rapid, labored breathing.[4,10] Similar to the development of the motion alerting algorithm, ethogram segments from subjects with and without health events were categorized into TP, FP, TN, and FN (Figure 2 A and B). We calculated the probability that the distribution of breathing rate values in the test window differed from the distribution in a rolling baseline window (Figure 2 C). A grid search was performed that varied the required magnitude of the distribution change, the number of breathing rate values, and the sizes of the baseline and test windows, to identify the highest performing algorithm. The following parameters were found to optimize performance of the algorithm: 1) a breathing rate threshold of 2.5 (Z-score); 2) a minimum baseline window of 55 breathing rate points with a rolling window that may extend to include the 4 previous days when these data are available; 3) a minimum test window of 10 breathing rate points within a 6-h period; and 4) an alerting frequency of 4 h (Figure 2 D). In the training dataset, this algorithm correctly identified 70.0% of all breathing rate segments with actual health events (recall), with 95.4% of alerts representing true health events (precision; F1 score = 0.81).

Retrospective performance evaluation. We evaluated the functional performance of automated alerting by retrospectively applying the optimized algorithms to motion and breathing rate ethograms collected for all studies conducted in our vivarium over a 1 mo period (Figure 3). Alerts were categorized as TP, FP, TN, and FN according to whether they corresponded to reported health concerns. A total of 331 subjects was surveyed over a 31-d time period (8908 total subject days). The entire study include 402 alerts (308 motion alerts, 94 breathing rate alerts), representing 4.5% of total subject days. Approximately 80.8% of the total alerts (325 of 402) were associated with a true health event (TP) as confirmed by reported clinical observations and measurements, whereas 19.2% (77 of 402) were not (FP). Reported adverse health observations included body weight loss, decreased BCS (≤2), labored breathing, decreased activity, hunched posture, squinted eyes, convulsions or seizures, presence of tumors or cysts, distended abdomen, skin tenting, and limp tails.

Of the 331 total subjects monitored, 56 subjects reached early endpoint. Alerting algorithms identified more than 2/3 of early endpoint subjects (47 of 56), including 7 of 9 found dead (FD) and 40 of 47 humane endpoint (HE). Motion alerts were more robust, identifying nearly all early endpoint subjects (7 of 9 FD and 39 of 47 HE), than the breathing rate alerting algorithm (1 of 9 FD, 5 of 47 HE); 9 early endpoint subjects (2 FD and 7 HE) were undetected by alerting algorithms (FN). There were 8508 study days with no alerts on subjects observed to be healthy (TN). The accuracy of automated alerting for the 331 subjects over 1 mo was 98.99%.

We further examined the performance of the motion and breathing rate alerting algorithms by subdividing the results of the retrospective analysis according to the 9 therapeutic models (Tables 1 and 2). The percentage of alerts that coincided with true health events was higher for motion than breathing rate (93.2% compared with 40.4%, respectively). Depending on the animal model, PPV ranged between 0% to 100%, and the accuracy was >98% for motion and >96% for the breathing rate algorithms. For motion, the bleomycin-induced lung fibrosis and ES2 cell induction oncology models had the highest PPV (90% to 100%), whereas the 4-way cross and C57BL/6J aging models had the lowest PPV (<50%). For breathing rate, the bleomycin-induced lung fibrosis and ES2 cell induction oncology models demonstrated high PPV (100%), whereas the other studies demonstrated lower PPV (<20%) or did not have breathing rate alerts at all. When further analyzed by documented causes of death, alerting identified 90% of subjects with BCS≤2 and 85.7% of subjects that were moribund prior to endpoint (Table 3). Alerting failed to identify 2 subjects that died shortly after a blood collection procedure.

---

breathing rate associated with a known health event (e.g., euthanasia, death). Inset legend indicates days as the unit for the *x* axis and breaths per minute as the unit for the *y* axis. (B) Categorize alerts. Similar to those for motion, the following benchmarking criteria were used: true positive (TP) if an alert was generated within the predefined alert window corresponding to a reported health event; false positive (FP) if an alert was generated on a reportedly healthy animal, true negative (TN) for each day no alert was generated on reportedly healthy animals; and false negative (FN) for the absence of an alert within the predefined alert window corresponding to a reported health event. Shaded red area depicts an alert. Inset legend indicates days as the unit for the *x* axis and breaths per minute as the unit for the *y* axis. (C) Identify parameters. To detect changes in breathing rate, we used Z-scores to calculate the probability of observing breathing rate values in the test window (i.e., day or night period being surveyed for alerts) within a given distribution (rolling baseline window). To create the alerting algorithm, we identified a number of parameters, including minimum number of observations and days in the rolling baseline window, minimum number of observations and hours in the test window, the breathing rate threshold (Z-score), and alerting frequency. When the Z-score for the test window is above or below the defined threshold, an alert is generated (red shaded area). Gray shaded areas depict dark cycle period. (D) Optimize parameter combination by grid search. To identify the optimal combination of parameters for the algorithm, we performed a grid search. The selected parameters (red circle with corresponding text box) with an F1 score of 0.81 were the following: 1) minimum baseline window of 55 observations with a rolling window that can extend to as many as 4 previous days (if data are available), 2) minimum test window of 10 observations within a 6-h period, 3) breathing rate threshold of 2.5 (Z-score), and 4) alerting frequency every 4 h. Examples of suboptimal parameters wherein Recall > Precision (blue square with corresponding text box) and Precision > Recall (green triangle with corresponding box) are illustrated also.
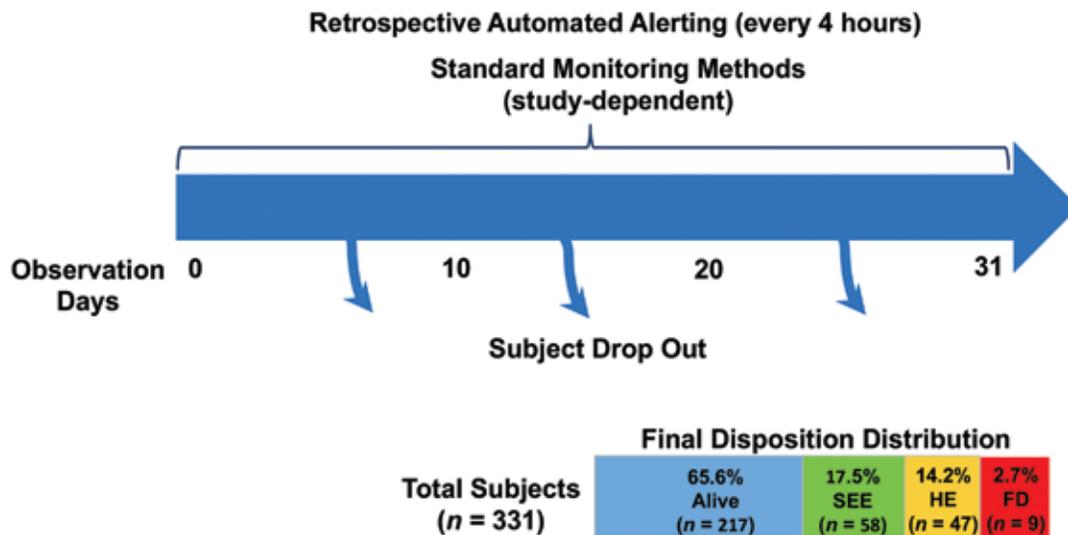
Figure 3. Retrospective performance evaluation of automated alerting demonstrates utility of platform in a live vivarium. Schematic of analysis timeline. To evaluate its performance, we retrospectively applied the automated alerting algorithms to data from 9 studies that were ongoing in a vivarium during a preselected 31-d observation period, which was selected to maximize the variety of therapeutic models running simultaneously. The total number of subjects observed during this time period was 331, although not all subjects were on-study at the same time for the entire duration of the analysis. At the end of the observation period, the final disposition distribution was as follows: 217 animals were alive, 58 had reached study endpoint euthanasia (SEE), 47 had reached a humane endpoint (HE), and 9 were found dead (FD).

