Neuroscience Applied 3 (2024) 104058

Contents lists available at ScienceDirect

Neuroscience Applied

journal homepage: www.journals.elsevier.com/neuroscience-applied

Review

Assessment of quality of life and wellbeing in mouse preclinical research – A scoping review

A. Sanz-Moreno [a,*,1], P. da Silva-Buttkus [a,1], C.B. Terwee [b], M. Raess [c], H. Fuchs [a], V. Gailus-Durner [a], M. Hrabě de Angelis [a,d,e]

[a] Institute of Experimental Genetics, German Mouse Clinic, Helmholtz Zentrum München, 85764, Neuherberg, Germany
[b] Department of Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, 1081 HV, Netherlands
[c] INFRAFRONTIER GmbH, 85764, Neuherberg, Germany
[d] German Center for Diabetes Research (DZD), 85764, Neuherberg, Germany
[e] School of Life Sciences Weihenstephan, Technische Universität München, 85354, Freising, Germany

ARTICLE INFO

Keywords:
Quality of life
Wellbeing
Mice
Scoping review
Translational research

ABSTRACT

Mouse preclinical research is of great scientific interest to understand the mechanisms of human diseases and test potential therapeutic interventions. Researchers characterize biological and physiological traits, behaviors and disease symptoms using standardized phenotypic protocols in the context of *in vivo* mouse studies. However, the procedures applied do not always fully translate to reported outcomes in clinical trials. Quality of life (QoL) and wellbeing (WB) are particularly relevant outcomes in human medicine in general, and in neurology in particular, that are routinely measured by patient self-reports but rarely monitored in mouse research. In this novel scoping review, we have identified and described the instruments/tests and outcomes used to assess QoL and WB in recent mouse research (spanning 13 years). We found that WB was stated to be measured more frequently in murine studies (77 publications fulfilled our selection criteria) than QoL (only 13 articles). Instruments measuring WB were commonly used in neurology but less frequently in behavior and psychiatric research articles. Interestingly, we found a high variability of QoL and WB instruments/tests used as well as outcomes measured in the reviewed mouse studies. In addition, among similar parameters tested, we observed variable methodological procedures and mouse sample sizes. Thus, there is a lack of consensus on how to measure QoL and WB in the mouse research field. For ensuring a better translation from mouse to human, outcomes that are important in clinical trials (e.g., QoL and WB) should be measured in mouse studies. Finally, we would like to point out that a proper standardization of QoL and WB assessment protocols, for instance through a modified Delphi consultation survey, should be pursued by the mouse research community.
Review registration: The study was registered on the PROSPERO Database (registration number CRD42018103507)

1. Introduction

In biomedical clinical research, measuring quality of life (QoL) and wellbeing (WB) outcomes is of imperative importance to include patients' concerns in the evaluation of the impact of disease and the repercussions of interventions, including benefits as well as undesirable effects influencing therapeutic decision-making (Haraldstad et al., 2019; van Agteren et al., 2021). Overall, the burden of disability caused by neurological diseases has risen in the past decades (GBD 2016 Neurology Collaborators, 2019) and therefore, focus on understanding QoL and WB instruments may facilitate interventions in this area (Ziegeler et al., 2023). Generally, QoL and WB outcomes are measured through self-report questionnaires, also called patient-reported outcome measures (PROMs) (Pouwer, Snoek, van der Ploeg, Adèr and Heine, 2000; Sears et al., 2014). According to the World Health Organization (WHO), QoL is defined as *"individuals' perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept incorporating in a complex way the persons' physical health,*

* Corresponding author.
E-mail address: adrian.sanzmoreno@helmholtz-munich.de (A. Sanz-Moreno).
[1] contributed equally to this work.

https://doi.org/10.1016/j.nsa.2024.104058
Received 15 December 2023; Received in revised form 17 February 2024; Accepted 11 March 2024
Available online 19 March 2024
2772-4085/© 2024 The Authors. Published by Elsevier B.V. on behalf of European College of Neuropsychopharmacology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Abbreviations

QoL	quality of life
WB	wellbeing

psychological state, level of independence, social relationships, personal beliefs and their relationships to salient features of the environment" (World Health, 1998). The WHOQOL-100 is a QoL assessment that encompasses these six domains subdivided into 24 facets (World Health, 1998). In the medical field, often only those aspects of QoL that are related to health are measured (commonly called health-related quality of life; HRQL). The Neurology-QoL questionnaire, that measures HRQL in different neurological disorders, was developed to increase consistency in clinical neurology practice and research (Cella et al., 2011). Another key goal of medical treatment is WB, also termed "subjective WB", which denotes how people experience and evaluate their lives (Stone and Mackie, 2013). The WHO-5 WB Index is a questionnaire that measures WB in primary health care patients, consisting of five simple and non-invasive questions, used across a range of study fields (Europe, 1998; Topp et al., 2015). A recent study by King, 2019 elaborated on the relationship between WB and different brain regions, a helpful clue in targeting neurological conditions. In the human literature, the terms QoL and WB are often used interchangeably, although there are also some differences worth mentioning. WB questionnaires usually produce a unique overall score (e.g., The World Health Organization (WHO)-5 Well-Being Index), while QOL questionnaires frequently result in separate scores for aspects of symptoms and functioning (e.g., World Health Organization's Quality of Life Questionnaire (WHOQOL)) (Costa et al., 2021).

However, despite the importance of QoL and WB measurements in patients, there is no unanimity regarding their assessment criteria in a veterinary scientific context (Christiansen and Forkman, 2007; McMillan, 2000; Yeates and Main, 2009). How are we assessing QoL and WB in animals? The situation is further complicated by the fact that QoL and WB are rarely defined in animal research literature and often used as equivalents in conjunction with other terms such as animal health or welfare (Christiansen and Forkman, 2007; Campos-Luna et al., 2019; Mullan, 2015; Yeates, 2018). We noted that animal QoL is described in the literature, among other definitions, as a "*multidimensional, experiential continuum comprising an array of affective states broadly classifiable as comfort-discomfort and pleasure states*" (McMillan, 2000) or the "*balance between the positive and negative experiences that animals have*" (Mellor, 2016). WB has been characterized as a "*state of clinical health or the absence of disease or injury*" in which the animals perform natural behaviours by eliminating "*things as pain, suffering, fear, anxiety and frustration*" (Hetts, 1991) or a state that can be *"properly maintained not only by eliminating pain, distress and behavioural abnormalities, but also by allowing an animal to perform species-specific behaviours*" (Yeates, 2018). Consequently, considering these experiences, challenges remain in developing widely applicable instruments/tests to assess QoL and WB in animals.

