

Welfare and Scientific Considerations of Tattooing and Ear Tagging for Mouse Identification

Johnny V Roughan* and Tatum Sevenoaks

Ear tagging is perceived as less painful or stressful than tattooing and therefore is generally considered less harmful or costly to welfare. However, ear tags are more difficult to read than tattoos and can fall out, and mice usually require restraint for the tag numbers to be read accurately. We assessed the welfare and scientific implications of tattooing by using a commercial device compared with restraint in a device versus ear tagging. Male and female BALB/c mice ($n = 32$) underwent procedures after 1 wk of tail or nonaversive (tunnel) handling to determine whether tunnel handling reduced anxiety. Pain was evaluated using both the Mouse Grimace Scale (MGS) and manual and automated behavior analyses; light–dark preference testing and voluntary interaction with the handler's hand were used to assess anxiety. Tail inflammation after tattooing was quantified using bioluminescent imaging, and ear tag and tattoo misidentification rates were estimated from volunteer staff records. Tunnel handling reduced anxiety compared with tail handling. According to the MGS, tattooing was not more painful than ear tagging but caused significant tail inflammation and more agitation and anxiety. However, all tattoos were read correctly without handling, whereas all ear tagged mice needed restraint, and at least 25% of the tag codes were misread. Handling stress together with identification errors at this rate represent potentially serious concerns regarding the scientific integrity of data from studies using ear tagging. These concerns are unlikely to arise with tattooing. Although tattooing was stressful, so were restraint and ear tagging. However, considering the other major advantages of tattooing, the total costs associated with tattooing were not substantially greater than for ear tagging.

Abbreviations: FAU, facial action unit; ICC, intraclass correlation coefficient; MGS, Mouse Grimace Scale

DOI: 10.30802/AALAS-JAALAS-18-000057

Laboratory mice prefer social housing, so a method of identifying individual animals within each cage is usually necessary. Commonly stated guidelines on choosing an identification method (for example, the UK Animals [Scientific Procedures] Act, 1986)⁴³ are that it should be appropriate to the study and cause only momentary discomfort or pain and no lasting harm. However, the results of a recent systematic review indicate a lack of consensus about which methods have the least adverse effects on welfare.⁴² Sufficient information is not available to assess the influence on welfare of different methods of identification. For example, although most establishments in the United Kingdom view fur clipping or shaving and using marker pens as virtually benign, both require reapplication, resulting in additional handling or restraint and probably stress.^{14,19} Ear notching is the next most common choice,²⁸ possibly because the marks are permanent and because any welfare concerns are offset by the ability to perform genotyping using the excised tissue. However, additional lifting, handling, or restraint may still be required to confirm an animal's identity, the cumulative effect of which may be to exacerbate stress to a greater extent than is currently appreciated. Preventing stress has important scientific implications because stress influences numerous physiologic systems and compromises biochemical stability.² Depending on its timing and severity, restraint stress can lead to animals becoming variably able to cope with subsequent challenges during scientific testing.⁴² Stress and anxiety are therefore

potentially major sources of statistical 'noise',³⁹ contributing to diminished welfare in individual mice and the need to use more animals. However, the factor with arguably the greatest potential to damage studies is the possibility of misidentification. The scale of this problem is uncertain because the true rates of misidentification are, by their nature, unknown; but if sufficiently frequent, misidentification would lead not only to animals being wasted but also to serious concerns about study integrity,¹⁷ possibly surpassing those arising from some initial discomfort or anxiety. For instance, although ear tagging probably does not represent a significantly painful event, tags can fall out or be torn out if snagged on apparatus or as a result of fighting. In addition, fighting can cause ear notches to become indistinct or torn, leading not only to additional harm but less confident identification. Ear punches can even reseal in some strains.⁴⁰ Although radio-frequency identification devices (that is, microchips) generally avoid the need to handle animals, restraint can become necessary should the implant migrate. Overall, many factors should be considered before choosing the most appropriate identification method.

Although misidentification issues are less likely with tattoos, tattooing is used much less frequently than ear marking or tagging.²⁸ This difference may be due to staff training requirements^{5,8} or because humans know tattoos can be painful, tattooing is perceived as more likely to be detrimental to animal welfare than ear tagging or notching. However, despite the risk of pain, inflammation, and infection, almost 40% of American teenagers have at least one tattoo.²⁹ Although in most cases these involve more substantial inking than would be needed for animal identification purposes, whether tattooing causes similar inflammation and pain in mice is presently unknown.

Received: 16 May 2018. Revision requested: 09 Jul 2018. Accepted: 14 Aug 2018.
Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, Tyne and Wear,
United Kingdom.

*Corresponding author. Email: j.v.roughan@newcastle.ac.uk

Recently marketed as the Labstamp (Somark Innovations, San Diego, CA), a new, partially automated system is available for tail tattooing in mice; including those with either pale or pigmented skin. Because using this device requires minimal staff training and provides an easily recognizable number that reportedly can be read without handling, the manufacturer's claim is that, overall, tattooing by using this device is preferable to most other permanent methods.²⁰ To date, only one study has assessed the effect of this system on mouse welfare.³⁷ That study used behavior, fecal corticosteroid analyses and the Mouse Grimace Scale (MGS) to assess stress and pain.²⁵ Tattooing was thought to cause only mild pain and was viewed as stressful, although not more so than anesthesia. Although the study included an ear notched group (only for genotyping purposes) and because that group of mice was anesthetized, ear notched, and then tattooed, the study could not determine the relative influences of tattooing and ear notching.

When choosing a method of identification, one should consider not only whether it could cause pain but also whether it may cause undue stress or anxiety. For welfare reasons and to reduce experimental noise, stress and anxiety should be minimized in so far as possible.³⁹ Although limiting stress and anxiety in mice has been difficult, mice can be made less anxious by using nonaversive (tunnel) rather than conventional tail handling.^{14,19} In the present investigation, one goal was to evaluate the practical application of tunnel handling, and another was to determine whether the method represents a welfare refinement that also achieves "more robust scientific outcomes"¹⁹ and more reliable behavior data.¹⁵ We used established methods of pain and anxiety testing to assess the relative welfare effects (costs) of tattooing compared with ear tagging. Potential scientific concerns (scientific costs) were assessed by estimating misidentification rates when volunteer staff read tattoo or ear tag numbers. In addition, our study provided an opportunity to test the use of luminol as a more financially viable means of imaging inflammation and one that potentially causes less aversion to mice than in our previous investigation.⁴¹

Materials and Methods

All work was carried out in accordance with the UK Animals (Scientific Procedures) Act (1986) and was approved by the local Animal Welfare and Ethical Review Body.

Animals and husbandry. Female ($n = 16$) and male ($n = 16$) BALB/cAnNCrl mice (age, 10 to 13 wk; weight: 27.0 ± 0.5 g [males], 19.0 ± 1.0 g [females]) were certified free of the common pathogens listed on the vendor's website⁶ and were donated by Charles River UK (Margate, Kent, United Kingdom). These mice were selected because they are a relatively common strain that is thought to be more prone to anxiety than others. Mice were randomly assigned to type 2 IVC (Arrowmights, Hereford, United Kingdom) so that each cage contained either 4 male or 4 female mice. Hardwood bedding (Aspen, BS and S, Edinburgh, United Kingdom), a cardboard tube, a chew block, Sizzle Nest (B and K Universal, Hull, United Kingdom) food (R and M no. 3, SDS, Essex, United Kingdom) and tap water were provided. The holding room was kept at 22 ± 2 °C, $55\% \pm 10\%$ humidity, and on a 12:12-h light:dark cycle (lights on, 0700 to 1900). The IVC rack had 40 air changes per hour. Mice were allowed a 2-wk settling period. Cages were cleaned between 0900 and 1000. Starting 4 d after arrival, mice were weighed on alternate days during the settling period, thus providing 5 baseline weight records as an assurance of normal weight maintenance. On the first day that the 4 mice in each cage were weighed, their tails were marked with red, blue, green, or a black nontoxic marker

pen. These markings were renewed every 2 d without handling the mice. All mice were tail-handled during the 2 cleaning cycles undertaken during the settling phase, and to prevent subsequent disruption to data collection, alternate cages (1, 3, 5, and 7) were cleaned on Monday and the remainder on Wednesday mornings. All husbandry was undertaken by the same person (TS).

Experimental design. The animal numbers in the current study were based on previous work, in which a minimum of 16 mice was sufficient to detect a difference in the effect of tunnel handling compared with tail handling (that is, tunnel handling reduced anxiety).^{14,19} Eight cages were positioned on 2 rows of an IVC rack so that throughout the study, each cage was always adjacent to and above or below one containing mice of the opposite sex. The 4 mice in each cage were randomly allocated for restraint or tattooing and tail or tunnel handling. Figure 1 illustrates the study design according to the rack placement of each cage and the order in which each procedure was applied to each mouse. This balanced arrangement was a precaution against exposing cages to different pheromone levels or visual cues from mice of the opposite sex and was maintained throughout the study. Figure 1 also shows how the study was blocked and counterbalanced so that equal numbers of tail- and tunnel-handled mice of each sex were tattooed or restrained each week. Two cages, one containing 4 male mice and one containing 4 female mice, began procedures each week. The sex of the mice used was alternated weekly. That is, if a cage of male mice began on Monday, it was females 2 d later (Wednesday), and then the following Monday began with female mice, and so on. Figure 2 shows this procedure pattern and the data collection schedule for the first 4 cages, which then was replicated. To standardize tissue injury from tattooing, only 2 numeric codes were used, with digits of identical occurrence (010 or 100). Equal numbers of male and female mice and of tunnel- or tail-handled mice were assigned to receive each digit sequence, but in week 4, 2 mice received the wrong code.