The performance evaluation revealed differences in accuracy and predictive value depending on the therapeutic model. Given that many of the studies were in the middle of execution, we wanted to understand whether performance improved when investigated from study start to end. To this end, we performed 2 additional retrospective case study analyses of models in which unexpected mortality and animal welfare as a result of disease onset are challenging to monitor manually: 1) a genetic rodent model of rapid aging and 2) a lung metastasis model, respectively.

Retrospective case study 1: motion alerts in a rodent model of rapid aging. Mice deficient in the DNA excision-repair gene *Ercc1* (excision repair cross complementing gene) demonstrate accelerated aging features.[37,40] Automated alerting was retrospectively applied to monitor *Ercc1Δ/−* mice and WT controls (Figure 4 A). A total of 48 subjects over a 68-d time period was surveyed (1791 total subject days). The results of this retrospective case study are summarized in Figure 4 B and Table 4. Overall, there were 38 alerts (28 motion alerts and 10 breathing rate alerts), representing 2.1% of the total subject days. A total of 21 of 38 alerts (55.3%) coincided with true abnormal health observations (TP), whereas 44.7% (17 of 38) were not (FP). Reported adverse health observations included body weight loss, BCS≤2, labored breathing, and decreased activity. PPV for the motion and breathing rate alerting algorithms were 67.9% and 20%, respectively.

Of the 48 total subjects monitored, 12 *Ercc1Δ/−* subjects reached early endpoint. The majority of early endpoint subjects (10 of 12) were identified by alerting algorithms as soon as 5 d and as late as a few hours prior to endpoint (2 of 2 FD and 8 of 10 HE; Figure 4 C). All HE subjects were euthanized due to moribundity as determined according to manual clinical observations (e.g., squinted eyes, hunched posture, decreased activity). The motion alerting algorithm identified nearly all early endpoint subjects (2 of 2 FD and 8 of 10 HE) compared with the breathing rate alerting algorithm (0 of 2 FD and 3 of 10 HE). Two HE subjects were not detected by alerting algorithms (FN). There were 1751 study days with no alerts on subjects observed to be healthy (TN). The accuracy of automated alerting for this rapid aging model was 98.9%.

To determine whether alerts were associated with unhealthy subjects, we subdivided the alerts according to genotype (Figure 5 A). On average, *Ercc1Δ/−* mice received more motion alerts than WT controls (88% of *Ercc1Δ/−* females, 29% of WT females, 43% of *Ercc1Δ/−* males, 11% of WT males). Regardless of sex, there were no differences in the number of breathing rate alerts between genotypes. We also examined normalized motion across their final 10 d for WT, surviving *Ercc1Δ/−* (subjects that reached study endpoint at 68 d), and nonsurviving *Ercc1Δ/−* mice (FD and HE; Figure 5 B). There were significant differences among groups during specific days ($F_{27,430} = 6.61$, $P \leq 0.0001$). Follow-up pairwise comparisons revealed that, unlike WT and *Ercc1Δ/−* mice that survived until study end, *Ercc1Δ/−* mice that died unexpectedly or reached humane endpoint (all due to moribundity) showed significantly decreased motion 1 to 2 d prior to their endpoint (compared with day −10, $P \leq 0.001$).

Retrospective case study 2: motion and breathing rate alerts in a rodent model of lung metastasis. In the CT26 colon carcinoma cell model of lung metastasis, mice inoculated intravenously—but not subcutaneously—develop tumor metastasis in the lungs, resulting in physiologic changes and increased mortality.[12,43] automated alerting was retrospectively applied to monitor for signs of disease or early endpoint in mice receiving intravenous or subcutaneous inoculations containing different titrations of CT26 cells (Figure 6 A). A total of 29 subjects over a 28-d time period were surveyed (825 total subject days). The results of this retrospective analysis are summarized in Figure 6 B and Table 5. Overall, there were 53 alerts (33 motion alerts and 20 breathing rate alerts), representing 6.4% of the total subject days. Approximately 84.9% of the total alerts (45 of 53) were associated with reported abnormal health observations (TP), whereas 15.1% (8 of 53) were not (FP). Reported adverse health observations included presence of tumors, rapid breathing (mild to moderate), skin tenting, distended abdomen, squinted eyes, penile prolapse, and rough hair coat. PPV for the motion and breathing rate alerting algorithms were 100% and 60.0%, respectively.

Table 1. Retrospective performance evaluation: performance of motion alerting algorithm according to study (*n* = 9)

Type of study or model	*n*	Alerts for subject with health event (TP)	Alerts for subjects with no health event (FP)	Positive predictive value	Unalerted subjects (FN)	Accuracy (%)	Frequency of alerts (% total subject days)
Respiratory (bleomycin induction)	17	10	1	90.9	0	99.7	3.7
Phenotyping (aging)	24	0	0	na	0	100	0
Kidney disease (nephrectomy induction)	20	5	2	71.4	1	99.5	1.1
Aging (C57BL/6J)	117	22	3	88.0	2	99.9	0.7
Aging (4-way cross)	27	8	11	42.1	1	98.6	2.3
Aging (C57BL/6J)	39	0	4	0	3	99.4	0.3
Phenotyping (metabolic)	1	0	0	na	0	100	0
Phenotyping (CNS)	32	0	0	na	0	100	0
Oncology (ES2 cell induction)	54	242	0	100	2	99.8	23.8
Summary (total no. of alerts [%])		287 of 308 (93.2%)	21 of 308 (6.8%)				

FN, false negative; FP, false positive; TP, true positive
A health event was defined by any of the following and was documented through clinical observations at the time of study execution: 1) early endpoint (i.e., animal found dead or at humane endpoint), 2) adverse response to an invasive procedure, or 3) expected signs of disease induction.