The house mouse (*Mus musculus*) is the most broadly used vertebrate species in experimental research laboratories, education and testing (Carbone, 2021). Consequently, 3.9 million mice were reported to have been employed for experimental purposes in the European Union and Norway in 2020 alone ("Summary Report on the statistics on the use of animals for scientific purposes in the EU and Norway (2020). https://circabc.europa.eu/ui/group/8ee3c69a-bccb-4f22-89ca-277e35de7c63/library/10ad28d6-e17e-4367-b459-20883402cfcc/details"). Specifically, genetically modified mouse models are of great scientific interest to understand the mechanisms of disease and to develop personalized treatments. Mouse phenotyping clinics worldwide employ a broad-range of phenotyping screening tests; researchers characterize biological and physiological traits, behaviors and disease symptoms with observable, measurable and analytical evaluations using standardized instruments and protocols, well described in many disease research areas (Hölter et al., 2015a; Hölter et al., 2015b). However, overall QoL and WB assessment criteria are rarely and not clearly reported in mouse studies (McMillan, 2000). Good health is known to decrease inter-mouse variability and increase replicability in preclinical stage trials across laboratories (Friese, 2013; Poole, 1997; Prescott and Lidster, 2017). Furthermore, there are strict governmental guidelines to adhere to the 3 R Principle (Replacement, Reduction, Refinement) and to protect the use of mice for scientific research purposes (European directives 2010/63). Attention is paid to ascertain that these legal and ethical obligations are followed, avoiding increased unnecessary risk of stress, pain or suffering during housing and phenotyping testing in research facilities (Zintzsch et al., 2017). A good example is the implementation of the Grimace scale to assess pain in the laboratory mouse (Langford et al., 2010). As such, we wanted to know which specific tests are currently being conducted in mouse research, in contrast to more general QoL and WB observations mainly deducted from cage behaviors. In addition, it remains unclear whether the tests, applied in the context of QoL and WB *in vivo* studies, are translatable to the self-reported outcomes used in later phases of human clinical trials. The assessment of QoL and WB in mouse and human studies may probably have common threads but, due to the ability to self-report by humans, there are still obvious differences and limitations (McMillan, 2000).

The aim of this study was to conduct a scoping review to identify the most commonly used QoL and WB instruments/tests and outcomes being measured in recent mouse scientific research. We aimed to promote awareness in the mouse and human research communities for those instruments/tests that are easy and inexpensive to implement, supplementing specific intervention-related instruments and measuring relevant disease outcomes for the benefit of the individual mouse but also to evaluate outcomes that are comparable to those used in later phases of clinical trials. In fact, a major goal of this study was to identify the methodology of QoL and WB assessment in mice (what and how is being measured) and to evaluate the extent of alignment with clinical studies.

2. Material and methods

2.1. Search strategy

Our study protocol was registered in the PROSPERO (International prospective register of systematic reviews) database: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018103507. Publications were identified according to the search strategy depicted in Table 1 and provided the initial dataset for this scoping review. Initially, publications measuring fatigue, an important outcome in human clinical trials, were also sought and will be addressed in a separate review.

Screening (abstracts and full texts) and data extraction were performed by two independent reviewers (AS-M and PS-B). Lack of consensus was resolved by discussion. A third reviewer with extensive experience in systematic reviews of measurement instruments (CT) provided methodological support throughout the whole process. The search was conducted in the Web of Science and MEDLINE (via PubMed) databases with the latest update on April 4th, 2022. Articles retrieved were transferred to EndNote. The publication library was imported to Covidence software (www.covidence.org), used for review management, that removed duplicates. Due to the high number of publications obtained for abstract screening, we limited our review to articles dated between August 2008 and August 2021.

2.2. Inclusion and exclusion criteria

All publications that used instruments/tests to measure QoL or WB in experimental mouse studies were included (see Table 1). Any publications that used other animals than mice (*Mus musculus*), that were not a

Table 1
Description of the search strategy used in the scoping review.

#1 search terms for mice: Mice [tiab] OR mouse [tiab].
#2 search terms for fatigue, quality of life, and wellbeing: HRQL [tiab] OR HRQoL [tiab] OR QL [tiab] OR QoL [tiab] OR "quality of life" [tiab] OR "life quality" [tiab] OR fatigue [tiab] OR wellbeing [tiab] OR "well being" [tiab] OR well-being [tiab].
#3 Combine #1 AND #2

primary research study (e.g., review, conference abstracts) or written in a language other than English were excluded.

2.3. Extraction of relevant data

Extracted data for this scoping review included the following: publication year, country, research area of the study, number of mice used, mouse strain, sex, description of the instruments/tests and outcomes measured. Outcomes refer to what is being measured (the concept, in this case QoL or WB), while instrument/test refers to how the outcome is being measured, e.g., the tests/observations that are used to produce a value or score. For instruments/tests, we quantified the number of publications in which they were employed. For outcomes, we grouped our extracted data in outcome categories, as often several outcomes (e.g., "nest location", "nest appearance", "time to integrate to the nest") related to a certain outcome category (e.g., "nest building") were mentioned in the same publication. Each of these outcomes measured ("findings") was quantified and the total number was assigned to the outcome category.

3. Results

3.1. Scoping review extraction features

Following removal of duplicates, a total of 7083 articles were identified (a flow diagram is depicted in Fig. 1). After abstract screening, the final number of full-texts reviewed was 1141 from which 576 fulfilled the inclusion criteria: 13 were related to QoL, 77 to WB and 486 to fatigue assessment (as mentioned in the *Methods* section, fatigue is the focus of another review). The publication years and the countries of the articles included in our review for both QoL and WB are shown in Fig. 2A and B. For QoL, there was an apparent increase in the number of articles in 2019-20 and publications were found from USA, China, Republic of Korea, Australia, Estonia, Italy and Spain (Fig. 2B). Equally, a higher number of WB studies were published in 2020–21 (considering that articles were only screened until August 2021). USA, Germany and Switzerland and UK appeared at the forefront of WB assessment (Fig. 2B). Studies measuring QoL in our search were predominantly from the Oncology field (Fig. 2C). Articles assessing WB were mainly found in the Laboratory Science field but the second study area was Neurology, followed by Oncology and Post-surgical stress (Fig. 2C). Regarding the number of animals used, there was variability in QoL studies while in WB most of the publications used less than 100 animals (Fig. 3A). The most common mouse strain employed was C57BL/6 and substrains, representing 31% and 60% of all strains in QoL and WB publications, respectively (Fig. 3B). While in articles measuring QoL parameters, mice were mainly male only (62%) or from both sexes (15%) or female only (15%), in WB studies proportions were more balanced: both sexes (45%), female only (22%) and male only (30%) (Fig. 3C). The sex of the mice was not stated in 8% and 3% of QoL and WB articles, respectively (Fig. 3C). The number of instruments/tests used in each publication for assessing QoL or WB is represented in Fig. 4A. In articles assessing QoL, the maximum number of instruments/tests employed was 5 in a given manuscript but in WB this number went up to 11 (Fig. 4A). In addition, the use of one instrument/test was frequently reported in the publications screened for both QoL and WB (Fig. 4A). The number of outcomes measured was lower in QoL publications than in WB ones, where a maximum of 27 outcomes were reported in one of the articles (Fig. 4B).