Preacclimation testing. The goal of preacclimation testing was to obtain baseline anxiety readings before beginning differential handling; preacclimation testing began after the settling phase. Each Monday or Wednesday before 1000, light–dark preference testing was followed by hand–approach avoidance (that is, voluntary interaction) testing. Male and female mice had separate sets of apparatus, and all equipment was cleaned with 70% ethanol between recordings. Each light–dark unit was a modified conditioned place-preference box (model CPP-3013AT, Med Associates, St Albans, VT) with black and white compartments fitted with infrared arrays to record compartment residence times and automatically controlled guillotine doors to control access. The black and white compartments were separated by a central gray (start) chamber that had a solid gray plastic floor. The floors of the black and white chambers were fitted with cardboard covered in adhesive black plastic (catalog no. 346-0002, D-C-Fix, Wilko, Nottinghamshire, United Kingdom); black plastic also was used to cover the inside of each compartment lid. The black compartment light was disabled, and the white compartment light was set to full intensity (100 mA, 28 V), thus creating dark and bright enclosures. Mice were lifted by the tail and placed into the gray chamber. The guillotine doors opened after 30 s for a 10-min choice trial and then closed. Voluntary interaction testing was then undertaken in a topless type 2 polycarbonate cage base (40 × 25 × 10 cm; Arrowmights, Hereford, United Kingdom) with the inside covered in the same black plastic. A videocamera (Legria HFM 506, Canon, Tokyo, Japan) was fixed at 90 cm above this cage center. Mice were lifted from the light–dark apparatus by their tail and placed

Week 1	Week 2	Week 3	Week 4
IVC Row 1			
Cage 1: ♂ (Tuesday)	Cage 3: ♀ (Tuesday)	Cage: 5 ♂ (Tuesday)	Cage 7: ♀ (Tuesday)
1. Restraint/Tail	9. Restraint/Tunnel	17. Restraint/Tail	25. Restraint/Tunnel
2. Tattoo/Tail	10. Tattoo/Tunnel	18. Tattoo/Tail	26. Tattoo/Tunnel
3. Restraint/Tunnel	11. Restraint/Tail	19. Restraint/Tunnel	27. Restraint/Tail
4. Tattoo/Tunnel	12. Tattoo/Tail	20. Tattoo/Tunnel	28. Tattoo/Tail
IVC Row 2			
Cage 2: ♀ (Thursday)	Cage 4: ♂ (Thursday)	Cage 6: ♀ (Thursday)	Cage 8: ♂ (Thursday)
5. Restraint/Tunnel	13. Restraint/Tail	21. Restraint/Tunnel	29. Restraint/Tail
6. Tattoo/Tunnel	14. Tattoo/Tail	22. Tattoo/Tunnel	30. Tattoo/Tail
7. Restraint/Tail	15. Restraint/Tunnel	23. Restraint/Tail	31. Restraint/Tunnel
8. Tattoo/Tail	16. Tattoo/Tunnel	24. Tattoo/Tail	32. Tattoo/Tunnel

Figure 1. The study design: cage layout across 2 rows of an IVC rack and order of applying restraint or tattooing procedures during each week of data recording. The 32 individually numbered mice initially underwent daily acclimation to tail or tunnel handling for 7 d prior to being tattooed or restrained. Eight mice (4 male and 4 female) underwent tattooing or restraint each week, and the sex that was tattooed or restrained was alternated between Tuesdays and Thursdays.

into the center of the cage. After approximately 10 s, the operator (TS) placed a gloved hand into the center of the cage, and approach-avoidance behavior was recorded for 3 min. Mice were then lifted by the tail and weighed before being returned to their home cage, which was returned to its previous position on the IVC rack in the holding room.

Handling acclimation. Handling acclimation began the day after preacclimation testing and was done every morning between 1000 and 1200 for 7 d. Mice assigned for tail handling were lifted from their home cage by the base of the tail, supported on the operator's forearm for 10 s, and then placed back into their home cage. The remaining 2 mice were tunnel handled for 60 s by using a plastiglas tunnel, by following the published instructions as closely as possible.¹⁹ Each cage had its own tunnel, which was removed when not in use. After handling acclimation, mice were only ever handled through their assigned method.

Recording after acclimation. After the 7 sessions of acclimation to tail or tunnel handling, videorecordings were made between 1200 and 1330 to determine whether handling differentially altered anxiety and to provide a baseline reading before assessing responses to tattooing, restraint, and ear tagging. Collection of behavior data for automated behavior analysis and close-up footage for manual processing and MGS recordings were therefore added to the schedule. For automated behavior analysis, the mice were first placed into a clear cage base (type 1144B, Techniplast UK, Northamptonshire, United Kingdom) containing only a thin (approximately 0.5 cm) layer of sawdust bedding. This set-up was placed 30 cm in front of a camera (Canon Legria HFM 506). To optimize behavior recognition, the cages were backlit by using an X-ray viewing box (model QUP/A4SL, Jencons Scientific, Hemel Hempstead, United Kingdom). The room lights were switched off, and mice were filmed for 10 min. The room lights were then turned on and the backlight off for an additional 3 min of manual close-up recording. By using the appropriate lifting method, mice were then placed into a plastiglas MGS cube (9 × 9 × 10 cm), where still images were captured using a high-speed camera (model EX-ZR1000, Casio, London, United Kingdom) for 5 min (or until at least 3 clear photographs were captured) as often as possible while mice faced the camera and were not grooming.³² They then underwent light-dark testing and then the approach-avoidance test followed by weighing, as described for preacclimation recording. Mice were then returned to the holding room.

Tattooing and restraint procedures. These procedures were done the day after preacclimation recording (that is, on Tuesday or Thursday afternoons between 1330 and 1530). Mice were weighed before each procedure. Tattoos were applied using a Labstamp Express machine (Somark Innovations, Tonsley, South Australia, Australia). Mice were held on a flat bench so their tail could be measured using a device provided by the manufacturer, and based on this the appropriate needle and paddle size were selected (both of which were supplied with the machine). After being laid on the paddle, the mouse was covered by a red plastic cover, and the tail was secured in the paddle spring-clamp mechanism. The tail was swabbed at the tattoo target site with the supplied tail oil. A magnetic slide containing a nontoxic black ink pad was then placed over the tail distal to the paddle clamp. When tattooing, the device was loaded with the appropriate needle before the assembly containing the mouse was inserted into the machine. The appropriate number (100 or 010) was entered and the start button pressed. The restraint procedure was identical to that for tattooing except that the machine had no needle. The approximately 30-s restraint or tattoo procedures were filmed using an iPhone 5S (Apple, Cupertino, CA) from a distance of approximately 10 cm. Afterward, mice were released into the center of their home cage (without handling) and given 5 min to explore, groom, and remove excess ink and tail oil. Because only one tattooing machine was available, mice of different sexes could not be tattooed or restrained separately; however, the equipment was cleaned with 70% ethanol between procedures.

Postprocedural recording. Recording after tattooing or restraint procedures was as described for postacclimation testing and began approximately 10 min after tattooing or restraint and was repeated at 24 h (that is, Wednesday and Friday afternoons).

Imaging. Tail inflammation was determined through bioluminescent imaging between 1700 and 1830 on the tattoo or restraint day and 24 h later. Luminol (sodium salt, Sigma-Aldrich, Gillingham, Dorset, United Kingdom) was diluted to a concentration of 25 mg/mL in 5 mL PBS. After weighing, pairs of tattooed or restrained mice were anesthetized in an acrylic chamber with 5% isoflurane in 2 L/min oxygen. They were then placed into an in-vivo imaging system (IVIS Spectrum 200, Perkin Elmer, Waltham, MA) in sternal recumbency on a 36 °C stage. Anesthesia was maintained by nose-cone delivery of 2% isoflurane in 500 mL/min oxygen. Tails were imaged within a 5×5-cm field of view. A background image was obtained using

Week	Day	Cage 1 ♂	Cage 2 ♀	Cage 3 ♀	Cage 4 ♂
1	Mon	PreA/Acc1			
	Tues	Acc2			
	Wed	Acc3	PreA/Acc1		
	Thurs	Acc4	Acc2		
	Fri	Acc5	Acc3		
	Sat	Acc6	Acc4		
	Sun	Acc7	Acc5		
2	Mon	PA	Acc6	PreA/Acc1	
	Tues	(Tat/Res) PP/I	Acc7	Acc2	
	Wed	PP/I (24h)	PA	Acc3	PreA/Acc1
	Thurs		(Tat/Res) PP/I	Acc4	Acc2
	Fri		PP/I (24h)	Acc5	Acc3
	Sat			Acc6	Acc4
	Sun			Acc7	Acc5
3	Mon		PA		Acc6
	Tues		(Tat/Res) PP/I		Acc7
	Wed		PP/I (24h)		PA
	Thurs				(Tat/Res) PP/I
	Fri				PP/I (24h)

Figure 2. An illustration of the days on which differential handling was done and the daily data collection schedule for the first 4 cages (3 study weeks). Two cages of 4 males and 2 cages of female mice entered the study each week, alternating sex between Mondays and Wednesdays. PreA (prehandling acclimation anxiety testing) was followed by 7 d of tail or tunnel handling of the mice in each cage (Acc1 through 7). Each cage then underwent Post (handling) Acclimation anxiety and baseline pain scoring (PA) before being tattooed or restrained (Tat/Res), and then the postprocedural anxiety and pain assessments (PP) were undertaken. The mice were imaged (I) the same evening, and the PP and imaging data collection was repeated at 24 h. Mice that had undergone restraint were ear tagged 2 wk later, immediately after which the PP recordings were repeated.

an open-filter scan; 1-min exposure. Mice were then injected intraperitoneally with 200 mg/kg luminol using an insulin syringe and splitting the dose between the left and right abdomen. Mice were replaced into the IVIS, and scans were made every 1 min for 20 min. A fresh luminol solution was made each week and was stored at approximately 4 °C but was warmed to room temperature prior to use.

Ear tagging procedure. The 2 mice in each cage that had previously undergone restraint were ear-tagged 2 wk later (thereby minimizing animal use). Mice were lifted according to their assigned method and restrained (scruffed) in the normal manner for intraperitoneal injection. A technician with extensive prior experience in ear-tag placement inserted a self-piercing and locking nickel-copper alloy ear tag (2.4 × 3.1 × 9.5 mm) into each mouse's left ear using an applicator (Kent Scientific, Torrington, CT). Pairs of sequentially numbered tags were chosen randomly from a batch numbered between 700 and 799. Data were then collected in the same manner as those that were obtained following tattooing or restraint, but recordings were only made on the day of the ear-tagging procedure. No data were collected at 24 h, and no imaging was performed.