Table 2. Retrospective performance evaluation: performance of breathing rate alerting algorithm according to study (*n* = 9)

Type of study or model	*n*	Alerts for subject with health event (TP)	Alerts for subject with no health event (FP)	Positive predictive value	Unalerted subjects (FN)	Accuracy (%)	Frequency of alerts (% total subject days)
Respiratory (bleomycin induction)	17	20	0	100	0	100	6.6
Phenotyping (aging)	24	0	2	0	0	99.7	0.3
Kidney disease (nephrectomy induction)	20	1	5	16.7	1	99.0	1.0
Aging (C57BL/6J)	117	2	10	16.7	2	99.7	0.3
Aging (4-way cross)	27	0	2	0	1	99.6	0.2
Aging (C57BL/6J)	39	0	37	0	3	96.7	3.0
Phenotyping (metabolic)	1	0	0	na	0	100	0
Phenotyping (CNS)	32	0	0	na	0	100	0
Oncology (ES2 cell induction)	54	15	0	100	2	99.8	1.5
Summary (total no. of alerts [%])		38 of 94 (40.4%)	56 of 94 (59.6%)				

FN, false negative; FP, false positive; TP, true positive
A health event was defined by any of the following and was documented through clinical observations at the time of study execution: 1) early endpoint (i.e., animal found dead or at humane endpoint), 2) adverse response to an invasive procedure, or 3) expected signs of disease induction.

Table 3. Retrospective performance evaluation of automated alerting: causes of death for mice found dead or at humane endpoint

Probable cause of death[a]	Found dead (*n* = 9)		Humane endpoint (*n* = 47)		Total (%) alerted
	Alerted	Not alerted	Alerted	Not alerted	
Body condition score ≤ 2	2		34	4	36 of 40 (90.0)
Moribund	1		5	1	6 of 7 (85.7)
Other (e.g., dermatitis distended abdomen, abnormal gait, mass)	4	1	1	1	5 of 7 (71.4)
Soon after blood collection (during daytime hours)		1		1	0 of 2 (0)
Total	7	2	40	7	

[a]Data from clinical observations or measurements at or near endpoint.

Of the 29 total subjects monitored, 3 subjects reached endpoint earlier than the scheduled study endpoint. All 3 early endpoint subjects were identified by automated alerting 1 to 5 d prior to endpoint (2 of 2 FD and 1 of 1 HE; Figure 6). The HE subject was euthanized due to respiratory distress as reported after direct clinical observation. The motion alerting algorithm identified all 3 early endpoint subjects (2 of 2 FD and 1 of 1 HE), whereas the breathing rate alerting algorithm identified 2 of the 3 subjects (1 of 2 FD, 1 of 1 HE). With no unalerted early endpoint subjects (FN) and 777 study days with no alerts on subjects observed to be healthy (TN), the accuracy of automated alerting for the lung metastasis model was 99%.
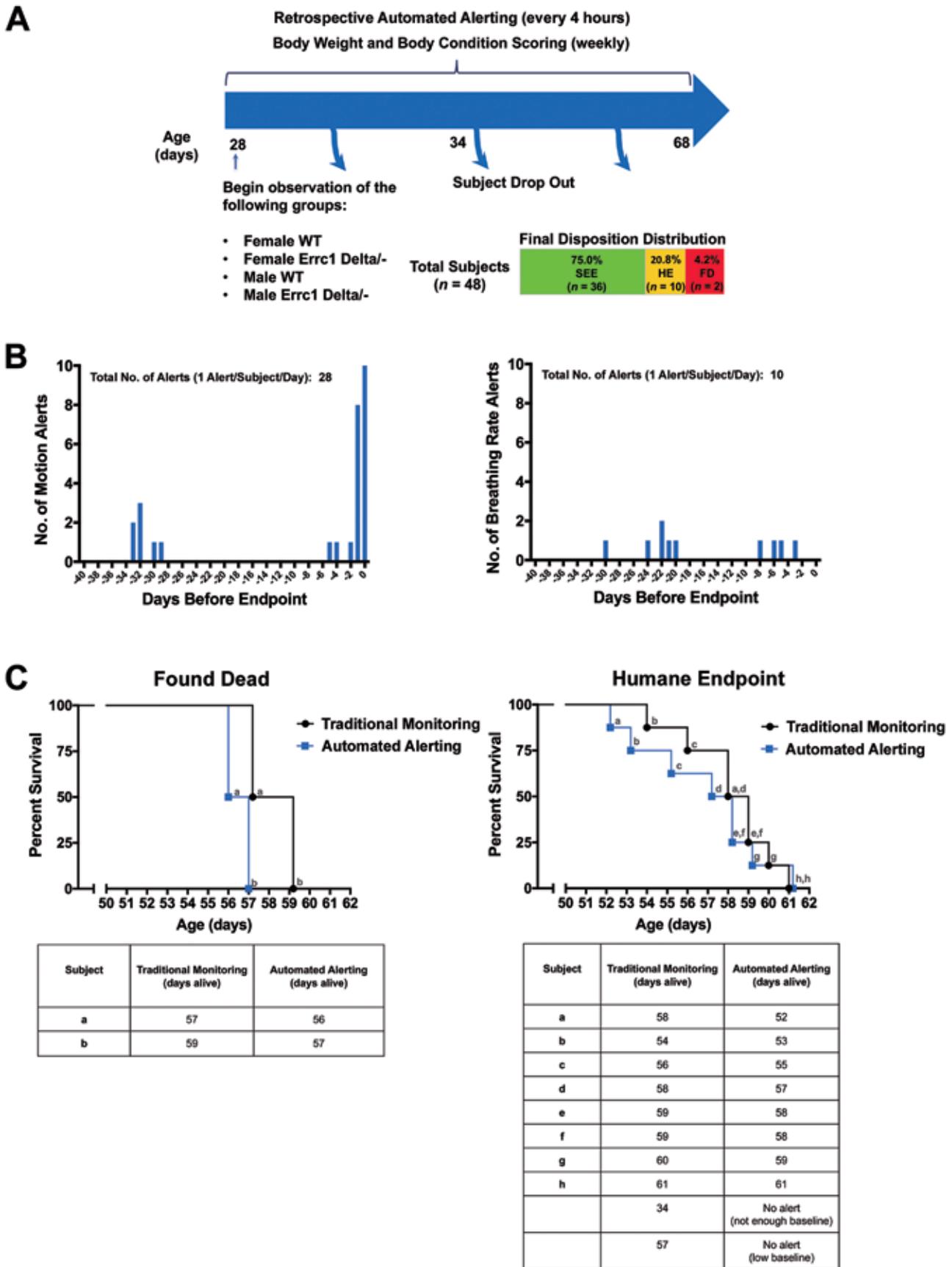
Figure 4. Retrospective case study 1: automated alerting detects endpoint subjects in the *Ercc1* mouse model of rapid aging. (A) Schematic of study timeline. Shortly after weaning, single-housed male and female wild-type (WT) and excision repair cross complementing gene 1 (*Ercc1*) mutant mice (*Ercc1*$^{\Delta/-}$) were observed for signs of accelerated aging ($n = 48$; female WT, $n = 14$; female *Ercc1*$^{\Delta/-}$, $n = 8$; male WT, $n = 19$;

Table 4. Retrospective case study 1: performance of automated alerting (n = 48 mice)