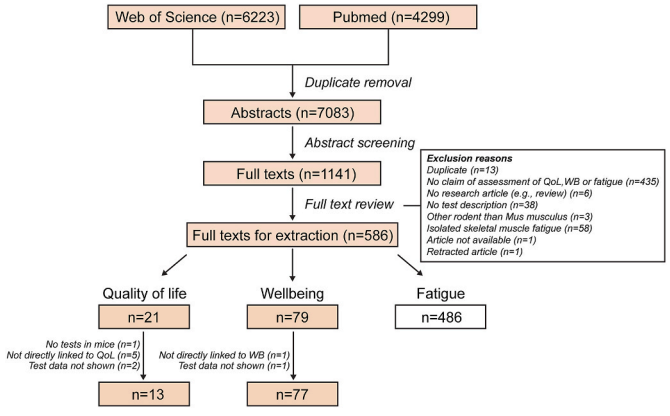

Fig. 1. Scoping review flow diagram showing the number of eligible articles used in abstract screening, full text review and data extraction. The reasons for exclusion of the reviewed articles are also depicted.

3.2. Instruments/tests and outcomes measured in articles assessing QoL

The number of articles which specified the assessment of QoL parameters in mice was low (only 13 publications). From these, the instruments/tests more frequently used were body weight scale, and food intake measurement, appearing in 4 and 3 articles each, respectively. Following, body condition score, cage activity behavior and tumor size appeared in 2 articles each (Table 2). A more detailed description of the instruments/tests used in all publications (including scoring details) is shown in Supplementary Table A1. Regarding the outcomes measured for determining mouse QoL in the publications screened, the most common outcome category was related to home cage activity parameters. Up to 7 findings from the articles reviewed were included in this category (e.g., vertical and stereotypic counts, active or walking time, etc.) (Table 3). Other less frequent outcomes measured were body weight, food consumption and tumor size with 5, 4 and 3 findings in the same outcome category, respectively (Table 3). Further categorization revealed that most QoL outcomes belonged to behavioral (38%), physiological (29%) and pathological (20%) readouts (Supplementary Table A3).

3.3. Instruments/tests and outcomes measured in articles assessing WB

There was a great heterogeneity in instruments/tests used and outcomes measured in WB, probably due to the larger number of articles from which data were extracted. Interestingly, we found that the most common instrument/test used to assess WB in mice from the publications screened was the body weight scale (Table 4). Up to 32 articles out of the total 77 that fulfilled our selection criteria (42%) measured body weight. Other common instruments/tests used were nest building behavior, burrowing behavior and corticosterone metabolites (found in 22, 13 and 10 publications out of the 77, respectively) (Table 4). Finally, with 9 publications each, cage activity monitoring (with video or infrared sensors) and food intake were also recurrent. A brief description of all instruments/tests retrieved and scoring details is available in Supplementary Table A2.

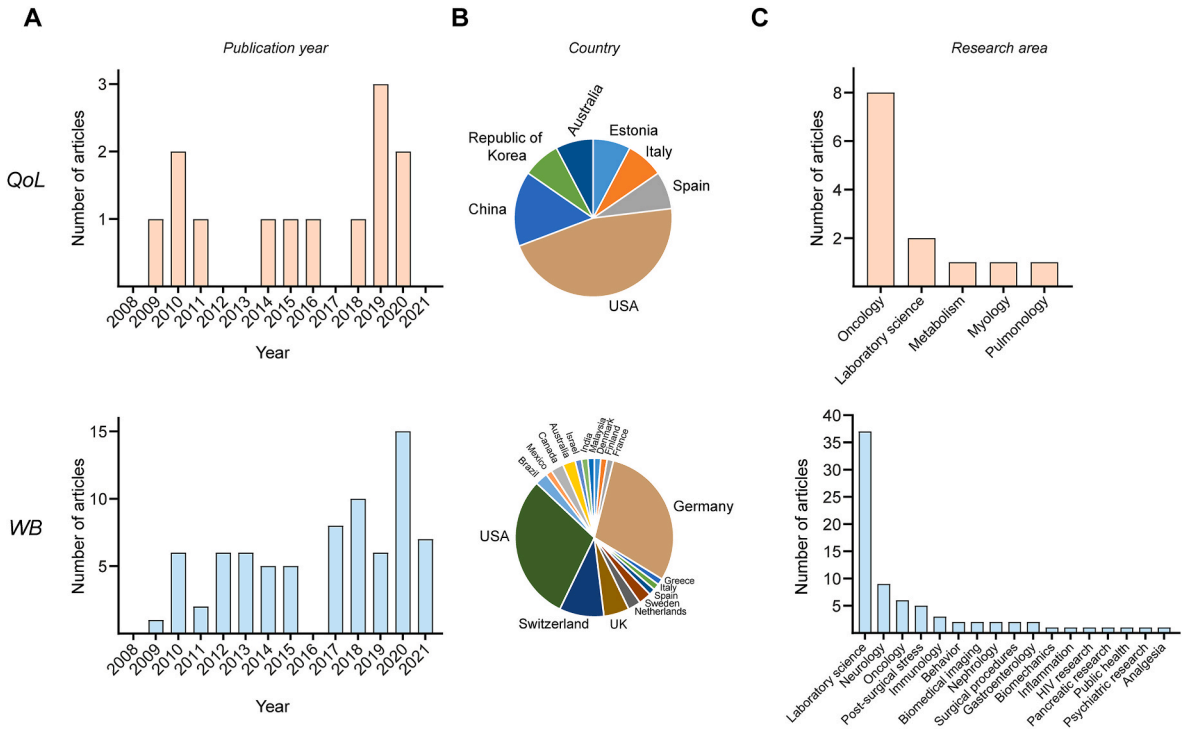

Fig. 2. Number of included articles by publication year, country and research area in quality of life (QoL) and wellbeing (WB) assessments.