Identification exercise. The last 4 cages of mice were used in an exercise to simulate routine animal identification. There were 10 volunteer ear-tag and tattoo readers, which were divided into 2 groups depending upon their experience in laboratory animal care. Three senior staff (one Named Animal Care and Welfare Officer and 2 veterinarians) were in the expert class, and the remaining 7, with no or minimal experience, were in the novice class (an undergraduate student, and an office administrator, 3 new scientists, and 2 junior technicians). Wearing eyeglasses as needed, volunteers selected one of the 4 cages, used his or her preferred handling method (tunnel or tail) to lift each mouse,

read and wrote down the tag or tattoo number, and replaced each mouse back into its home cage. Volunteers then repeated these steps until all ear tags and tattoos were read.

Data processing and statistical analysis. Body weight changes were calculated between before and after handling acclimation and from before to after each postprocedural data collection session. Weights after ear tagging were adjusted to compensate for the additional weight of the tag (0.25 g). The combined frequency of 'tail-looks' and 'struggling' was used to measure agitation during tattooing or restraint.³⁷ Briefly, when a mouse turned its head toward its tail, it was counted as a tail look. Struggling was recorded when a mouse made vigorous attempts at escape separated by brief (1 to 2 s) quiescent periods. The MGS was applied during the tattooing and restraint procedures, but only 3 MGS facial action units (FAU; orbital tightening, nose bulge, and ear position) could be seen clearly enough through the paddle cover and were assessed as present (score, 1) or absent (0). The 10-min recordings were processed using HomeCageScan software (Clever Sys, Reston, VA) to determine the frequency of walking, rearing, and grooming. Walking and rearing occurred at equivalent frequencies so were averaged to provide a general measure of behavioral activity. Grooming was assessed separately, as mean bout duration (that is, duration divided by frequency). One person (CB) performed all manual analysis of the 3-min segments of close-up footage in a random manner, scoring the occurrence of 2 pain-specific behaviors ('staggering' and 'twitching')⁴⁶ and the frequency of grooming 'errors' that have previously been linked to stress.²² These were when the normal cephalocaudal sequence was broken; that is, when the paw licking that normally begins the sequence is followed by nose and face or head washing, then the body and other fur is licked before proceeding to the legs, genitals, and tail. If the tail was groomed first or not last, it would then be counted as an error. Because the restraint material could not be distinguished from baseline (because no identification marks could be seen), the footage was scored blindly; blinded scoring was not possible once mice had a tattoo or an ear tag.

The MGS data were analyzed by selecting 3 photographs of each mouse at each time point, which were published on a Google site. Selection was random from those taken after restraint or tattooing, and, in an attempt to improve blinding after ear tagging, photographs were chosen in which the tag was least or not visible. Volunteer scorers were recruited through email and asked to rate their mouse husbandry experience. If they had used the MGS previously and had more than 5 y of experience, they were experts. Novices had no husbandry experience and no prior knowledge of the MGS. Instruction was limited to the material on the NC3Rs website.³⁶ The Google site had 336 photographs in 8 sections that could be scored in any order. The results were analyzed similarly to a previous study.⁴¹ A median score was calculated for each participant for each FAU over the 3 photographs of each mouse at each time point. Median scores for each FAU were then averaged across all participants in the expert or novice groups at each time point. These data underwent internal consistency testing (Cronbach α) to determine whether all FAU were essential to scale consistency and whether consistency was maintained across the 4 assessment times. Accordingly, at each time point, α values were calculated for novices or experts, first including all scale elements (all 5 FAU included) and subsequently with individual FAU omitted. Individual FAU were then summed to create an overall participant MGS score of each mouse at each time point. The intraclass correlation coefficient (ICC) was then used to establish whether, within their respective groups, the

MGS scores between the novices and within the experts agreed at each assessment time (2-way random model, mean of k raters, absolute agreement).²⁴ The average score for each mouse from all experts was then compared with the equivalent score from novices to determine whether the novice and expert groups provided consistent MGS scores, by recalculating ICC values at each assessment time (2-way random effects, single rater, absolute agreement). The expert and novice scores per mouse were pooled to create a global average MGS, which was used to assess procedure-related changes.

The results of light–dark testing were exported to Excel (Microsoft, Redmond, WA). The number of entrances into and movements within each chamber were calculated and then the relative preference for each chamber was determined according to total residence time. This was calculated as the proportion of time in each compartment relative to all others; for example, where t = time, preference for the black chamber was calculated as $Black_t / (Black_t + Gray_t + White_t)$. Preference for the black chamber relative to the white was determined by subtracting the white preference result from the black preference result for each mouse. The voluntary interaction footage was found to be unsuitable for automated processing, so was assessed manually and blindly according to handling method but blinded evaluation was not possible once mice had been tattooed, restrained, or ear tagged. Approach–interaction was the proportion of time during each trial that any part of the mouse excluding the tail was touching or within 1 cm of the operator’s hand or wrist, with the remainder of time being avoidance. The imaging data were processed by overlaying a 1×3-cm region of interest over the tattoo site or the same tail region in the restrained mice. Another 1×1-cm region of interest was placed over the tail-clamp site. Living Image software (version 4, PerkinElmer, Beaconsfield, Buckinghamshire, United Kingdom) was used to quantify peak signal intensity (total flux, photons/cm²/s) within each region of interest after subtraction of the background (preluminol) intensity. The proportionate numbers of tattoo or tag read errors were calculated for the expert and novice groups. All statistical analyses used SPSS software (version 23, IBM, Armonk, NY). Paired t tests were used to compare weight changes. The remaining data underwent ANOVA with procedure (tattoo or restraint), sex, and handling as between-subjects factors and time as the within-subjects (repeated) factor for the pre- to postprocedure analyses. Being derived from the same subjects, the ear tag data were compared with those after restraint by using paired t tests, but independent-samples t tests were used when comparing the ear-tagging results with those after tattooing. Regression was used to evaluate whether inflammation predicted the light–dark and voluntary interaction (primarily stress-indicative) responses relative to the pain-targeted findings (MGS and manual pain scoring). The coefficient of variation was used to determine whether there was any indication of more consistent behavioral results following tunnel handling. A coefficient of variation was calculated for each behavioral measurement type (that is, automated behavioral activity scoring and grooming bout length, grooming errors, voluntary interaction, and light–dark data) across each of the 3 postacclimation time points. The resulting sets of values were then compared between handling groups using ANOVA, with P values corrected for multiple comparisons, which was equivalent to performing individual Levene tests for equality of between-groups variance. All data are reported as mean \pm SEM.

Results

Body weight. Once procedures began, male mice were heavier than female mice (29.0 ± 1.1 compared with 22.0 ± 1.3 g; $F_{1,30} = 978, P < 0.001$). The change in body weight from before to after handling acclimation was no different between tunnel- and tail-handling groups (0.6 ± 0.1 compared with 0.5 ± 0.1 g, respectively); therefore, normal weight was maintained. Although restraint or tattooing caused no significant weight losses, these mice lost 0.6 ± 0.1 g from before until after the series of postprocedural tests ($t_{11} = 4.2, P = 0.001$) and approximately the same during the tests the next day (0.5 ± 0.1 g; $t_{11} = 10.4, P < 0.001$). Mice to be ear tagged had gained an average of 1.7 ± 0.7 g by the time of the procedure (2 wk after undergoing restraint). There was no difference in the effect of ear tagging on these mice compared with their weight changes after restraint, and no difference in the effect of ear tagging relative to tattooing, but behavioral testing again caused statistically significant although clinically negligible losses (0.9 ± 0.2 g; $t_{11} = 10, P < 0.001$).

Response to tattooing or restraint. All mice had flattened ears regardless of the ongoing procedure, but only 3 mice showed any another detectable sign of facial grimacing (that is, semi-closed eyes) during the restraint or tattooing procedures. All of these mice were male: one tail handled before restraint and one from each handling group during tattooing. In addition, one of these mice vocalized during tattooing, but none did so during restraint. Tattooing caused greater agitation than restraint (18 ± 8 compared with 6 ± 6 episodes, respectively; $F_{1,24} = 18.4, P < 0.001$). There were no significant sex-associated differences, although males more often appeared agitated during restraint than females (10 ± 8 episodes during restraint in males versus 3 ± 3 in females). However, this difference was not apparent during tattooing. Handling had no effect. Agitation due to ear tagging could not be scored because mice had to be restrained manually.

Automated behavior analysis. Postprocedural behavior changes in male mice were overall no different than those in female mice. Activity levels (Figure 3 A) declined relative to baseline after both tattooing and restraint and remained depressed for at least 24 h (time factor was significant; $F_{1,24} = 46.3, P < 0.001$). In addition, ear tagging was associated with reduced postprocedural activity compared with baseline, but this effect was no different than the previous response to restraint (2 wk earlier) and no greater than the effect of tattooing. Compared with tail-handled mice, tunnel-handled mice were generally more active at all assessment times ($F_{1,24} = 8.1, P = 0.009$). Grooming was no more frequent after tattooing than after restraint or ear tagging, and although bouts were longer after tattooing than restraint ($F_{2,48} = 6.9, P = 0.009$), this difference was present postprocedurally only and not 24 h later (Figure 3 B). In addition, grooming was more prolonged after tattooing than after ear tagging ($F_{1,30} = 4.9, P = 0.033$). The frequency and duration of grooming activities exceeded baseline after ear tagging (frequency: $t_{15} = 3.8, P = 0.001$; duration: $t_{15} = 4.7, P < 0.001$) but no differently than following restraint. The automated behavior analysis had no capacity to determine whether grooming was normal or tail-directed.