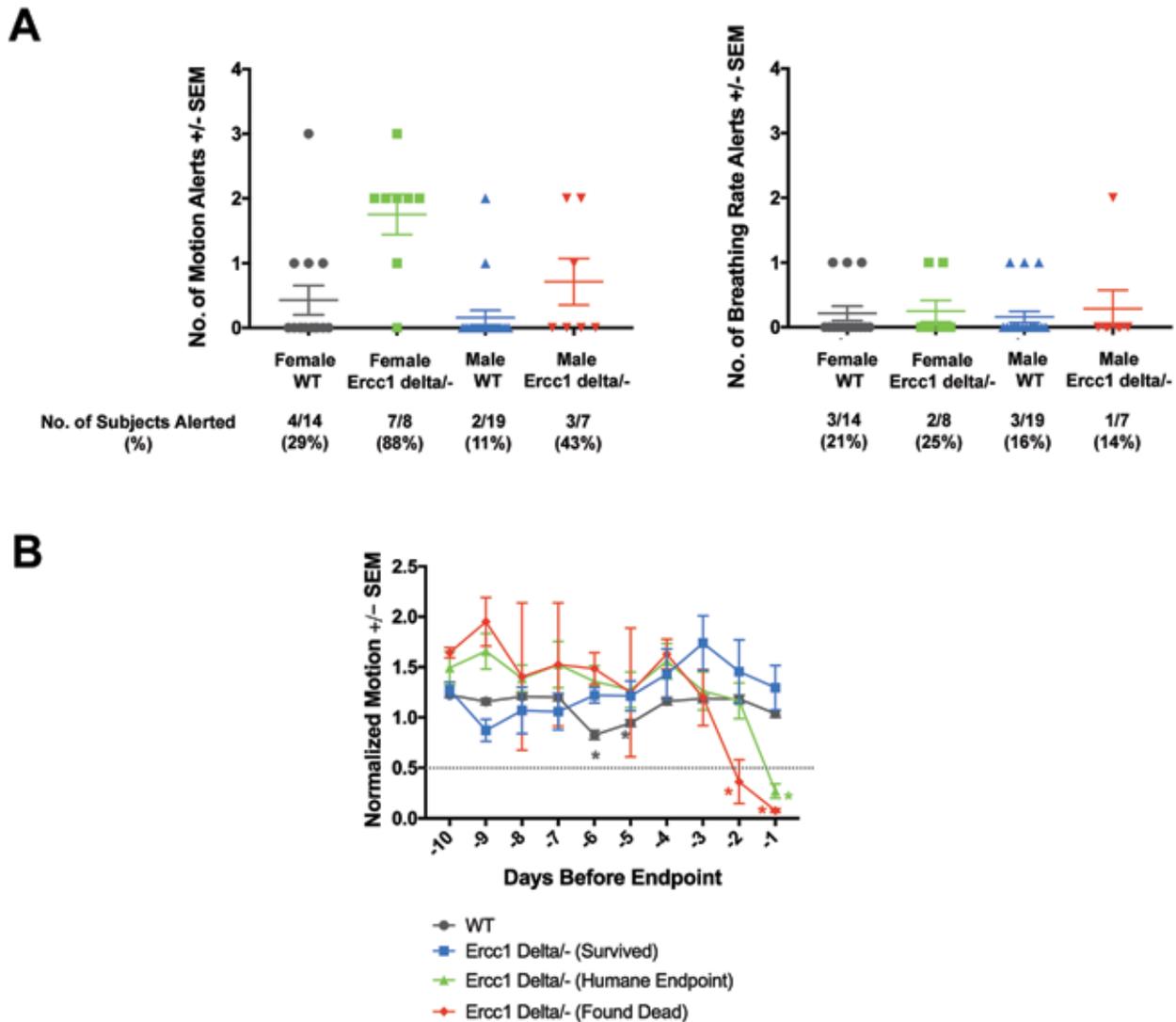|	n	Alerts for subject with health event (true positive)	Alerts for subject with no health event (false positive)
Motion alerts	28	19 (67.9%)	9 (32.1%)
Breathing rate alerts	10	2 (20%)	8 (80%)
Total	38	21 (55.3%)	17 (44.7%)

Figure 5. Retrospective case study 1: automated alerting detects signs of disease in the *Ercc1* mouse model of rapid aging. (A) Number of motion (right panel) and breathing rate (left panel) alerts according to sex and genotype and with corresponding incidences of alerts. Almost all Female *Ercc1*$^{\Delta/-}$ mice (7 of 8, 88%) received alerts. Furthermore, these mice had more alerts than female WT controls. There were no differences in breathing rate alerts between WT and *Ercc1*$^{\Delta/-}$ mice, irrespective of sex. Data are shown as mean ± SEM (error bars). (B) Normalized motion relative to endpoint (death or euthanasia). Closer to endpoint, motion decreased significantly for *Ercc1*$^{\Delta/-}$ subjects that were found dead (FD; n = 2) or humanely euthanized (HE; n = 9) compared with WT controls (n = 33) and *Ercc1*$^{\Delta/-}$ subjects that survived until study end (n = 3). Data given as mean ± SEM (error bars). *, $P \leq 0.05$ compared with day –10.