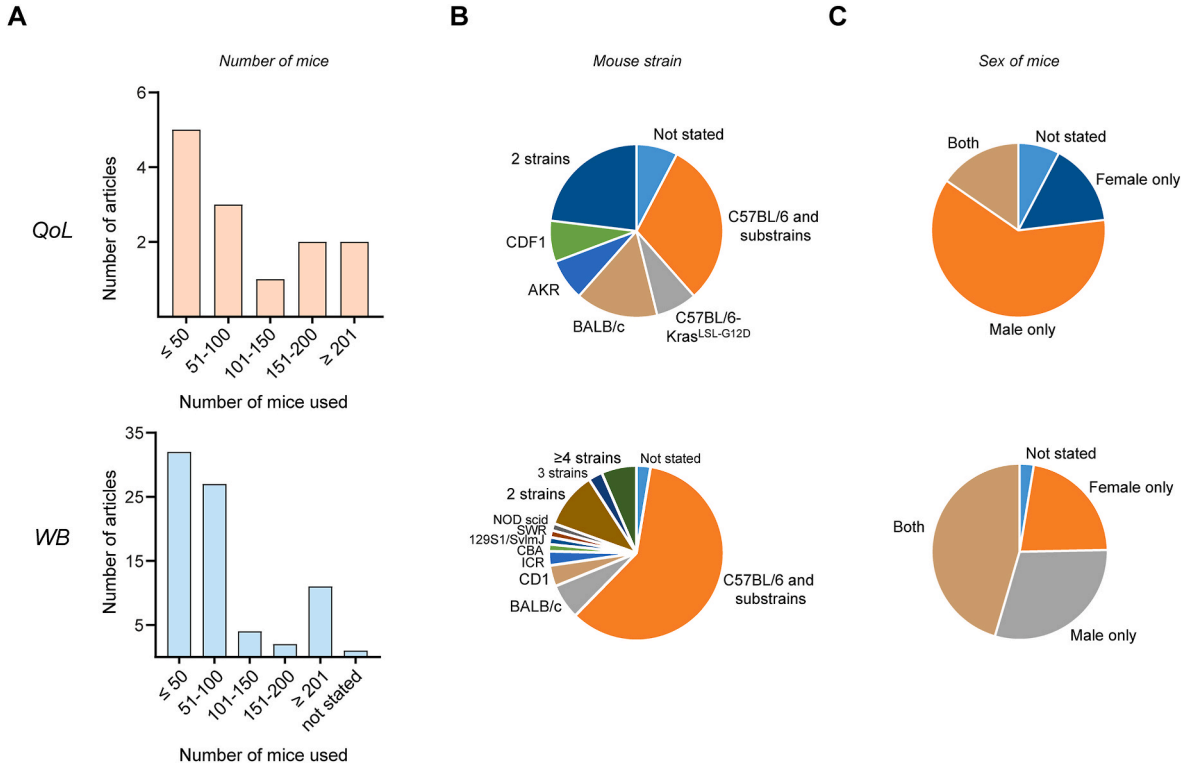

Fig. 3. Number of articles by number of mice used and frequency of strain and sex included in quality of life (QoL) and wellbeing (WB) assessments.

With respect to the outcomes measured in publications determining mouse WB, nest building was the most frequent with 39 findings in this outcome category (Table 5). Coming close in second place was body weight, also very common with 36 findings. Other less usual outcome categories linked to WB were home cage activity, food consumption, and burrowing behavior (Table 5). As indicated in Supplementary Table A3, WB outcomes from the behavioral field accounted for 53% of the outcome readouts of Table 5, followed by physiological (29%) and clinical chemistry/biochemistry (9%) outcomes.

Regarding why QoL and WB are measured in mouse research, we can categorize two major groups of motivations focusing either on the state of the mouse alone (e.g., describing how nest building behavior is

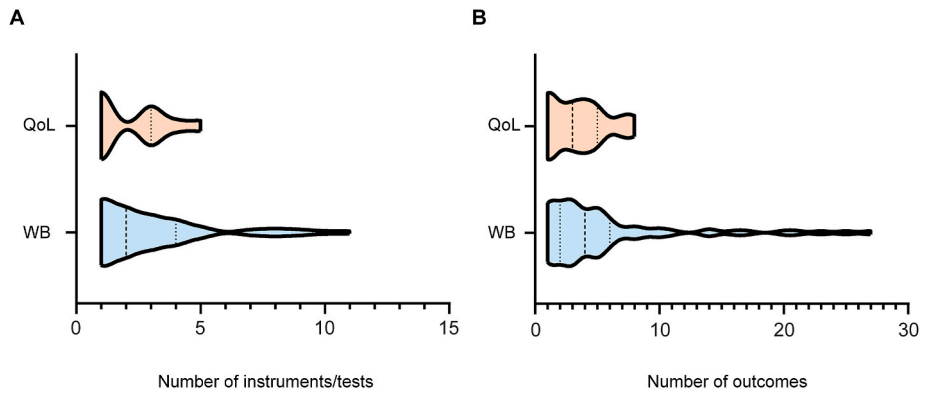

Fig. 4. Violin plots of the number of instruments/tests (A) and outcomes (B) in the articles identified in the review. In the X-axis the number of instruments/tests and outcomes are depicted. The Y-axis shows the frequency distribution of the number of publications (wider sections representing more articles that reported the particular number of instruments/tests or outcomes labelled in the X-axis).

Table 2
Summary of instruments/tests used in quality of life (QoL) assessment in mice.

Instrument/test	Number of articles the instrument/test was used
Body weight scale (no details and digital)	4
Food intake	3
Body condition score	2
Cage activity behavior with infrared sensor	2
Tumor diameter (no details and digital caliper)	2
Discomfort symptoms (mouse observation)	1
Elevated plus maze test	1
Forced swim test	1
Fur condition	1
Glucometer	1
Intruder-resident paradigm	1
Life expectancy	1
Observation (bleeding from penis)	1
Quality of life index (based on clinical signs)	1
Quality of life scores (based on clinical signs)	1
Rotarod test	1
Thermometer – Body temperature	1
Treadmill (motorized)	1
Tumor inhibition rates (formula)	1
Tumor weight	1
Wheel running activity	1

Table 3
Summary of the outcome categories of quality of life (QoL) assessment in mice.

Outcome categories	Outcome number per category
Home cage activity parameters (vertical & stereotypic counts, active or walking time)	7
Body weight	5
Food consumption	4
Tumor size	3
Body conditioning score (BCS)	2
Entries in the elevated plus-maze	2
Bleeding from the penis	1
Blood glucose	1
Life expectancy	1
Quality of life (QOL) scores	1
Rotarod performance (latency to fall)	1
Body temperature	1
Piloerection	1
Diarrhea or constipation	1
Posture	1
Tremors	1
Closed eyes	1
Red tears (chromodacryorrhea)	1
Level of interest of the resident-intruder	1
Forced swimming test time	1
Running distance (in running wheel)	1
Tumor inhibition rates	1
Tumor weight	1
Fur condition	1
Water consumption	1
Average speed (in emergence test)	1
Total distance (in emergence test)	1
Resting time (in emergence test)	1

related to mouse WB) or having the more translational objective of measuring outcomes to improve disease treatment in humans (e.g., outcome measurement of drug testing in mice before clinical trials with patients). Interestingly, we found that only 32% (25 out of 77) of our WB selected articles, in contrast to 85% (11 out of 13) of our QoL included publications, aimed to have a more translational focus. Thus, QoL assessment in mice seems more frequently linked to translational research than WB.

4. Discussion

To the best of our knowledge, this is the first extensive and rigorous literature search that aimed to identify and describe existing instruments/tests in publications in which the authors claimed to assess QoL and WB in experimental mice. This was an objective of the COordinated Research infrastructures Building Enduring Life-sciences services (CORBEL) interdisciplinary initiative towards harmonization of instruments and outcomes between mouse and human to infer better clinical trial translation ('narrowing the gap') from which other preclinical (Harman et al., 2020) and clinical (Harman et al., 2017; Harman et al., 2019) studies were published. For both QoL and WB, we found a high variability of instruments/tests used and outcomes measured in the reviewed mouse studies encompassing several clinical areas, such as neurology and oncology. In addition, among similar parameters tested, we observed variable methodological procedures, sample sizes and mouse sex used for the experiments. These issues clearly indicate that there is a lack of consensus on describing as well as with measuring QoL and WB in the mouse research field.