Manual behavior analysis. There were insufficient occurrences of pain-specific behavior (fewer than 2 identifiable occurrences per mouse) to conduct any meaningful analyses. The manual behavior assessment, therefore, focused on bouts of abnormal grooming (that is, errors). This behavior increased in all mice (Figure 4), but more ‘errors’ were recorded in tattooed mice compared with those that underwent restraint at both the first and 24 h assessments (time×procedure interaction, $F_{2,48} = 18, P < 0.001$). Scores after ear tagging (5.8 ± 3.2) were not significantly

different from those after restraint (4.4 ± 2.6) but were lower than those at the same time after tattooing (10.3 ± 3.5 , $F_{1,30} = 14.3$, $P = 0.001$). By 24 h, the tattooed mice were still grooming abnormally, similarly to those that had just been ear tagged. Although there were no major sex-associated differences, at 24 h restrained male mice showed more abnormal grooming than restrained females (5.5 ± 1.8 compared with 2.3 ± 1 ; $F_{1,14} = 20$, $P = 0.001$) but not tattooed males relative to tattooed female mice (8.3 ± 1.7 compared with 6.3 ± 2.5 ; $P = 0.1$). In addition, males tended to score higher than females in response to ear tagging, albeit nonsignificantly. Handling method had no significant effect on grooming behavior.

MGS. Of the 22 volunteers who responded to the initial email, 6 novices and 6 experts completed the online MGS scoring exercise. The initial reliability analysis used the Cronbach α to establish the internal validity of the 5 FAU of the MGS scale. Over the 4 assessment times, the α values for the overall scale for experts were high and had a narrow range (0.94 to 0.88), and the same was apparent for novices (0.78 to 0.87). As Table 1 shows, dropping individual FAU had little effect on the overall α values at each time point in the scores of the experts or novices, indicating that all FAU represented equally valuable MGS constructs and that none warranted exclusion. Average ICC values over each of the 4 assessment times indicated good to high agreement in the scores produced by the experts (ranging from 0.9 at the postacclimation assessment to 0.62 after ear tagging). Compared with those from experts, the novice scores indicated equal or greater consistency (ranging between 0.9 after tattooing to 0.74 after ear tagging). Because these values were not improved by removing the score of any particular participant, the 6 assessors in each group scored consistently. The ICC values obtained by comparing the collective expert and novice group scores indicated high overall agreement at each time point (ranging from 0.7 to 0.84); therefore agreement between novices and experts was generally good at all assessment times. This level of agreement indicated that the overall average MGS score was acceptable to use to assess any postprocedural, handling-associated, or sex-related MGS differences. Mice intended for tattooing tended ($P = 0.068$) to have higher preacclimation scores than those assigned for restraint. After this evaluation point, there was no overall difference in grimacing between the tattooed mice and those that were restrained (Figure 5 A). However, whereas scores relative to baseline decreased after tattooing (that is, approximately 30 min afterward), they increased after restraint ($F_{1,24} = 5.2$, $P = 0.031$), contrary to our expectation. Ear-tagged mice grimaced more than those that were tattooed ($F_{1,30} = 5$, $P = 0.033$) and seemed to be more severe compared with when these mice had been restrained. Overall MGS scores (Figure 5 B) were generally higher in tail- compared with tunnel-handled mice after handling acclimation (3.2 ± 0.3 compared with 2.3 ± 0.1 ; $t_{30} = 2.2$, $P = 0.036$) and throughout all subsequent testing ($F_{1,24} = 10.2$), $P = 0.004$). In addition, male mice grimaced more than females (3.4 ± 0.7 compared with 2.7 ; $F_{1,24} = 10.1$, $P = 0.004$; Figure 5 C), and regardless of the procedure undertaken, grimacing increased over successive test days ($F_{2,48} = 7.7$, $P = 0.001$; Figure 5).

Light–dark preference testing. The preference for each of the 3 compartments was similar regardless of whether this attribute was evaluated by using residence time, exploration, movements, or entrance counts. Therefore, the proportionate residence times were the only measure used to assess any procedure-related light–dark preference changes. The baseline (preacclimation) readings showed an initial preference bias for the white chamber in mice intended for tunnel handling. After this effect was sub-

tracted from all subsequent readings, there were no differences in chamber preference according to the procedure undertaken, sex, or handling method. This result occurred because mice largely remained in the gray start chamber. The average times spent in the gray, white, and black compartments over all trials were 367 ± 27 , 149 ± 19 , and 83 ± 15 s, respectively.

Voluntary interaction. Figure 6 shows the mean percentage of trial time that tail- or tunnel-handled mice spent engaged with the assessor's hand at each time point. Whereas tail-handled animals showed either no change or slight avoidance, approach behavior increased in tunnel-handled mice (time \times handling interaction; $F_{3,72} = 8.1$, $P < 0.001$). There was no difference depending on sex or whether mice had undergone tattooing or restraint and ear tagging. The apparent antineophobic effect of tunnel handling was long lasting and robust; apart from 2 weekly cage cleans (when differential handling was used), mice were not handled during the 2 wk between the restraint and tagging procedures, yet mice that had been tunnel handled remained considerably less fearful than those that were tail handled. These mice spent an average of $84\% \pm 2\%$ of the posttagging trial interacting with the handler's hand relative to only $10\% \pm 2\%$ in mice that had been handled by their tail ($F_{1,14} = 46$, $P < 0.0001$). Only tunnel-handled mice ever made full hand contact. On one occasion, an animal had to be retrieved from the assessor's sleeve, and another tore a glove by nonaggressive biting.

Imaging. Mice were imaged after tattooing or restraint only. The tails of tattooed mice developed inflammatory signals that initially were 5-fold greater than in restrained mice ($113 \pm 11 \times 10^3$ compared with $20.5 \pm 3.7 \times 10^3$ photons per second; $F_{1,30} = 66$, $P < 0.0001$; Figure 7 A). The following 24 h saw a proportionally similar decline after both procedures, but signals were still substantially more intense in tattooed mice than in those that were restrained ($82 \pm 14 \times 10^3$ p/s compared with $8 \pm 2 \times 10^3$ photons per second; $F_{1,30} = 26$, $P < 0.0001$). The tails of tunnel-handled mice were more inflamed than those that were tail-handled ($F_{1,24} = 6.1$, $P = 0.021$), as were those of male compared with female mice ($F_{1,24} = 4.3$, $P = 0.048$; Figure 7 B). Signals at the base of the tail (where it was fixed by the paddle clamp) exceeded background ($F_{1,24} = 103$, $P < 0.001$) similarly after restraint or tattooing. This tail-base inflammation had largely subsided by 24 h with no sex- or handling-associated differences, but at 24 h, signals were still above background ($F_{1,24} = 49$, $P < 0.001$). The tail base tended ($P = 0.054$) to become more inflamed after tattooing than restraint. There were no differences in intensity in mice tattooed with 100 as compared with 010, and signals remained stable over the 4 study weeks, suggesting the needle had not become blunted.

Regression analysis. Stepwise multiple regression was used to identify the extent to which the anxiety or pain measures predicted inflammation severity after restraint and tattooing. A significant proportion of the variation in signal intensity ($R^2 = 0.51$; $F_{2,29} = 16$, $P < 0.0001$) was predicted by the degree of agitation ($\beta = 0.48$, $t = 3.4$, $P = 0.002$) and abnormal grooming ($\beta = 0.39$, $t = 2.8$, $P = 0.008$), but only the day of the procedure and not at 24 h. This pattern is illustrated in Figure 8, where tail signal intensity is plotted against the postprocedural average frequencies of abnormal grooming and agitation. Although inflammation from tattooing might have been the major contributor, inflammation after placement of the magnetic ink slide (that is, restraint) was associated with slightly increased agitation and abnormal grooming also.

Coefficient of variation analysis. There was no significant reduction in the variability of behavior data from tunnel compared

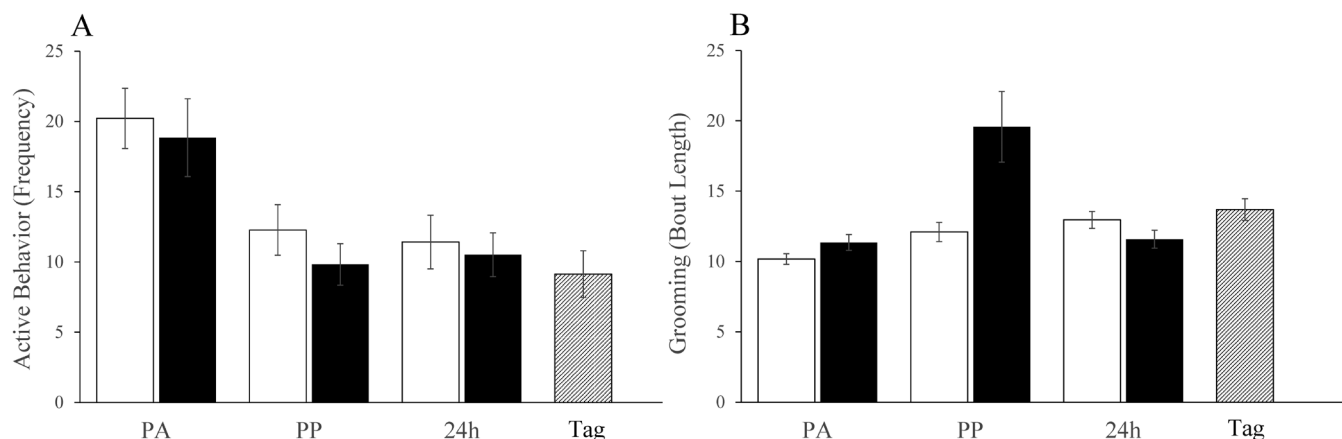


Figure 3. Results of automated behavior analysis (HCS) postacclimation (PA) to differential handling, postprocedurally (PP) and at 24 h in tattooed mice (black bars) or those that underwent restraint (white bars) followed 2 wk later by ear tagging (hatched bar; Tag). (A) Mean active behavior (rearing and walking) frequency (total in 10 min) decreased at successive time points without significant difference between procedures. (B) The mean duration (in seconds) of bouts of grooming behavior during a 10-min period. Compared with baseline (PA), bouts of grooming were significantly ($P = 0.009$) longer postprocedurally after tattooing than restraint, but restraint and ear tagging had equivalent effect.