male *Ercc1*$^{\Delta/-}$, n = 7). At study end, the final disposition distribution was as follows: 36 mice had reached study endpoint euthanasia (SEE), 5 reached a humane endpoint (HE), and 7 were found dead (FD). For the case study, the automated alerting algorithms were retrospectively applied to determine their utility in detecting diseased mice and those at early endpoints. (B) Frequency distributions of motion (left panel) and breathing rate (right panel) alerts aligned to endpoint (day 0). Closer to death or euthanasia, motion alerts increased, whereas breathing rate alerts were scattered throughout the study. (C) Survival curve and corresponding table of subjects that were FD (left panel) or reached HE (right panel) prior to study end. The actual survival curve using traditional monitoring (black circles) represents the death day of each subject, and the automated alerting curve (blue squares) represents the day when the subject first received a 'true' alert (i.e., an alert related to disease or a health event). Lowercase letters correspond to the same subject data point in both curves. In every case, automated alerting predicted the actual event by at least 1 d and, in several cases, by at least 2 d.
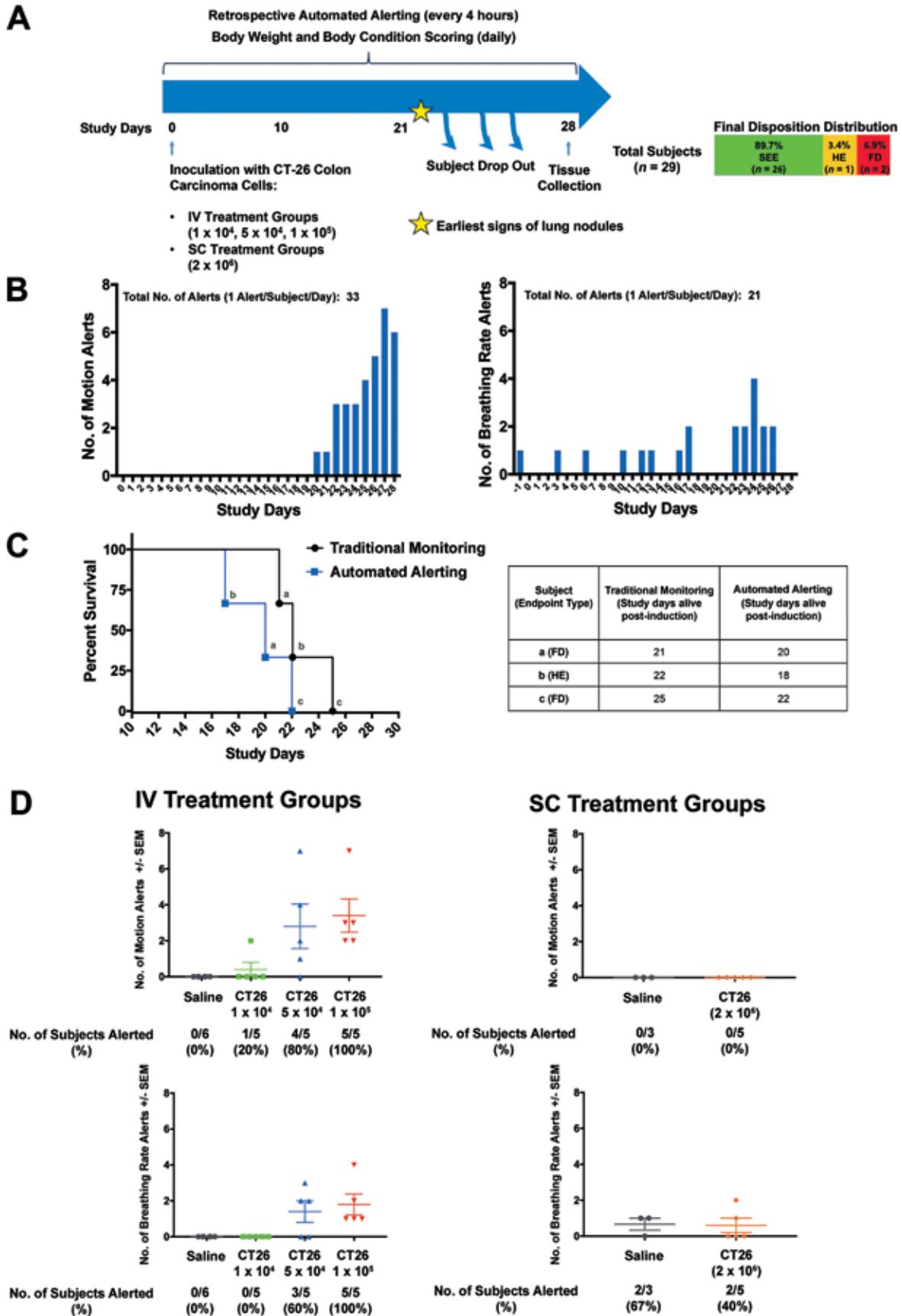
Figure 6. Retrospective case study 2: automated alerting detects diseased and endpoint subjects in a CT26 colon carcinoma cell induction rodent model of lung metastasis. (A) Schematic of study timeline. Single-housed male, BALB/C mice were inoculated intravenously with either saline or various doses of CT26 colon carcinoma cells (1×10^4, 5×10^4, or 1×10^5 cells, $n = 5$ or 6 per treatment group). To control for the effects of cell tumor growth without the development of metastasis, an additional cohort of mice was inoculated subcutaneously with either saline or 2×10^6

Table 5. Retrospective case study 2: performance of automated alerting (*n* = 29 mice)

	n	Alerts for subject with health event (true positive)	Alerts for subject with no health event (false positive)
Motion alerts	33	33 (100%)	0 (0%)
Breathing rate alerts	20	12 (60.0%)	8 (40.0%)
Total	53	45 (84.9%)	8 (15.1%)

To determine whether alerts were associated with increased tumor burden in the lungs, we subdivided the alerts according to treatment group (Figure 6 D). The largest numbers of motion and breathing rate alerts were associated with mice receiving the highest density of cells (1×10^5); all 5 mice (100%) in this group received alerts. The group given the intermediate dose (5×10^4 cells) generated an intermediate number of alerts (4 of 5 mice or 80% received alerts), and the low-density inoculation group (1×10^4 cells) received a similar number of alerts as saline controls (1 of 5 mice; 20% of alerts received). Mice that received subcutaneous inoculations generated few alerts, the number of which was similar to that from mice inoculated with saline.

Discussion

Several manually collected health indicators are used to robustly detect animals nearing humane endpoint for specific rodent models, including infectious disease,[6,9,34] oncology,[25,39,42] aging,[27,33,41] exposure to radiation,[23,24] and sepsis.[13,14] Regardless of the health indicator used, researchers and animal technicians must closely monitor study subjects to minimize pain and unexpected deaths. Here, we tested the hypothesis that noninvasive continuous collection of digital biomarkers (motion and breathing rates) from the home cage can be used to monitor animal welfare and that algorithms based on these biomarkers can be used to automatically detect changes in health condition and alert care takers and investigators accordingly.

The creation and evaluation of the alerting algorithms were conducted retrospectively on data previously collected, to maximize the number of subjects and animal models that could be assessed without requiring the use of additional animals. Motion and breathing rate data from subjects with and without health events were used to first identify common signatures of health decline and then to create and optimize the algorithms to detect these relevant digital biomarker changes with high recall and precision. Performance of the algorithms were evaluated using 2 strategies: 1) cross-sectional analysis of studies occurring simultaneously over a 31-d time period in our vivarium and 2) analysis of the entire duration of 2 specific case studies.

Cross-sectional analysis were performed with the goal of testing the generalizability of the developed alerting algorithms across animal models (e.g., aging, oncology, respiratory), mouse diversity (e.g., strain, sex, age), and study design (e.g., genetic, induction, short-term, long-term), as well as simulating the functionality of automated alerting if deployed broadly in an animal facility. Case study analyses were performed to determine the utility and sensitivity of automated alerting in detecting both humane endpoints and disease onset and progression. To confirm the validity of the alerts (i.e., determine whether alerts were for true health events), all alerts were compared with clinical observations and measurements collected throughout the study.

The accuracy of automated alerting exceeded 90% when applied to studies across a broad range of therapeutic areas. Depending on the animal model, the percentage of alerts per total subject days ranged between less than 1% to 2% in a number of aging models to greater than 20% in an oncology model. Motion alerts had a high PPV across most animal models evaluated. Reduced activity in animals is a general sign of pain or sickness, which can result from disease, a response to a procedure, or approaching a humane endpoint.[3,8,27] Motion alerts were most sensitive in models in which animal health declines rapidly, such as in the bleomycin-induced lung fibrosis model, ES2 cell induction oncology model, and CT26 carcinoma cell induction model of lung metastasis. In contrast, motion alerts performed less well in models where animal health deteriorated slowly, such as in the rapid aging or other aging models. This result is not unexpected, given that the alerting algorithms initially were developed to identify rapid decreases in motion. Further work is necessary to characterize and build algorithms that can better detect decreases in motion occurring over prolonged periods of time.