One of the most unexpected findings of this scoping review was the low number of publications assessing QoL in mouse studies. This illustrates the challenge of adopting human QoL assessments, such as the Neuro-QoL tool, which have a pivotal importance and are required in clinical studies (Cella et al., 2011; Haraldstad et al., 2019). Instruments/tests of QoL assessment in mice should be predictive of response to therapeutic treatments in research or preclinical testing. However, the

Table 4
Summary of the top 25 instruments/tests used for wellbeing (WB) assessment in mice.

Instrument/test	Number of articles the instrument/test was used
Body weight scale (not defined and technical device defined)	32
Nest building behavior/nest complexity scoring	22
Burrowing behavior/performance test	13
Corticosterone metabolites (fecal, hair)	10
Cage activity behavior with video or infrared sensor	9
Food intake/feed disappearance	9
Mouse Grimace Scale score from photos	7
Blood chemistry analyses	6
Cage activity behavior without video I	6
Hemogram/blood count	6
Necropsy - organ weights	5
Light-dark box (Maze)	4
Open field test with automatically measured behavioral parameters	4
Time to integrate into nest test/explorative test score	4
DEXA dual-energy X-ray absorptiometry	3
Free exploratory paradigm	3
Fur scoring/coat state	3
Observation of survival rates	3
Rotarod test	3
Social interaction test	3
Vocalization or floating behavior linked to handing while performing Morris water maze test	3
Water consumption	3
Wheel running - voluntary activity recorded using a automatic counter	3
Body condition score	2
Body temperature (transponder or non-contact infrared thermometry)	2

Table 5
Summary of the top 25 outcome categories of wellbeing (WB) assessment in mice.

Outcome categories (Wellbeing)	Outcome number per category
Nest building	39
Body weight	36
Home cage activity	16
Food consumption	15
Burrowing behavior	14
Fur/skin appearance	12
Corticosterone metabolites (fecal, hair, serum)	12
Body posture	11
Aggressive behavior/barbering	10
Water consumption	10
Grooming behavior	10
Mouse grimace scale (MGS) score	9
Time mobile/immobile	8
Hemoglobin content	7
General behavioral observations	7
Defecation or stool consistency	7
Vocalization	7
Gait analysis	6
Light-dark maze performance	6
Organ weights (thymus, adrenal, heart, testes, kidney)	6
Eyes (e.g. discharge, color, closure, orbital tightness)	6
Body temperature	6
Resting bouts	6
White Blood Cells (WBCs) counts	6
Social behaviors	6

concept of QoL from a veterinary viewpoint is poorly understood and often equated to lack of disease in general. Further, QoL assessment is not a priority in preclinical studies because laboratories focus principally on disease-specific measurements (e.g., showing tumor growth after drug administration but neglecting the possible side-effects of the drug on the mice by QoL measures). Although we are aware of discrepancies in both the biology (Howe et al., 2018; Leenaars et al., 2019) and the possibility to self-report of mice and humans, with this review we intend to raise awareness for the unfavorable "translational methodological gap" between mouse experiments and clinical research. Also, to stress the necessity to bring both fields closer by developing standard protocols for QoL assessment in mice.

The top 3 instruments/tests used to assess QoL - body weight scale, food intake quantification and body condition score - are inexpensive and rapid to perform in a laboratory, require no additional equipment, training or costly software and are therefore highly practical and easily integrated in a mouse phenotyping facility (Ahloy-Dallaire et al., 2019). However, the associated outcomes are not used to assess human QoL in clinical research because they do not address how a patient feels and functions in daily life. The challenge is to find instruments/tests that can be easily implemented but that can primarily encompass measures that matter in human QoL such as feelings, functions and natural behaviors, beyond disease itself (the core domains of QoL are feelings of pain, anxiety, depression, fatigue, and physical, mental and social function), resulting in a better clinical translation. With this objective in mind, one can argue that body weight, food intake and body condition score are not appropriate measures of QoL in mice and therefore specific knowledge of human QoL by animal scientists could be beneficial to better define which QoL instruments/tests should be used in preclinical research.

Similarly, the body weight scale was the top instrument/test for WB assessment in the mouse research publications screened. The body weight outcome is commonly used to determine humane endpoints in mice (with a weight loss of 20% or more) but it has been shown that it is highly dependent on the context and nature of the experiments (Talbot et al., 2020). As stated in this study, a loss of less than 20% of weight can indicate "*clinical deterioration*" in mice depending on the experimental setting (e.g., intracranial glioma model), so other parameters should be additionally measured for determination of humane endpoints (Talbot et al., 2020). Other WB instruments/tests from Table 4 seem more informative about how mice feel and function. For example, nest building behavior has been linked to mouse wellbeing in a number of publications (Gaskill et al., 2013; Jirkof, 2014) and it is an easy to implement and affordable test (Deacon, 2006). In our literature search, nest building behavior was found as a sensitive WB functional measure of clinical stroke deficit (Yuan et al., 2018). Also, burrowing behavior and automated measurements of home cage activity can be used to detect behavioral abnormalities that might be missed by simple visual inspection from the investigator (Voikar and Gaburro, 2020) and can be critical for determining compromised WB, such as disease onset (Richardson, 2015) or aggression (Theil et al., 2020). Interestingly, the Mouse Grimace Scale (MGS), able to determine the occurrence or severity of pain in mice by inspection of facial features (Whittaker et al., 2021), was used in only 7 out of 77 publications in the context of WB. Its utilization follows the 3 R principle for humane research, in which refinement prioritizes those methods that minimize the suffering or distress of the mice (Jirkof et al., 2019; Tannenbaum and Bennett, 2015). Therefore, the use of standardized protocols for QoL or WB assessment can help the investigator identify unwanted physical and mental health suffering in any given experimental setting.

Although there was a considerable overlap in the instruments/tests used to examine QoL and WB in the reviewed literature, many authors agree on WB outcomes such as nest building behavior, burrowing behavior, corticosterone metabolites, grooming behavior and Mouse Grimace Scale that were not included in the QoL outcomes. These results agree with the general view that WB in animals is associated with a "broad behavioural repertoire" (Baumans, 2005) and its assessment requires a "thorough knowledge of the specific behaviour and biological needs of that particular species".

Another important observation is that for many of the instruments/

tests screened, we found a lack of methodological information: how often the instrument/test was used or cut-off values for outcomes measured (e.g., body weight was used but no mention to level of body weight loss or gain that compromised QoL or WB). The ARRIVE guidelines (Animal Research: Reporting of *In Vivo* Experiments), a checklist of information that should be reported in publications including *in vivo* experiments, were inconsistently followed in the reviewed literature (Percie du Sert et al., 2020). Also, for identical outcomes observed in our review, we found a high variability in the methodology applied, demonstrated for instance with the use of manual or automated observation when studying cage behavior. It would be valuable to dissect these QoL and WB measures and study their validity and reliability.