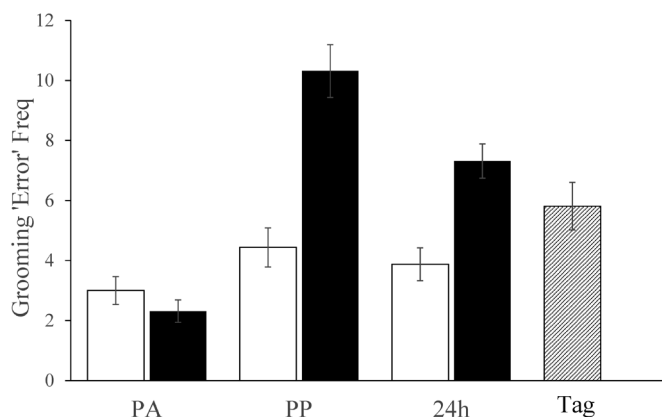


Figure 4. Manual behavior analysis results illustrating the frequency (Freq) of grooming errors (in 3 min) at each time point in tattooed mice (black bars) or those undergoing restraint (white bars) followed by ear tagging (hatched bar). Greater abnormality of grooming microstructure (indicating greater anxiety) was detected in tattooed mice lasting at least 24 h ($P < 0.001$). In addition, anxiety was indicated after ear tagging, and although its effect was no worse than after restraint, it was less than after tattooing ($P = 0.001$). PA, postacclimation; PP, postprocedure; Tag, after ear tagging.

with tail-handled mice. The largest group score separation came from the voluntary interaction test where tunnel handled mice had a lower coefficient estimate of 0.61 ± 0.16 compared with 0.78 ± 0.2 in mice handled by the tail.

Identification exercise. All tattoos were read correctly without handling, often without removal of the cage food and water hopper, but all volunteers needed to lift mice to read their ear tags. All but 2 volunteers used tail lifting followed by supporting the mouse on their forearms. The 2 that did not were in the novice group and chose to use a tunnel. The percentage of tag numbers read incorrectly was 45% in the experienced group compared with 25% in the novice group. Figure 9 A shows a mouse with ear tag number 799, and Figure 9 B shows a mouse tattooed with 100.

Discussion

Our intention was to evaluate a new, partially automated system for tattooing mice that was said to be a more welfare-friendly option than either manual tattooing or ear tagging, and one offering scientific advantages making it currently the

best method for permanent mouse identification.²⁰ The study was designed to reveal the relative effects of each method on welfare (that is, pain or anxiety) as well as less immediately apparent concerns or benefits, such the need for handling and ease of identification. Tattooed mice were compared with those undergoing restraint and then ear tagging. Pain was assessed using the MGS and both automated and manual analysis of behavior recordings, and anxiety was evaluated using light-dark choice⁴ and voluntary interaction¹⁹ testing. Because nonaversive (tunnel) handling is now thought to be an effective means of reducing anxiety in mice^{14,15,19} and because this practice might allow mice to cope better in circumstances eliciting pain or anxiety, we also sought to establish whether using this method might lessen any welfare-related problems arising through the application of restraint, tattooing, or ear tagging. According to their welfare effects and potential to harm study validity, we had doubts about the acceptability of both ear tagging and tattooing, but contrary to both popular opinion^{8,28,44} and our initial expectations, tattooing was not found to be substantially more harmful than ear tagging.

All mice struggled throughout the approximately 30-s automated tattooing cycle, and although this behavior was much less vigorous during restraint, mice still showed some obvious distress. Postprocedural testing began with a 10-min recording for automated behavior analysis and then a further 3 min of recording for closer manual inspection. The resulting automated behavioral analysis data showed that all mice became less active (Figure 3 A), and bouts of grooming behavior, although not more frequent, became longer after tattooing than after restraint or ear tagging (Figure 3 B). We have reported similar effects as evidence of pain after surgery,^{41,45} but, only because they were accompanied by more convincing evidence during detailed manual analysis. Considering how different the current procedures were, it was not surprising that few such signs were recorded. The manual assessment included the grooming microstructure, where cephalocaudal sequence errors are thought to indicate anxiety or stress more insightfully than gross measures of grooming frequency or duration.^{22,23} Accordingly, although the increase in the length of grooming bouts evident from the automated behavioral analysis subsided 24 h after tattooing, errors continued to be detected until 24 h after tattooing, longer than after either restraint or ear tagging (Figure 4). If these errors accurately reflected anxiety or stress, this result suggested

Table 1. Results of Cronbach α internal consistency testing of the MGS data for all expert or novice assessors at each time point

	Overall Cronbach α	Cronbach α when listed facial action unit is deleted				
		Cheek	Ear	Nose	Orb	Whisker
Experts						
After acclimation	0.922	0.881	0.931	0.891	0.917	0.899
After procedure	0.93	0.907	0.927	0.905	0.921	0.914
At 24 h	0.885	0.839	0.906	0.836	0.88	0.837
After ear tagging	0.942	0.915	0.943	0.921	0.936	0.929
Novices						
After acclimation	0.841	0.831	0.849	0.768	0.789	0.8
After procedure	0.872	0.852	0.829	0.84	0.865	0.843
At 24 h	0.783	0.718	0.796	0.744	0.752	0.703
After ear tagging	0.877	0.836	0.85	0.819	0.885	0.872

Cheek, cheek bulge; ear, ear position; nose, nose bulge; orb, orbital tightening; whisker, whisker change

The data show overall scale consistency (proportion of explained variability) for each group and the extent to which consistency changed after deleting each facial action unit. Larger values relative to the overall Cronbach α estimate indicate facial action units with lowered validity compared with others; for example, deleting ear position from the scores produced by experts at 24 h improved scale consistency from 0.885 to 0.906.

that stress or anxiety was more severe and lasted longer after tattooing than after restraint or ear tagging.

The MGS has revealed pain in mice in a variety of different nociceptive assays,²⁵ after surgery,^{11,26,27,31,41} and during disease development,^{9,35} so we initially considered it the assessment technique with the greatest potential to establish whether pain occurred during the restraint or tattooing procedures. However, the mice were partially obscured by the red paddle cover of the tattooing apparatus and were more often 'tail-looking' rather than looking at the camera, so the only MGS data that could be used for pain assessment were those obtained from photographs taken 20 to 25 min after the procedures and on the following day. These data underwent rigorous reliability analysis. The first stage was to assess the internal validity of the MGS scale elements (Cronbach α) and then to determine the consistency of scores produced by the groups of 6 volunteer expert or novice scorers (ICC analysis). Although most authors cite that Cronbach values should be close to 0.7, in important decision-making it has been argued that the value should exceed 0.8.³⁸ There was only one instance where α values fell below 0.8 (0.78 in the novice scores at 24 h after tattooing), therefore all FAU were included in the subsequent ICC analyses. This evaluation found consistent agreement both within and between the novice and experts groups. This finding suggested that prior MGS scoring experience was largely immaterial and that the online MGS instructions had provided sufficient training. Experience was ignored therefore, and the overall average MGS score was used to assess procedure-related effects. From the assumption that the MGS accurately reflects pain, the main result was that mice were no different after tattooing or restraint but were more painful after ear tagging (Figure 5 A). This outcome contradicts our expected result and possibly previous MGS findings after ear notching, where ear-notched mice showed either no response or one so momentary that the authors thought it might have been missed.³² Although our MGS photographs were taken longer after each procedure, 20 to 25 min was still within a time-frame whereby the effect of nociceptive stimuli of moderate duration of effect can supposedly be captured.²⁵ Unless ear tagging or notching have vastly different effects, perhaps the response to ear notching was missed previously, but not by recording too late but because the photographs were taken before any nociception or painful inflammation developed. One might speculate that ear tagging was presently perceived as more harmful or painful than ear notching or punching due to the prolonged

irritation by the tag once the procedure ended. Alternatively, our analysis might have been more accurate due to our greater number of scorers, thereby reducing the effect of any false negatives or positives.¹³

In addition, the attempt to blind the scorers regarding the presence of an ear tag might have failed, leading to increased pessimism and thus elevated scores. To gauge whether this happened, the overall MGS scores were recomputed with ear position excluded and although they remained highest after ear tagging, more widespread bias regarding the appearance of these mice cannot be discounted. Resolving this issue would require a carefully designed direct comparison between ear-notched and -tagged mice, which was not possible with the resources available for the current study. Furthermore, MGS scores increased over successive test sessions, suggesting, as mentioned in the original article, that this scale may be stress-sensitive.²⁵ Given that MGS scores were significantly lower in tunnel-handled mice (Figure 5 B), it was tempting to say the mice were made less anxious, but this effect was not seen in differentially handled CBA or DBA/2 mice.³⁴ The MGS has been shown to be strain-dependent and sex-sensitive,³² and as in the latter study, we also found that male mice grimaced more than females (Figure 5 C). However, in both cases, there was a troubling degree of score variation even before the tattooing procedures began (Figure 5 A, preacclimation time point). The effectiveness of the MGS likely depends on pain duration and intensity,²⁵ on recording methodology (that is retrospective or cageside),³² whether the procedure requires anesthesia,³⁰ and probably what the ongoing behavioral activity of the mouse is.²⁵ It therefore seems too susceptible to extraneous influences, echoing the concerns of other authors about its face validity.¹³ Therefore, although the current MGS results had high internal consistency and reliability, it remains uncertain whether they accurately represented postprocedural pain or stress/anxiety, or showed that this was lessened by tunnel handling.