Breathing rate alerts had a lower PPV than motion across most animal models evaluated. Changes in breathing rate may be more specific to particular conditions, such as respiratory diseases or diseases involving tumors or masses that interfere with breathing.[20,30,36] In general, the PPV of the breathing rate alerting algorithms were low (≤20%), with the exception of 3 disease models: the bleomycin-induced lung fibrosis model, ES2 cell

---

CT26 cells (*n* = 3 to 5 per treatment group). Mice were monitored daily for signs of disease to determine humane and study endpoint. Mice were euthanized and tissues were collected on study day 28 (*n* = 29). The final disposition distribution was as follows: 26 mice reached study endpoint euthanasia (SEE), 1 reached a humane endpoint (HE), and 2 were found dead (FD). The star indicates the earliest observations of lung nodules, indicating lung metastasis, in a separate cohort of intravenously inoculated mice. For the case study, the automated alerting algorithms were retrospectively applied to determine their utility in detecting diseased mice and those at early endpoint. (B) Frequency distributions of motion (left panel) and breathing rate (right panel) alerts. As the study progressed, motion and breathing rate alerts increased, with a total of 33 and 21 alerts, respectively. (C) Survival curve and corresponding table of subjects that were FD or reached HE prior to study end. The actual survival curve using traditional monitoring (black circles) represents the death day of individual subjects, and the automated alerting curve (blue squares) represents the day when each subject first received a 'true' alert (i.e., an alert related to disease or a health event documented by using standard clinical observations or measurements). Lowercase letters correspond to the same subject data point in both curves. Of the 3 animals that reached HE or FD, the alerts predicted the actual event at least 3 d earlier. (D) Number of motion (top panel) and breathing rate (bottom panel) alerts in the intravenous (left panel) and subcutaneous (right panel) treatment groups, with corresponding incidences of alerts. Motion and breathing rate alerts occurred in all 5 mice (100%) inoculated with the highest dose of CT26 cells (1×10^5). Furthermore, this treatment group received more alerts than any other treatment group. Subcutaneously inoculated mice did not show differences in the number of alerts. Data are shown as mean ± SEM (error bars).

induction oncology model, and CT26 carcinoma cell induction model of lung metastasis. In these studies, wherein subjects are expected to show signs of lung disease, PPV for the breathing rate algorithms were 60% to 100%. In the other models, wherein a majority of subjects are not expected to show signs of respiratory illness, the lower PPV is likely due to larger fluctuations in breathing rate. For example, breathing rate may show intermittent increases due to brief periods of intense physical activity.[22] Creating additional breathing rate features that incorporate and account for motion status may help to distinguish such events and further refine the breathing rate alerting algorithm.

Automated alerting identified subjects that unexpectedly died or required euthanasia based on standard humane endpoint criteria. Across all 3 retrospective analysis, automated alerting identified 80% to 100% of early endpoint subjects: 1) 83.9% (47 of 56) in the retrospective performance evaluation; 2) 83.3% (10 of 12) in the rapid aging case study, and 3) 100% (3 of 3) in the lung metastasis case study. Levels of detection were similar for animals that were FD or HE. The algorithms detected the majority of animals that were humanely euthanized due to low BCS (90% in the retrospective performance evaluation) and moribundity (85.7% in retrospective performance evaluation and 80% in rapid aging case study), 2 common humane endpoint criteria used in most animal facilities.[32,35] In general, alerts were generated on early endpoint subjects as soon as 5 d to as late as a few hours prior to recorded death.

Although the majority of subjects that reached an early endpoint had triggered an alert, the alerting algorithms missed a subset of subjects (11 of 71 from all 3 retrospective studies combined). Further investigation revealed a few reasons for the absence of alerts. Six subjects had low levels of baseline activity, either at the beginning of observation or for long periods of time prior to endpoint. Because the baseline window was used as a reference for comparison of changes in motion and breathing rate, the algorithms was less robust in detecting changes when baseline is low. Future refinements can incorporate additional parameters that detect changes in absolute motion and breathing rate. Three subjects demonstrated abnormal circadian rhythms. Although activity during the dark cycle appeared normal or only slightly reduced, activity during the light cycle, when rodents are often at rest, was higher than average. Because the motion algorithm detects reductions in activity during the dark cycle only, future refinements can include training algorithms to recognize deviations in circadian patterns. Finally, 2 subjects did not have an opportunity to receive an alert because they were immediately euthanized due to an adverse response to a blood collection procedure (Table 3).

Overall, we found that automated assessment of behavioral and physiologic measures provides objective indicators of animal welfare. Automated alerting could be an effective tool to increase the timeliness of critical veterinary interventions and scientific decisions. For example, in aging studies, survival and health span measurements are standard endpoints, and numerous subjects are observed in parallel for long periods of time.[1] The ability to notify personnel about animals with declining health allows researchers to efficiently manage subjects in need of attention with minimal disturbance of healthy animals. Early notification may be particularly important in studies with high morbidity and mortality, such as oncology studies, to prevent prolonged pain or distress, as well to make better scientific decisions regarding the prognosis of these subjects.[10,20,39,42]

In addition to the potential for better identifying humane endpoints, automated alerting may identify and notify researchers of disease onset in some models. For example, in the lung metastasis study, the first motion alerts and an increased frequency in breathing rate alerts, especially in the animals receiving the highest concentration of CT26 cells, appeared to coincide with the earliest signs of tumor nodules in the lung. Further optimization of alerting algorithms to identify disease onset could potentially be used to improve study execution by: 1) providing automated confirmation that disease was induced in study animals and 2) indicating when therapeutic dosing should begin.

To determine the broader utility of the alerting algorithms, we simulated the implementation of automated alerting throughout the animal facility. Although we have not yet tested its use in actual practice, preliminary results from the simulation suggest that automated alerting would improve animal welfare without a substantial increase in workload. During our 1 mo (31 d) simulation, a total of 402 alerts were generated across the 331 subjects on study (approximately 13 alerts daily). More than 80% of the alerts were associated with health decline due to disease onset or animals nearing humane endpoint, and less than 20% were false alerts. Importantly, researchers and animal care staff would have received warning of declining health in 7 of the 9 animals that were found dead unexpectedly, significantly improving the chances that these animals would have been euthanized.

Performing this analysis retrospectively does have some limitations. First, we were restricted by the study designs and schedules. We did not know the fate of all alerted subjects, because not all developed disease or reached endpoint prior to the end of the observation period or planned study end. Second, some of the alerts were challenging to review retrospectively because not all animals had clinical annotations close to the time of the alert, which complicated ascertaining whether the animal was healthy at that time. For example, in the rapid aging case study, motion alerts that occurred around the same time for 3 different subjects (6 alerts in total) were labeled as FP because the alert did not coincide with a known health event despite being associated with decreased motion. On further investigation, we found that these alerts coincided with a refill of supplemental feed (e.g., hydrogel, moistened food). Several studies have shown that routine laboratory and husbandry procedures can contribute to changes in animal behavior.[2,5,7,15] Although we cannot firmly determine retrospectively whether animal health was affected, we speculate that these animals were most likely not sick because no clinical annotations had been made during daily health checks. Automated alerting also gives researchers identification of changes in behavior and insights into potential effects of seemingly noninvasive procedures on study data. Altogether, these limitations do not hinder the goal of developing automated alerting: to complement existing health indicators by improving the efficiency and reliability of identifying sick animals through automating the process of collection and analysis of health indicator data, so as to limit pain and distress associated with preclinical research.