A striking observation in this study was that, despite QoL and WB functioning as vital concepts in the welfare and human care of animals in veterinary medicine (McMillan, 2000), they were rarely defined and often used interchangeably in the mouse field, as reported in human medicine (Sears et al., 2014; Costa et al., 2021). But, how do we develop common measures in mouse models that better translate into patients? In humans, the COMET initiative (Williamson et al., 2017) aims to facilitate the development of core outcome sets, which are sets of outcomes that should be measured and reported in all clinical trials in a specific condition. The literature is replete with core outcome set (COS) developmental processes to establish 'what' outcomes are important for a particular disease and 'how' these should be measured (Gargon et al., 2021). With this in view, a qualitative approach to evaluate commonalities between mouse studies and clinical trials is needed to facilitate the assessment of QoL and WB in mice by establishing a defined set of outcomes and measuring instruments/tests. Obviously, the aim to establish appropriate QoL and WB measures is complex, subjective and likely to have a risk of uncritical anthropomorphism, but we find it would narrow the translation gap and improve laboratory animal care (Morton et al., 1990; Sundberg and Schofield, 2018). In this respect, the structured process of the Delphi method is a useful tool as demonstrated in the work of Campos-Luna (Campos-Luna et al., 2019). With a modified Delphi consultation survey, the authors reported that hunched posture, coat condition and body condition score were the top 3 methods for evaluating mouse welfare in a laboratory mouse facility. The term welfare, also lacking consensus in an accepted general definition (Reimert et al., 2023), encompasses physical and mental WB, but also the concept of "natural living", in which animals can perform species-specific behaviors (Bracke and Hopster, 2006). Broom defined the welfare of an individual as "its state as regards its attempts to cope with its environment" (Broom, 1986). For the Model Animal Welfare Curriculum Planning Group (AVMA), good animal welfare requires disease prevention, responsible care, proper housing, management, nutrition, careful handling and humane euthanasia (Lord et al., 2017). While this report from Campos-Luna related to welfare, not the focus of the present scoping review as we concentrated in PROMs, in the future the same methodological approach, informed by our results, could be used for establishing consensus on the concepts of QoL and WB and how to measure them in mice. With this in mind, we identified comparable endeavors in recently published articles that created QoL or WB composite scores/indexes based on clinical observations (body weight, water/food intake, cage activity, etc.). We think this effort is positive and should be acknowledged and further developed in agreement with human measures (Bohnert et al., 2019; Roda et al., 2020).

One should highlight that this current review has some limitations. The main one is the restriction to a 13-year period. Also, we excluded abstracts in other languages than English or published as conference abstracts. We believe this might have excluded some mouse research facilities worldwide that assessed QoL and WB. In addition, articles that did not mention the terms WB and QoL in the abstract were not considered. Finally, it is important to recognize that a small number of WB publications derived from the same research group throughout the years. This resulted in an increase in the number of articles using the same instrument/test (see Supplementary Tables A1 and A2).

Finally, this scoping review is also relevant to understand how much of the mouse research studied the impact of disease or treatment for potential human applicability and how much was focused on the mouse *per se* with less of a translational focus. Importantly, we found that the majority of QoL publications (85%) included in the full-text data extraction stage had a translational aim. In contrast, more than two thirds of the publications claiming to assess WB used instruments/tests to determine mouse WB *per se*. This suggests that QoL is more frequently considered to be a measure of treatment effect in a translational context in mouse studies but, as already mentioned, the number of mouse research articles assessing QoL is very low. In this context, Hooijmans and colleagues (Hooijmans et al., 2014) strongly suggest the implementation of systematic reviews of animal experiments prior to initiating studies with patients to provide all available information to clinical research. This approach was further emphasized in a recent article despite the associated temporal and logistical limitations (Pound and Ritskes-Hoitinga, 2020). We have also observed an increased interest in QoL and WB measures from the scientific community as reflected by the boost in QoL and WB mouse publications in recent years. In our view, it is imperative to continue to improve the alignment of QoL and WB instruments/tests used in preclinical and clinical studies in order to guarantee better translation from mouse to human.

5. Conclusions

In a nutshell, there is a lack of consensus on how to measure QoL and WB in the mouse research field. For ensuring a better translation from mouse to human, we emphasize the need for measuring outcomes that matter to human health, like QoL and WB, in mice. Also, we believe that an effort for standardization of QoL and WB assessment protocols (tests and outcomes) should be pursued with high priority in the mouse research community.

Authors' contributions

PDSB designed the study, ASM performed the search strategy, ASM and PDSB collected, extracted and analyzed the data and wrote the manuscript, CBT provided valuable guidance in analysis, collection and writing. All authors read and approved the final manuscript.

Funding

This work was financially supported by the European Union's Horizon 2020 research and innovation program CORBEL (COordinated Research Infrastructures Building Enduring Life-science services) under grant agreement no. 654248.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the remaining members of the SCORE-IT study team (Serena Battaglia, Jacques Demotes-Mainard, Silvio Garattini, Nicola L Harman, Cecilia A C Prinsen and Paula R Williamson), and thank the INFRAFRONTIER GmbH. The authors are grateful to the Helmholtz Munich Zentralbibliotek librarians, who searched for article PDFs.

Appendix ASupplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nsa.2024.104058.

References

Ahloy-Dallaire, J., Klein, J.D., Davis, J.K., Garner, J.P., 2019. Automated monitoring of mouse feeding and body weight for continuous health assessment. Lab Anim 53 (4), 342–351. https://doi.org/10.1177/0023677218797974.

Baumans, V., 2005. Science-based assessment of animal welfare: laboratory animals. Rev Sci Tech 24 (2), 503–513.

Bohnert, B.N., Dörffel, T., Daiminger, S., Calaminus, C., Aidone, S., Falkenau, A., Artunc, F., 2019. Retrobulbar Sinus Injection of Doxorubicin is more Efficient than lateral Tail Vein Injection at Inducing experimental Nephrotic Syndrome in mice: a Pilot study. Lab Anim 53 (6), 564–576. https://doi.org/10.1177/0023677218824382.

Bracke, M.B.M., Hopster, H., 2006. Assessing the importance of natural behavior for animal welfare. J. Agric. Environ. Ethics 19 (1), 77–89. https://doi.org/10.1007/s10806-005-4493-7.

Broom, D.M., 1986. Indicators of poor welfare. Br. Vet. J. 142 (6), 524–526. https://doi.org/10.1016/0007-1935(86)90109-0.

Campos-Luna, I., Miller, A., Beard, A., Leach, M., 2019. Validation of mouse welfare indicators: a Delphi consultation survey. Sci. Rep. 9 (1), 10249 https://doi.org/10.1038/s41598-019-45810-y.

Carbone, L., 2021. Estimating mouse and rat use in American laboratories by extrapolation from Animal Welfare Act-regulated species. Sci. Rep. 11 (1), 493. https://doi.org/10.1038/s41598-020-79961-0.