Nonaversive handling was used to determine whether any anxiety or potentially pain-related outcomes could be blunted if not prevented, given that this relatively simple husbandry refinement can have a profoundly beneficial effect by reducing anxiety^{6,14,15,19} and can even optimize the consistency and reproducibility of research findings.^{12,18} As in the cited studies, the voluntary interaction assessment provided the clearest illustration of the benefits of tunnel handling (Figure 6). Tunnel-handled mice were substantially more interactive, in some cases

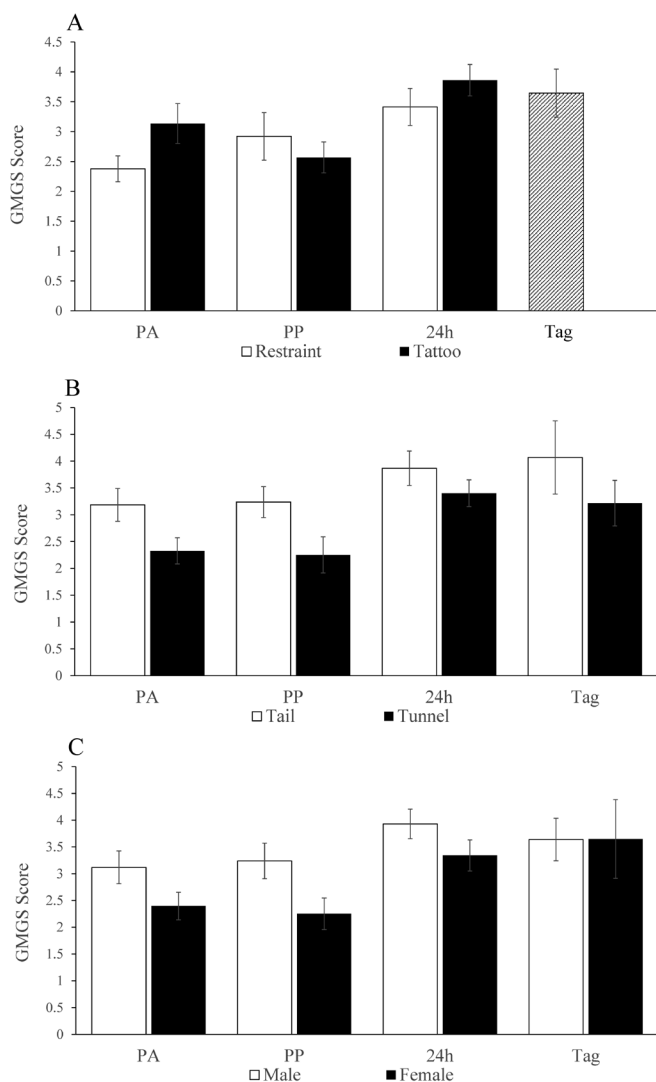


Figure 5. (A) Mean Global Mouse Grimace Scale (GMGS) score (all FAU) at each assessment time after each procedure, showing no overall difference between between the restraint (white bars) or tattooed (black bars) groups but more severe grimacing postprocedurally following ear-tagging than following tattooing (Tag compared PP; $P = 0.033$). (B) Mice handled by the tail grimaced more severely than those that were tunnel-handled, both before the tattooing or restraint procedures (PA) and at each subsequent time point, including when those mice that were restrained were ear-tagged ($P = 0.004$). (C) Male mice grimaced more than female mice. ($P = 0.004$). PA, postacclimation; PP, postprocedure; Tag, after ear tagging.

showing full-sustained hand contact, which was not seen in any tail-handled animal. Although the analysis of the coefficient of variation did not reveal any increased precision in the behavior data after tunnel handling, these mice were generally more explorative, suggesting an antineophobic effect that could be helpful to studies concerned with novelty testing.¹⁵ In addition, the data in Figure 6 suggest that tunnel handling overcame anxiety-like behavior after restraint or tattooing, perhaps in a similar way to how it dampens the effect of an aversive event, such as scruff restraint.¹⁹ Although differential handling did not occur daily during the 2 wk prior to ear tagging but was used only during cage cleaning, the tunnel-handled mice remained more willing to interact than mice that had been tail handled. The effect was therefore long-lasting also.

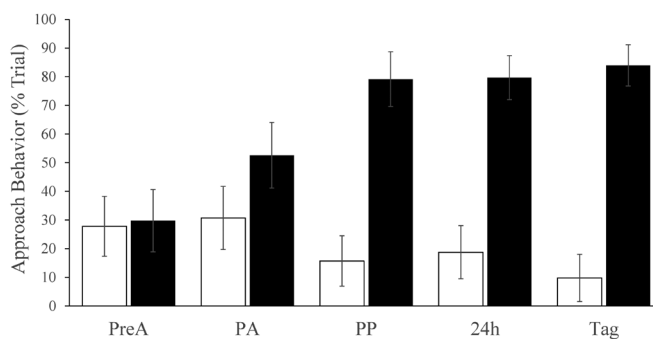


Figure 6. The percentage of time during the 3-min voluntary interaction trials when mice assigned to the tail or tunnel handling groups (respectively, white bars and black bars) were engaged in approach behavior. There was no significant difference in interaction times in mice that were tattooed compared with those that were restrained (so these data were combined). There was no difference prior to handling acclimation (PreA), but after 7 d of tunnel handling (PA), and after tattooing or restraint (PP, 24 h) or after ear tagging (Tag), the percentage of time spent interacting with the handler's hand (contacting or being within 1 cm of the hand or wrist) was significantly greater in mice that had been tunnel handled ($P < 0.001$); indicating that the tunnel handled mice were persistently less timid. White bars, tail handling; black bars, tunnel handling.

Imaging was used to assess inflammation and to identify any relationship between inflammation and pain severity. Using luminol to image inflammatory responses¹⁶ also provided a financially more cost-effective and possibly refined approach to real-time monitoring than in our previous COX2 study.⁴¹ Whereas all mice reacted badly to the COX2 probe injection and 2 died, mice responded to luminol injection as to any other innocuous substance. Signals were predictably strongest at the site of tattoo application (Figure 7 A), and inflammation severity was aligned with how agitated mice became during tattooing and the degree to which grooming microstructure was altered (Figure 8). This finding was surprising in that it suggested that the tails of some mice were injured more severely than others. Because the tattooing device we used applies ink to a precise depth, inflammatory responding should have been relatively uniform given the correct needle and paddle size. Albeit only marginally, the tails of male mice became more inflamed than those of female mice, as did those in mice that were tunnel handled (Figure 7 B). The sex-associated difference could have arisen if the male tails had been thicker than those of females, resulting in the needle going proportionately deeper in male mice. Although more precise measurements could have been taken, the tails were relatively uniform according to the data taken for paddle and needle size selection. Effects were not due to needle blunting or another longitudinal effect, because consistent signals were obtained over 4 wk, and the design had equal numbers of male and female and tail- or tunnel-handled mice that were tattooed on alternate days each week (Figure 1). The more intense inflammation in male mice was, therefore, more likely due to greater sensitivity. It is possible but unlikely that female progesterone had an antiinflammatory effect, through decreased production of cytokines.¹ Why tunnel handling resulted in more inflammation might be explained by improved tail circulation and consequently increased luminol distribution through lack of handling. If that scenario were true, given the lack of other negative effects of tunnel handling, it would seem to strengthen the argument that it is preferable to tail handling. In addition, there were stronger than expected signals at the point of contact with the ink slide in mice undergoing restraint,

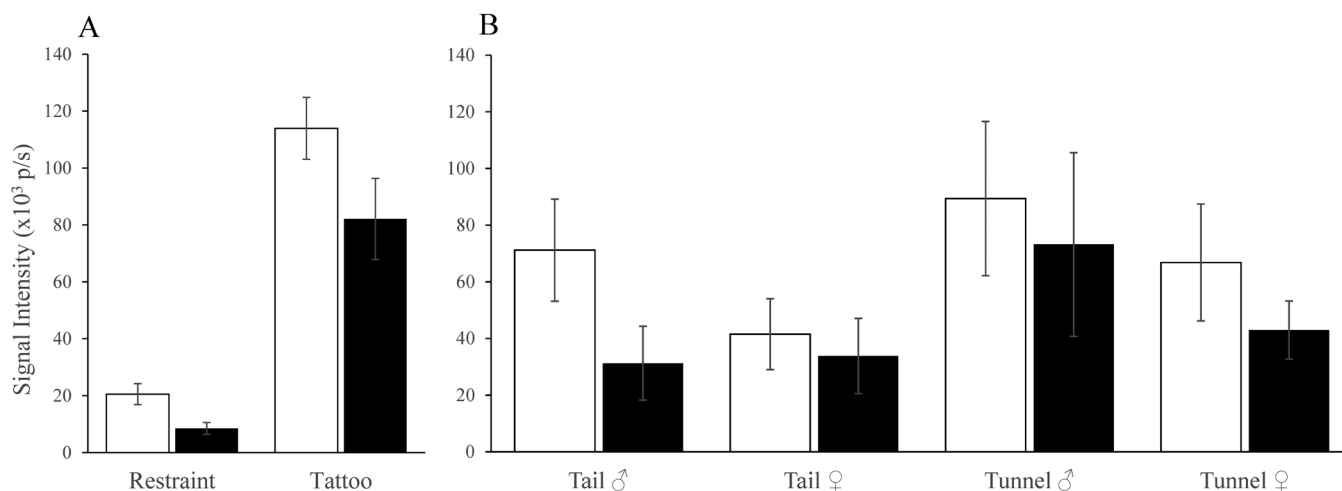


Figure 7. (A) Bioluminescent signal (inflammation) intensity ($\times 10^3$ photons per second [p/s]) from the tails of mice on the day of tattooing or restraint (white bars) and 24 h later (black bars). Inflammation was substantially more intense in tattooed mice postprocedurally ($P < 0.0001$), but above-background signals also occurred after restraint ($P < 0.001$). Inflammation resolved to a proportionately similar extent after each procedure, but the tails of tattooed mice remained more inflamed ($P < 0.0001$) at 24 h. (B) Inflammatory signal intensity in tail- or tunnel-handled male or female mice postprocedurally (white bars) and at 24 h (black bars). The tails of restrained or tattooed tunnel-handled mice were more inflamed ($P = 0.021$), and the tails of male mice were slightly more inflamed than those of female mice ($P = 0.048$).