Digital tools and technologies are emerging or being adapted for use in animal welfare management and endpoint detection. For example, researchers are using biotelemetry and infrared thermometry to detect hypothermia in animal models of sepsis associated with high mortality rates.[13,19,38] In agricultural husbandry, wearable sensors are being used to monitor animal health.[21,31] In preclinical research, machine-learning strategies[18] and digital platforms[21,28,29,31] that monitor animal behavior and physiology can be useful tools for building on the existing literature and for assessing animal health more efficiently by providing a more comprehensive picture of the wellbeing of an animal. The platform and associated data that

we used for the current study serve as only one example of how automated collection and analysis of digital health data can be used to improve animal welfare. Here we show how automated alerting, which uses near real-time physiologic metrics tied to vital life functions and normal animal behavior, represents a critical advancement toward merging traditional health indicators with technology. With this synergy, critical health conditions are less likely to be overlooked (particularly important against high capacity workloads), and every study subject is used to maximize responsible scientific advancement. The insights provided by automated alerting allow researchers to learn more about the patterns preceding sickness or death (or both) by using behaviorally and physiologically relevant biomarkers. This process will lead to further refinements to the algorithms, allowing for more informed decisions earlier that affect scientific integrity and animal welfare, with the ultimate consequence of improving the ability to predict and prevent animal suffering.

Acknowledgments

References

1. **Ackert-Bicknell CL, Anderson LC, Sheehan S, Hill WG, Chang B, Churchill GA, Chesler EJ, Korstanje R, Peters LL.** 2015. Aging research using mouse models. Curr Protoc Mouse Biol **5:**95–133. https://doi.org/10.1002/9780470942390.mo140195.

2. **Balcombe JP, Barnard ND, Sandusky C.** 2004. Laboratory routines cause animal stress. Contemp Top Lab Anim Sci **43:**42–51.

3. **Brenneis C, Westhof A, Holschbach J, Michaelis M, Guehring H, Kleinschmidt-Doerr K.** 2017. Automated tracking of motion and body weight for objective monitoring of rats in colony housing. J Am Assoc Lab Anim Sci **56:**18–31.

4. **Burkholder T, Foltz C, Karlsson E, Linton CG, Smith JM.** 2012. Health evaluation of experimental laboratory mice. Curr Protoc Mouse Biol **2:**145–165. https://doi.org/10.1002/9780470942390.mo110217.

5. **Febinger HY, George A, Priestley J, Toth LA, Opp MR.** 2014. Effects of housing condition and cage change on characteristics of sleep in mice. J Am Assoc Lab Anim Sci **53:**29–37.

6. **Franco NH, Correia-Neves M, Olsson IA.** 2012. How "humane" is your endpoint? Refining the science-driven approach for termination of animal studies of chronic infection. PLoS Pathog **8:**1–4. https://doi.org/10.1371/journal.ppat.1002399.

7. **Gerdin AK, Igosheva N, Roberson LA, Ismail O, Karp N, Sanderson M, Cambridge E, Shannon C, Sunter D, Ramirez-Solis R, Bussell J, White JK.** 2012. Experimental and husbandry procedures as potential modifiers of the results of phenotyping tests. Physiol Behav **106:**602–611. https://doi.org/10.1016/j.physbeh.2012.03.026.

8. **Gurkar AU, Niedernhofer LJ.** 2015. Comparison of mice with accelerated aging caused by distinct mechanisms. Exp Gerontol **68:**43–50. https://doi.org/10.1016/j.exger.2015.01.045.

9. **Hankenson FC, Ruskoski N, van Saun M, Ying GS, Oh J, Fraser NW.** 2013. Weight loss and reduced body temperature determine humane endpoints in a mouse model of ocular herpesvirus infection. J Am Assoc Lab Anim Sci **52:**277–285.

10. **Hawkins P, Brookes S, Bussell J, Dennison N, Ehall H, Farmer A-M, Langford T, Lelliott C, Lilley E, Ragan I, Ryder K, Wells S.** 2019. Avoiding mortality in animal research and testing. Avoiding mortality in animal research and testing. University of Cambridge: RSPCA Research Animals Department.

11. **Institute for Laboratory Animal Research.** 2011. Guide for the care and use of laboratory animals, 8th ed. Washington (DC): National Academies Press.

12. **Jarnicki AG, Lysaght J, Todryk S, Mills KH.** 2006. Suppression of antitumor immunity by IL-10 and TGF-β-producing T cells infiltrating the growing tumor: influence of tumor environment on the induction of $CD4^+$ and $CD8^+$ regulatory T cells. J Immunol **177:**896–904. https://doi.org/10.4049/jimmunol.177.2.896.

13. **Laitano O, Van Steenbergen D, Mattingly AJ, Garcia CK, Robinson GP, Murray KO, Clanton TL, Nunamaker EA.** 2018. Xiphoid surface temperature predicts mortality in a murine model of septic shock. Shock **50:**226–232. https://doi.org/10.1097/SHK.0000000000001007.

14. **Lilley E, Armstrong R, Clark N, Gray P, Hawkins P, Mason K, Lopez-Salesansky N, Stark AK, Jackson SK, Thiemermann C, Nandi M.** 2015. Refinement of animal models of sepsis and septic shock. Shock **43:**304–316. https://doi.org/10.1097/SHK.0000000000000318.

15. **Lim MA, Defensor EB, Mechanic JA, Shah PP, Jaime EA, Roberts CR, Hutto DL, Schaevitz LR.** 2019. Retrospective analysis of the effects of identification procedures and cage changing by using data from automated, continuous monitoring. J Am Assoc Lab Anim Sci **58:**126–141. https://doi.org/10.30802/AALAS-JAALAS-18-000056.

16. **Lim MA, Louie B, Ford D, Heath K, Cha P, Betts-Lacroix J, Lum PY, Robertson TL, Schaevitz L.** 2017. Development of the digital arthritis index, a novel metric to measure disease parameters in a rat model of rheumatoid arthritis. Front Pharmacol **8:**1–18. https://doi.org/10.3389/fphar.2017.00818.

17. **Marx JO, Brice AK, Boston RC, Smith AL.** 2013. Incidence rates of spontaneous disease in laboratory mice used at a large biomedical research institution. J Am Assoc Lab Anim Sci **52:**782–791.

18. **Mei J, Banneke S, Lips J, Kuffner MTC, Hoffmann CJ, Dirnagl U, Endres M, Harms C, Emmrich JV.** 2019. Refining humane endpoints in mouse models of disease by systematic review and machine learning-based endpoint definition. ALTEX **36:**555–571. https://doi.org/10.14573/altex.1812231.