Cella, D., Nowinski, C., Peterman, A., Victorson, D., Miller, D., Lai, J.S., Moy, C., 2011. The neurology quality-of-life measurement initiative. Arch. Phys. Med. Rehabil. 92 (10 Suppl. l), S28–S36. https://doi.org/10.1016/j.apmr.2011.01.025.

Christiansen, S.B., Forkman, B., 2007. Assessment of animal welfare in a veterinary context—a call for ethologists. Appl. Anim. Behav. Sci. 106 (4), 203–220. https://doi.org/10.1016/j.applanim.2007.01.004.

Costa, D.S.J., Mercieca-Bebber, R., Rutherford, C., Tait, M.A., King, M.T., 2021. How is quality of life defined and assessed in published research? Qual. Life Res. 30 (8), 2109–2121. https://doi.org/10.1007/s11136-021-02826-0.

Deacon, R.M., 2006. Assessing nest building in mice. Nat. Protoc. 1 (3), 1117–1119. https://doi.org/10.1038/nprot.2006.170.

W.H. O.R. O.f., 1998. Wellbeing Measures in Primary Health Care: the DepCare Project : Report on a WHO meeting Stockholm, Sweden 12-13 February 1998. WHO Regional Office for Europe.

Friese, C., 2013. Realizing potential in translational medicine: the Uncanny Emergence of care as science. Curr. Anthropol. 54 (S7), S129–S138. https://doi.org/10.1086/670805.

Gargon, E., Gorst, S.L., Matvienko-Sikar, K., Williamson, P.R., 2021. Choosing important health outcomes for comparative effectiveness research: 6th annual update to a systematic review of core outcome sets for research. PLoS One 16 (1), e0244878. https://doi.org/10.1371/journal.pone.0244878.

Gaskill, B.N., Karas, A.Z., Garner, J.P., Pritchett-Corning, K.R., 2013. Nest building as an indicator of health and welfare in laboratory mice. J. Vis. Exp. 82, 51012 https://doi.org/10.3791/51012.

GBD 2016 Neurology Collaborators, 2019. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 18 (5), 459–480. https://doi.org/10.1016/s1474-4422(18)30499-x.

Haraldstad, K., Wahl, A., Andenæs, R., Andersen, J.R., Andersen, M.H., Beisland, E., Helseth, S., 2019. A systematic review of quality of life research in medicine and health sciences. Qual. Life Res. 28 (10), 2641–2650. https://doi.org/10.1007/s11136-019-02214-9.

Harman, N.L., James, R., Wilding, J., Williamson, P.R., 2017. SCORE-IT (Selecting Core Outcomes for Randomised Effectiveness trials in Type 2 diabetes): a systematic review of registered trials. Trials 18 (1), 597. https://doi.org/10.1186/s13063-017-2317-5.

Harman, N.L., Sanz-Moreno, A., Papoutsopoulou, S., Lloyd, K.A., Ameen-Ali, K.E., Macleod, M., Williamson, P.R., 2020. Can harmonisation of outcomes bridge the translation gap for pre-clinical research? A systematic review of outcomes measured in mouse models of type 2 diabetes. J. Transl. Med. 18 (1), 468. https://doi.org/10.1186/s12967-020-02649-6.

Harman, N.L., Wilding, J.P.H., Curry, D., Harris, J., Logue, J., Pemberton, R.J., Williamson, P.R., 2019. Selecting Core Outcomes for Randomised Effectiveness trials in Type 2 diabetes (SCORE-IT): a patient and healthcare professional consensus on a core outcome set for type 2 diabetes. BMJ Open Diabetes Res Care 7 (1), e000700. https://doi.org/10.1136/bmjdrc-2019-000700.

Hetts, S., 1991. Psychologic well-being: Conceptual issues, behavioral measures, and Implications for Dogs. Vet. Clin. Small Anim. Pract. 21 (2), 369–387. https://doi.org/10.1016/S0195-5616(91)50039-5.

Hölter, S.M., Einicke, J., Sperling, B., Zimprich, A., Garrett, L., Fuchs, H., Wurst, W., 2015a. Tests for anxiety-related behavior in mice. Curr. Protoc. Mol. Biol. 5 (4), 291–309. https://doi.org/10.1002/9780470942390.mo150010.

Hölter, S.M., Garrett, L., Einicke, J., Sperling, B., Dirscherl, P., Zimprich, A., Wurst, W., 2015b. Assessing Cognition in mice. Curr. Protoc. Mol. Biol. 5 (4), 331–358. https://doi.org/10.1002/9780470942390.mo150068.

Hooijmans, C.R., IntHout, J., Ritskes-Hoitinga, M., Rovers, M.M., 2014. Meta-analyses of animal studies: an introduction of a valuable instrument to further improve healthcare. ILAR J. 55 (3), 418–426. https://doi.org/10.1093/ilar/ilu042.

Howe, J.R.t., Bear, M.F., Golshani, P., Klann, E., Lipton, S.A., Mucke, L., Silva, A.J., 2018. The mouse as a model for neuropsychiatric drug development. Curr. Biol. 28 (17), R909–r914. https://doi.org/10.1016/j.cub.2018.07.046.

Jirkof, P., 2014. Burrowing and nest building behavior as indicators of well-being in mice. J. Neurosci. Methods 234, 139–146. https://doi.org/10.1016/j.jneumeth.2014.02.001.

Jirkof, P., Rudeck, J., Lewejohann, L., 2019. Assessing affective state in laboratory Rodents to promote animal welfare-what is the progress in applied refinement research? Animals (Basel) 9 (12). https://doi.org/10.3390/ani9121026.

King, M.L., 2019. The neural correlates of well-being: a systematic review of the human neuroimaging and neuropsychological literature. Cognit. Affect Behav. Neurosci. 19 (4), 779–796. https://doi.org/10.3758/s13415-019-00720-4.

Langford, D.J., Bailey, A.L., Chanda, M.L., Clarke, S.E., Drummond, T.E., Echols, S., Mogil, J.S., 2010. Coding of facial expressions of pain in the laboratory mouse. Nat. Methods 7 (6), 447–449. https://doi.org/10.1038/nmeth.1455.

Leenaars, C.H.C., Kouwenaar, C., Stafleu, F.R., Bleich, A., Ritskes-Hoitinga, M., De Vries, R.B.M., Meijboom, F.L.B., 2019. Animal to human translation: a systematic scoping review of reported concordance rates. J. Transl. Med. 17 (1), 223. https://doi.org/10.1186/s12967-019-1976-2.

Lord, L.K., Millman, S.T., Carbone, L., Cook, N., Fisher, A., McKeegan, D., Patterson-Kane, E., 2017. A model curriculum for the study of animal welfare in colleges and schools of veterinary medicine. J. Am. Vet. Med. Assoc. 250 (6), 632–640. https://doi.org/10.2460/javma.250.6.632.

McMillan, F.D., 2000. Quality of life in animals. J. Am. Vet. Med. Assoc. 216 (12), 1904–1910. https://doi.org/10.2460/javma.2000.216.1904.