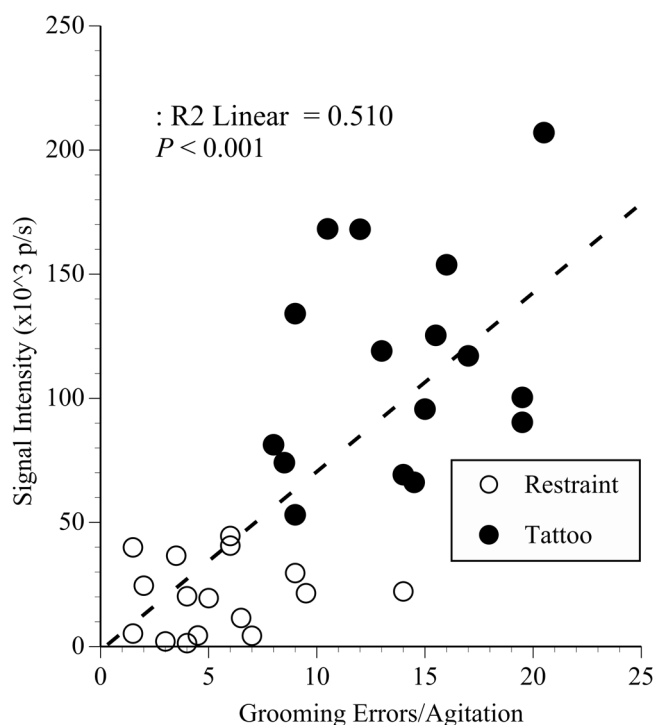


Figure 8. The results of regression analysis of tail signal intensity ($\times 10^3$ photons per second [p/s]) after restraint or tattooing plotted against the average frequency of agitation and grooming errors. Mice that struggled most and groomed poorly developed more inflamed tails, especially after tattooing; suggesting agitation, and grooming microstructure predicted the degree of anxiety caused by tattooing.

presumably due to the pressure points created by the ink slide magnet. Signals were registered at the point of application of the tail-clamp also. Removing these pressure points seems an obvious means of refining the Lapstamp.

The light–dark test results were disappointing in that they did not reveal differential responses to either the identification procedures or handling methods. The persistent preference of mice for the gray chamber suggested a strong phobia for both

the white and black compartments. The apparatus was a modified version that used previously,³ with a start compartment between 2 equal-sized light and dark boxes rather than only 2 differently lit boxes; an arrangement chosen to avoid bias due to mice having no initial choice.¹⁰ Given that male and female mice had separate equipment that was cleaned meticulously, the likeliest reason for their lack of exploration was that the choice compartments had the original grid or steel rod floors fitted with plastic-covered cardboard sheets. These materials might have been aversive for olfactory or tactile reasons; in hindsight, we might have used more robust coverings, both black and white, or used an elevated plus maze for that stage of anxiety testing.

The volunteer ear-tag and tattoo readers had varied prior husbandry experience, but none had previously used ear tags or tattoos. Although our setup was supposed to represent routine mouse identification, under normal circumstances those responsible for the task would have anticipated the codes that they would be presented with (for example, on a cage label). In addition, there were only 2 tattoo codes to read (010 or 100), which were the same across all 4 cages, making reading more predictable. A better design would have had ear tags or tattoos bearing the same sequential numbers; however, we considered standardizing the degree of tail injury in the tattooed mice to be more essential to study success. Nevertheless, the ear-tag recognition error rate was sufficiently high ($\geq 25\%$) that it might have prompted serious concerns regarding validity in another type of research study. The most frequent errors were when numbers were read inverted and backward, for example, 799 was read as 661 (Figure 9 A) and 798 as 861. Using these numbers in the current study was somewhat unfortunate but a result of random selection. Furthermore, the identification exercise highlighted potential scientific concerns because all ear-tagged mice had to be handled, as would likely be the case routinely. In addition, one tag fell out after the study ended. By contrast, all tattoos were read correctly without handling of mice (Figure 9 B) and often before removal of the cage hopper, both of which are attributes that would be practically advantageous for both welfare reasons and in circumventing unnecessary stress.

Our within-subjects design, whereby the mice that initially underwent restraint were then ear tagged, was beneficial in re-

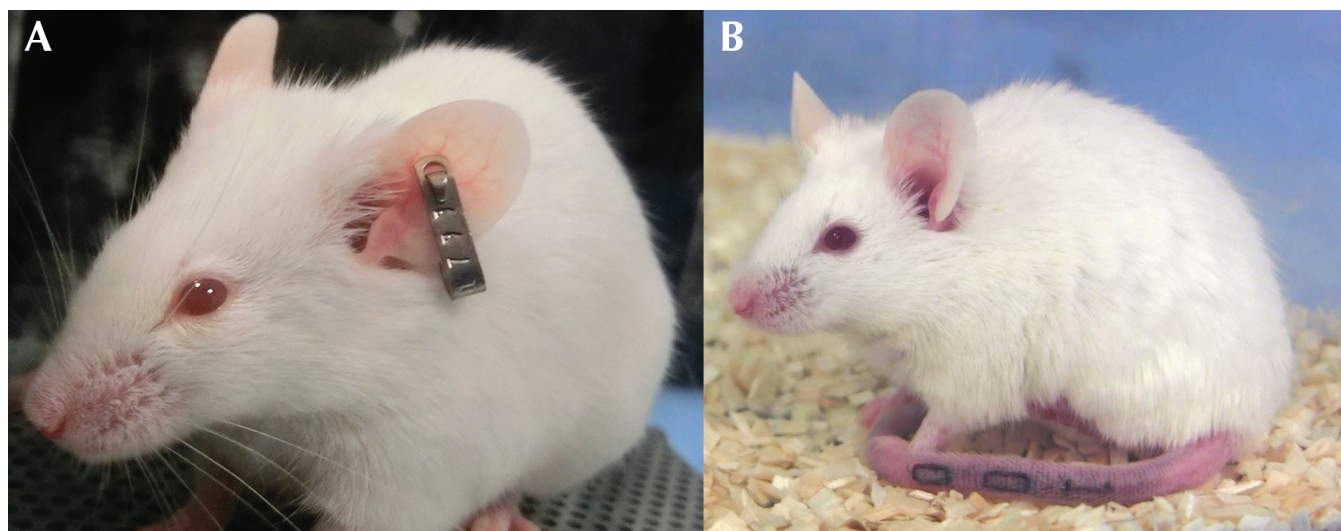


Figure 9. (A) A mouse bearing an ear tag numbered 799, which was sometimes read in reverse and inverted, that is, misidentified and recorded as 661. (B) A mouse tattooed with the digits 100, to illustrate how the tail tattoos were more easily read and correctly identified.

ducing animal numbers but could have introduced bias through escalating anxiety. The mice lost weight, not as a result of the restraint or identification procedures, but due to the burden of sequential testing. The mice underwent a sequence of different behavioral assays (both pain-orientated and those meant to detect anxiety). This repetitive testing was done to try to gain clear evidence regarding the relative effects of identification methods on mouse welfare. Given that no mouse lost more than 1 g, we felt there were no major welfare concerns arising from this. Although mice could have lost this weight via normal defecation or urination, other potential evidence of escalating stress through repeated testing was present, such as the increasing MGS scores (Figure 5). Therefore, in any future testing of the tattooing system we used here or a similar device, we would limit exposure to only the most effective assessment methods. According to our current results, we would increase our focus on manual behavior analyses and possibly combine these with naturalistic behavior assessments, which can be undertaken in the home cage.²¹

The current study highlights the difficulty of choosing any method of identification based on the balance of welfare and scientific concerns. Our data suggested that tattooing was not necessarily more painful than ear tagging but was more stressful. The anxiogenic effects of tattooing were longer-lasting than those of restraint or ear tagging, but probably persisted for only 1 to 2 d rather than several days. Quantitative assessment of the relative welfare compared with scientific benefits or disadvantages of different identification methods is a challenging task. However, the longer-term benefits of tattooing are derived from the application of a truly permanent easily readable mark. Reducing both the need to handle mice and the chances of misidentification would seem to balance concerns regarding increased anxiety or stress or even any brief, initial pain caused by tattooing. Overall, we conclude that there is probably little justification for choosing ear tagging over tattooing. Improvements to the tattooing device that promote better overall utility would be modifications that allow the mice to be held securely but less intrusively, make the device less noisy, and allow the mice to be seen more clearly during the procedure so that their wellbeing could be assessed more effectively.

Acknowledgments

The work was funded by the Institute of Neuroscience, Newcastle University. We thank Saimir Luli for assistance during imaging, Charlie Batchelor for designing the MGS website, and Jennifer Murray who placed the ear tags. The Wellcome Trust provided the IVIS machine (grant number 087961).

References

1. Aksoy AN, Tokar A, Celik M, Aksoy M, Halici Z, Aksoy H. 2014. The effect of progesterone on systemic inflammation and oxidative stress in the rat model of sepsis. *Indian J Pharmacol* **46**:622–626. <https://doi.org/10.4103/0253-7613.144922>.
2. Balcombe JP, Barnard ND, Sandusky C. 2004. Laboratory routines cause animal stress. *Contemp Top Lab Anim Sci* **43**:42–51.
3. Belzung C, Le Pape G. 1994. Comparison of different behavioral test situations used in psychopharmacology for measurement of anxiety. *Physiol Behav* **56**:623–628. [https://doi.org/10.1016/0031-9384\(94\)90311-5](https://doi.org/10.1016/0031-9384(94)90311-5).
4. Bourin M, Hascoët M. 2003. The mouse light–dark box test. *Eur J Pharmacol* **463**:55–65. [https://doi.org/10.1016/S0014-2999\(03\)01274-3](https://doi.org/10.1016/S0014-2999(03)01274-3).
5. Castelhana-Carlos MJ, Sousa N, Ohl F, Baumans V. 2010. Identification methods in newborn C57BL/6 mice: a developmental and behavioural evaluation. *Lab Anim* **44**:88–103. <https://doi.org/10.1258/la.2009.009044>.
6. Charles River Laboratories. [Internet]. 2017. UK health reports by species. [Cited: 14 January 2018]. Available at: <https://www.criverson.com/HealthData/uk/H62A520M.pdf>.
7. Clarkon JM, Dwyer DM, Flecknell PA, Leach MC, Rowe C. 2018. Handling method alters the hedonic value of reward in laboratory mice. *Sci Rep* **8**:1–8. <https://doi.org/10.1038/s41598-018-20716-3>.
8. Dahlborn K, Bugnon P, Nevalainen T, Raspa M, Verboost P, Spangenberg E. 2013. Report of the Federation of European Laboratory Animal Science Associations Working Group on Animal Identification. *Lab Anim* **47**:2–11. <https://doi.org/10.1177/002367712473290>.
9. Duffy SS, Perera CJ, Makker PGS, Lees JG, Carrive P, Moalem-Taylor G. 2016. Peripheral and central neuroinflammatory changes and pain behaviors in an animal model of multiple sclerosis. *Front Immunol* **7**:1–16.
10. Ennaceur A. 2014. Tests of unconditioned anxiety—pitfalls and disappointments. *Physiol Behav* **135**:55–71. <https://doi.org/10.1016/j.physbeh.2014.05.032>.
11. Faller KME, McAndrew DJ, Schneider JE, Lygate CA. 2015. Refinement of analgesia following thoracotomy and experimental myocardial infarction using the Mouse Grimace Scale. *Exp Physiol* **100**:164–172. <https://doi.org/10.1113/expphysiol.2014.083139>.