19. **Mei J, Riedel N, Grittner U, Endres M, Banneke S, Emmrich JV.** 2018. Body temperature measurement in mice during acute illness: implantable temperature transponder versus surface infrared thermometry. Sci Rep **8:**1–10. https://doi.org/10.1038/s41598-018-22020-6.

20. **Mendoza A, Gharpure R, Dennis J, Webster JD, Smedley J, Khanna C.** 2013. A novel noninvasive method for evaluating experimental lung metastasis in mice. J Am Assoc Lab Anim Sci **52:**584–589.

21. **Neethirajan S.** 2017. Recent advances in wearable sensors for animal health management. Sens Biosensing Res **12:**15–29. https://doi.org/10.1016/j.sbsr.2016.11.004.

22. **Nicolò A, Massaroni C, Passfield L.** 2017. Respiratory frequency during exercise: the neglected physiological measure. Front Physiol **8:**1–8. https://doi.org/10.3389/fphys.2017.00922.

23. **Nunamaker EA, Anderson RJ, Artwohl JE, Lyubimov AV, Fortman JD.** 2013. Predictive observation-based endpoint criteria for mice receiving total body irradiation. Comp Med **63:**313–322.

24. **Nunamaker EA, Artwohl JE, Anderson RJ, Fortman JD.** 2013. Endpoint refinement for total body irradiation of C57BL/6 mice. Comp Med **63:**22–28.

25. **Paster EV, Villines KA, Hickman DL.** 2009. Endpoints for mouse abdominal tumor models: refinement of current criteria. Comp Med **59:**234–241.

26. **Peng Q, Mechanic J, Shoieb A, Pardo ID, Schaevitz L, Fenyk-Melody J, Vitsky A, Boucher M, Somps C, Cook JC, Liu CN.** 2019. Circulating microRNA and automated motion analysis as novel methods of assessing chemotherapy-induced peripheral neuropathy in mice. PLoS One **14:**1–17. https://doi.org/10.1371/journal.pone.0210995.

27. **Ray MA, Johnston NA, Verhulst S, Trammell RA, Toth LA.** 2010. Identification of markers for imminent death in mice used in longevity and aging research. J Am Assoc Lab Anim Sci **49:**282–288.

28. **Redfern WS, Tse K, Grant C, Keerie A, Simpson DJ, Pedersen JC, Rimmer V, Leslie L, Klein SK, Karp NA, Sillito R, Chartsias A, Lukins T, Heward J, Vickers C, Chapman K, Armstrong JD.** 2017. Automated recording of home cage activity and temperature of individual rats housed in social groups: The Rodent Big Brother

project. PLoS One **12:**1–26. https://doi.org/10.1371/journal.pone.0181068.

29. **Richardson CA.** 2015. The power of automated behavioural homecage technologies in characterizing disease progression in laboratory mice: A review. Appl Anim Behav Sci **163:**19–27. https://doi.org/10.1016/j.applanim.2014.11.018.

30. **Rumsey WL, Bolognese B, Davis AB, Flamberg PL, Foley JP, Katchur SR, Kotzer CJ, Osborn RR, Podolin PL.** 2017. Effects of airborne toxicants on pulmonary function and mitochondrial DNA damage in rodent lungs. Mutagenesis **32:**343–353.

31. **Rushen J, Chapinal N, de Passillé AM.** 2012. Automated monitoring of behavioural-based animal welfare indicators. Anim Welf **21:**339–350. https://doi.org/10.7120/09627286.21.3.339.

32. **Toth LA.** 2000. Defining the moribund condition as an experimental endpoint for animal research. ILAR J **41:**72–79. https://doi.org/10.1093/ilar.41.2.72.

33. **Trammell RA, Cox L, Toth LA.** 2012. Markers for heightened monitoring, imminent death, and euthanasia in aged inbred mice. Comp Med **62:**172–178.

34. **Trammell RA, Toth LA.** 2011. Markers for predicting death as an outcome for mice used in infectious disease research. Comp Med **61:**492–498.

35. **Ullman-Culleré MH, Foltz CJ.** 1999. Body condition scoring: a rapid and accurate method for assessing health status in mice. Lab Anim Sci **49:**319–323.

36. **Vanoirbeek JA, Rinaldi M, De Vooght V, Haenen S, Bobic S, Gayan-Ramirez G, Hoet PH, Verbeken E, Decramer M, Nemery B, Janssens W.** 2010. Noninvasive and invasive pulmonary function in mouse models of obstructive and restrictive respiratory diseases. Am J Respir Cell Mol Biol **42:**96–104. https://doi.org/10.1165/rcmb.2008-0487OC.

37. **Vermeij WP, Dolle ME, Reiling E, Jaarsma D, Payan-Gomez C, Bombardieri CR, Wu H, Roks AJ, Botter SM, van der Eerden BC, Youssef SA, Kuiper RV, Nagarajah B, van Oostrom CT, Brandt RM, Barnhoorn S, Imholz S, Pennings JL, de Bruin A, Gyenis A, Pothof J, Vijg J, van Steeg H, Hoeijmakers JH.** 2016. Restricted diet delays accelerated ageing and genomic stress in DNA-repair-deficient mice. Nature **537:**427–431. https://doi.org/10.1038/nature19329.

38. **Vlach KD, Boles JW, Stiles BG.** 2000. Telemetric evaluation of body temperature and physical activity as predictors of mortality in a murine model of staphylococcal enterotoxic shock. Comp Med **50:**160–166.

39. **Wallace J.** 2000. Humane endpoints and cancer research. ILAR J **41:**87–93. https://doi.org/10.1093/ilar.41.2.87.

40. **Weeda G, Donker I, de Wit J, Morreau H, Janssens R, Vissers CJ, Nigg A, van Steeg H, Bootsma D, Hoeijmakers JH.** 1997. Disruption of mouse ERCC1 results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. Curr Biol **7:**427–439. https://doi.org/10.1016/S0960-9822(06)00190-4.

41. **Whitehead JC, Hildebrand BA, Sun M, Rockwood MR, Rose RA, Rockwood K, Howlett SE.** 2013. A clinical frailty index in aging mice: comparisons with frailty index data in humans. J Gerontol A Biol Sci Med Sci **69:**621–632. https://doi.org/10.1093/gerona/glt136.

42. **Workman P, Aboagye EO, Balkwill F, Balmain A, Bruder G, Chaplin DJ, Double JA, Everitt J, Farningham DA, Glennie MJ, Kelland LR, Robinson V, Stratford IJ, Tozer GM, Watson S, Wedge SR, Eccles SA; Committee of the National Cancer Research Institute.** 2010. Guidelines for the welfare and use of animals in cancer research. Br J Cancer **102:**1555–1577. https://doi.org/10.1038/sj.bjc.6605642.

43. **Wu G, Kim D, Park BK, Park S, Ha JH, Kim TH, Gautam A, Kim JN, Lee SI, Park HB, Kim YS, Kwon HJ, Lee Y.** 2016. Anti-metastatic effect of the TM4SF5-specific peptide vaccine and humanized monoclonal antibody on colon cancer in a mouse lung metastasis model. Oncotarget **7:**79170–79186. https://doi.org/10.18632/oncotarget.13005.