Mellor, D.J., 2016. Updating animal welfare thinking: Moving beyond the "five Freedoms" towards "A life worth living". Animals (Basel) 6 (3). https://doi.org/10.3390/ani6030021.

Morton, D.B., Berghardt, G.M., Smith, J.A., 1990. Animals, science, and ethics – Section III. Critical anthropomorphism, animal suffering, and the ecological context. Hastings Cent. Rep. 20 (3), S13–S19.

Mullan, S., 2015. Assessment of quality of life in veterinary practice: developing tools for companion animal carers and veterinarians. Vet Med (Auckl) 6, 203–210. https://doi.org/10.2147/vmrr.S62079.

Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M.T., Baker, M., Würbel, H., 2020. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. PLoS Biol. 18 (7), e3000410 https://doi.org/10.1371/journal.pbio.3000410.

Poole, T., 1997. Happy animals make good science. Lab Anim 31 (2), 116–124. https://doi.org/10.1258/002367797780600198.

Pound, P., Ritskes-Hoitinga, M., 2020. Can prospective systematic reviews of animal studies improve clinical translation? J. Transl. Med. 18 (1), 15. https://doi.org/10.1186/s12967-019-02205-x.

Pouwer, F., Snoek, F.J., van der Ploeg, H.M., Adèr, H.J., Heine, R.J., 2000. The well-being questionnaire: evidence for a three-factor structure with 12 items (W-BQ12). Psychol. Med. 30 (2), 455–462. https://doi.org/10.1017/s0033291700001719.

Prescott, M.J., Lidster, K., 2017. Improving quality of science through better animal welfare: the NC3Rs strategy. Lab Anim (NY) 46 (4), 152–156. https://doi.org/10.1038/laban.1217.

Reimert, I., Webb, L.E., van Marwijk, M.A., Bolhuis, J.E., 2023. Review: towards an integrated concept of animal welfare. Animal 17 (Suppl. 4), 100838. https://doi.org/10.1016/j.animal.2023.100838.

Richardson, C.A., 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.

Roda, E., Luca, F., Iorio, C.D., Ratto, D., Siciliani, S., Ferrari, B., Rossi, P., 2020. Novel medicinal Mushroom Blend as a Promising supplement in integrative Oncology: a Multi-Tiered study using 4T1 Triple-negative mouse Breast Cancer model. Int. J. Mol. Sci. 21 (10) https://doi.org/10.3390/ijms21103479.

Sears, L.E., Agrawal, S., Sidney, J.A., Castle, P.H., Rula, E.Y., Coberley, C.R., Harter, J.K., 2014. The well-being 5: development and validation of a diagnostic instrument to improve population well-being. Popul. Health Manag. 17 (6), 357–365. https://doi.org/10.1089/pop.2013.0119.

Stone, A.A., Mackie, C., 2013. Introduction. In *subjective well-being: measuring Happiness, suffering, and other Dimensions of experience*. Panel on Measuring Subjective Well-Being in a Policy-Relevant Framework. National Academies Press (US).

Summary Report on the Statistics on the Use of Animals for Scientific Purposes in the EU and Norway, 2020. https://circabc.europa.eu/ui/group/8ee3c69a-bccb-4f22-89ca-277e35de7c63/library/10ad28d6-e17e-4367-b459-20883402cfcc/details.

Sundberg, J.P., Schofield, P.N., 2018. Living inside the box: environmental effects on mouse models of human disease. Dis Model Mech 11 (10). https://doi.org/10.1242/dmm.035360.

Talbot, S.R., Biernot, S., Bleich, A., van Dijk, R.M., Ernst, L., Häger, C., Zechner, D., 2020. Defining body-weight reduction as a humane endpoint: a critical appraisal. Lab Anim 54 (1), 99–110. https://doi.org/10.1177/0023677219883319.

Tannenbaum, J., Bennett, B.T., 2015. Russell and Burch's 3Rs then and now: the need for clarity in definition and purpose. J Am Assoc Lab Anim Sci 54 (2), 120–132.

Theil, J.H., Ahloy-Dallaire, J., Weber, E.M., Gaskill, B.N., Pritchett-Corning, K.R., Felt, S. A., Garner, J.P., 2020. The epidemiology of fighting in group-housed laboratory mice. Sci. Rep. 10 (1), 16649 https://doi.org/10.1038/s41598-020-73620-0.

Topp, C.W., Østergaard, S.D., Søndergaard, S., Bech, P., 2015. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother. Psychosom. 84 (3), 167–176. https://doi.org/10.1159/000376585.

van Agteren, J., Iasiello, M., Lo, L., Bartholomaeus, J., Kopsaftis, Z., Carey, M., Kyrios, M., 2021. A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. Nat Hum Behav. https://doi.org/10.1038/s41562-021-01093-w.

Voikar, V., Gaburro, S., 2020. Three Pillars of automated home-cage phenotyping of mice: novel findings, refinement, and Reproducibility based on literature and

experience. Front. Behav. Neurosci. 14, 575434 https://doi.org/10.3389/fnbeh.2020.575434.
Whittaker, A.L., Liu, Y., Barker, T.H., 2021. Methods used and Application of the mouse Grimace scale in biomedical research 10 Years on: a scoping review. Animals (Basel) 11 (3). https://doi.org/10.3390/ani11030673.
Williamson, P.R., Altman, D.G., Bagley, H., Barnes, K.L., Blazeby, J.M., Brookes, S.T., Young, B., 2017. The COMET Handbook: version 1.0. Trials 18 (Suppl. 3), 280. https://doi.org/10.1186/s13063-017-1978-4.
World Health, O., 1998. Programme on mental health : WHOQOL user manual. Retrieved from Geneva. https://apps.who.int/iris/handle/10665/77932.
Yeates, J., 2018. Naturalness and animal welfare. Animals (Basel) 8 (4). https://doi.org/10.3390/ani8040053.
Yeates, J., Main, D., 2009. Assessment of companion animal quality of life in veterinary practice and research. J. Small Anim. Pract. 50 (6), 274–281. https://doi.org/10.1111/j.1748-5827.2009.00755.x.
Yuan, D., Liu, C.L., Wu, J., Hu, B.R., 2018. Nest-building activity as a reproducible and long-term stroke deficit test in a mouse model of stroke. Brain and Behavior 8 (6). https://doi.org/10.1002/brb3.993.
Ziegeler, B., W, D.S., Vinton, A., Mulukutla, S., Shaw, C., Carne, R., 2023. Neurological health: not Merely the absence of disease: current wellbeing instruments across the Spectrum of neurology. Am. J. Lifestyle Med. 17 (2), 299–316. https://doi.org/10.1177/15598276221086584.
Zintzsch, A., Noe, E., Reißmann, M., Ullmann, K., Krämer, S., Jerchow, B., Rülicke, T., 2017. Guidelines on severity assessment and classification of genetically altered mouse and rat lines. Lab Anim 51 (6), 573–582. https://doi.org/10.1177/0023677217718863.