12. Ghosal S, Nunley A, Mahbod P, Lewis AG, Smith EP, Tong J, D'Alessio DA, Herman JP. 2015. Mouse handling limits the impact of stress on metabolic endpoints. *Physiol Behav* **150**:31–37. <https://doi.org/10.1016/j.physbeh.2015.06.021>.
13. Golledge HD, Thomas A. 2016. Score sheets and analgesia. *Lab Anim* **50**:411–413.
14. Gouveia K, Hurst JL. 2013. Reducing mouse anxiety during handling: effect of experience with handling tunnels. *PLoS One* **8**:1–8. <https://doi.org/10.1371/journal.pone.0066401>.
15. Gouveia K, Hurst JL. 2017. Optimising reliability of mouse performance in behavioural testing: the major role of nonaversive handling. *Sci Rep* **7**:1–12. <https://doi.org/10.1038/srep44999>.
16. Gross S, Gammon ST, Moss BL, Rauch D, Harding J, Heinecke JW, Ratner L, Piwnica-Worms D. 2009. Bioluminescence imaging of myeloperoxidase activity in vivo. *Nat Med* **15**:455–461. <https://doi.org/10.1038/nm.1886>.
17. Holman C, Piper SK, Grittner U, Diamantaras AA, Kimmelman J, Siegerink B, Dirnagl U. 2016. Where have all the rodents gone? The effects of attrition in experimental research on cancer and stroke. *PLoS Biol* **14**:1–12. <https://doi.org/10.1371/journal.pbio.1002331>.
18. Hurst J. [Internet]. 2014. Taming anxiety and variation in laboratory mice. [Cited 11 January 2018]. Available at: <https://www.nc3rs.org.uk/taming-anxiety-and-variation-laboratory-mice>.
19. Hurst JL, West RS. 2010. Taming anxiety in laboratory mice. *Nat Methods* **7**:825–826. <https://doi.org/10.1038/nmeth.1500>.
20. Somark Innovations. [Internet]. 2015. Mouse identification matters: LabStamp ensures better science and better business. [Cited 14 January 2018]. Available at: <https://crackit.org.uk/mouse-identification-matters-labstamp-ensures-better-science-and-better-business>.
21. Jirkof P. 2014. Burrowing and nest building behavior as indicators of wellbeing in mice. *J Neurosci Methods* **234**:139–146. <https://doi.org/10.1016/j.jneumeth.2014.02.001>.
22. Kalueff AV, Tuohimaa P. 2004. Grooming analysis algorithm for neurobehavioural stress research. *Brain Res Brain Res Protoc* **13**:151–158. <https://doi.org/10.1016/j.brainresprot.2004.04.002>.
23. Kalueff AV, Tuohimaa P. 2005. Mouse grooming microstructure is a reliable anxiety marker bidirectionally sensitive to GABAergic drugs. *Eur J Pharmacol* **508**:147–153. <https://doi.org/10.1016/j.ejphar.2004.11.054>.
24. Koo TK, Li MY. 2016. A guideline of selecting and reporting intra-class correlation coefficients for reliability research. *J Chiropr Med* **15**:155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>. Erratum. *J Chiropr Med* 2017. <https://doi.org/10.1016/j.jcm.2017.10.001>
25. Langford DJ, Bailey AL, Chanda ML, Clarke SE, Drummond TE, Echols S, Glick S, Ingrao J, Klassen-Ross T, Lacroix-Fralish ML, Matsumiya L, Sorge RE, Sotocinal SG, Tabaka JM, Wong D, van den Maagdenberg AMJM, Ferrari MD, Craig KD, Mogil JS. 2010. Coding of facial expressions of pain in the laboratory mouse. *Nat Methods* **7**:447–449. <https://doi.org/10.1038/nmeth.1455>.
26. Leach MC, Klaus K, Miller AL, Scotto di Perrotolo M, Sotocinal SG, Flecknell PA. 2012. The assessment of post-vasectomy pain in mice using behaviour and the Mouse Grimace Scale. *PLoS One* **7**:1–9. <https://doi.org/10.1371/journal.pone.0035656>.
27. Matsumiya LC, Sorge RE, Sotocinal SG, Tabaka JM, Wieskopf JS, Zaloum A, King OD, Mogil JS. 2012. Using the Mouse Grimace Scale to reevaluate the efficacy of postoperative analgesics in laboratory mice. *J Am Assoc Lab Anim Sci* **51**:42–49.
28. Mazlan N, López-Salesansky N, Burn C, Wells D. 2014. Mouse identification methods and potential welfare issues: a survey of current practice in the UK. *Animal Technology and Welfare* **13**:1–10.
29. Mercer M. [Internet]. 2017. Explosion in tattooing, piercing tests state regulators. [Cited 14 June 2017]. Available at: <https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2017/06/14/explosion-in-tattooing-piercing-tests-state-regulators>
30. Miller A, Kitson G, Skalkoyannis B, Leach M. 2015. The effect of isoflurane anaesthesia and buprenorphine on the mouse grimace scale and behaviour in CBA and DBA/2 mice. *Appl Anim Behav Sci* **172**:58–62. <https://doi.org/10.1016/j.applanim.2015.08.038>.
31. Miller AL, Kitson GL, Skalkoyannis B, Flecknell PA, Leach MC. 2016. Using the mouse grimace scale and behaviour to assess pain in CBA mice following vasectomy. *Appl Anim Behav Sci* **181**:160–165. <https://doi.org/10.1016/j.applanim.2016.05.020>.
32. Miller AL, Leach MC. 2014. Using the Mouse Grimace Scale to assess pain associated with routine ear notching and the effect of analgesia in laboratory mice. *Lab Anim* **49**:117–120. <https://doi.org/10.1177/0023677214559084>.
33. Miller AL, Leach MC. 2015. The Mouse Grimace Scale: A clinically useful tool? *PLoS One* **10**:1–10. <https://doi.org/10.1371/journal.pone.0136000>.
34. Miller AL, Leach MC. 2016. The effect of handling method on the Mouse Grimace Scale in 2 strains of laboratory mice. *Lab Anim* **50**:305–307. <https://doi.org/10.1177/0023677215622144>.
35. Mittal A, Gupta M, Lamarre Y, Jahagirdar B, Gupta K. 2016. Quantification of pain in sickle mice using facial expressions and body measurements. *Blood Cells Mol Dis* **57**:58–66. <https://doi.org/10.1016/j.bcmd.2015.12.006>.
36. NC3Rs. [Internet]. 2018. Mouse Grimace Scale (MGS): The Manual. [Cited 01 January 2017]. Available at: <https://www.nc3rs.org.uk/mouse-grimace-scale>.
37. Nedved L. 2013. Evaluating the need for inhalant anesthesia for tail tattooing in mice. AALAS National Meeting. Baltimore, Maryland. 27–31 October 2013. *J Am Assoc Lab Animal Sci* **52**:617.
38. Nunnally J, Bernstein I. 1994. *Psychometric theory*, 3rd ed. New York: McGraw-Hill.
39. Parker RMA, Browne WJ. 2014. The place of experimental design and statistics in the 3Rs. *ILAR J* **55**:477–485. <https://doi.org/10.1093/ilar/ilu044>.
40. Rajnoch C, Ferguson S, Metcalfe AD, Herrick SE, Willis HS, Ferguson MWJ. 2003. Regeneration of the ear after wounding in different mouse strains is dependent on the severity of wound trauma. *Dev Dyn* **226**:388–397. <https://doi.org/10.1002/dvdy.10242>.
41. Roughan JV, Bertrand HG, Isles HM. 2015. Meloxicam prevents COX2-mediated post-surgical inflammation but not pain following laparotomy in mice. *Eur J Pain* **20**:231–240. <https://doi.org/10.1002/ejp.712>.
42. Sadler AM, Bailey SJ. 2016. Repeated daily restraint stress induces adaptive behavioural changes in both adult and juvenile mice. *Physiol Behav* **167**:313–323. <https://doi.org/10.1016/j.physbeh.2016.09.014>.
43. United Kingdom Legislation. [Internet]. 1986. *Animals Scientific Procedures Act*. [Cited: 14 January 2019] Available at: http://www.legislation.gov.uk/ukpga/1986/14/pdfs/ukpga_19860014_en.pdf
44. Wever KE, Geessink FJ, Brouwer MAE, Tillema A, Ritskes-Hoitinga M. 2017. A systematic review of discomfort due to toe or ear clipping in laboratory rodents. *Lab Anim* **51**:583–600. <https://doi.org/10.1177/0023677217705912>.
45. Wright-Williams S, Flecknell PA, Roughan JV. 2013. Comparative effects of vasectomy surgery and buprenorphine treatment on faecal corticosterone concentrations and behaviour assessed by manual and automated analysis methods in C57 and C3H mice. *PLoS One* **8**:1–13. <https://doi.org/10.1371/journal.pone.0075948>.
46. Wright-Williams SL, Courade JP, Richardson CA, Roughan JV, Flecknell PA. 2007. Effects of vasectomy surgery and meloxicam treatment on faecal corticosterone levels and behaviour in 2 strains of laboratory mouse. *Pain* **130**:108–118. <https://doi.org/10.1016/j.pain.2006.11.003>.