**Title: 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.**

**Authors:**

Nicola L Harman1, https://orcid.org/0000-0001-6958-6466

Adrián Sanz-Moreno2, https://orcid.org/0000-0001-9478-5432

Stamatia Papoutsopoulou3, https://orcid.org/0000-0001-6665-8508

Katie A Lloyd4, https://orcid.org/0000-0002-2325-8050

Kamar E Ameen-Ali5, https://orcid.org/0000-0002-7583-4099

Malcolm Macleod6  https://orcid.org/0000-0001-9187-9839

Paula R Williamson1 https://orcid.org/0000-0001-9802-6636

Affiliations

1. Department of Health Data Science, Institute of Translational Medicine, University of Liverpool, Liverpool, UK, L69 3GL
2. German Mouse Clinic, Institute of Experimental Genetics, HMGU, Neuherberg 85764, Germany
3. Cellular and Molecular Physiology, Institute of Translational Medicine, University of Liverpool, Liverpool, UK, L69 3GL
4. Clinical Translational Research Innovation Centre (CTRIC), Altnagelvin Hospital, University of Ulster, Londonderry/Derry, UK, BT47 6SB
5. Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, UK NE4 5PL
6. Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

Corresponding Author

Nicola L Harman

Tel: +44 151 795 8771

Email: [n.harman@liv.ac.uk](mailto:n.harman@liv.ac.uk)

Address: Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, Liverpool, UK, L69 3GL

**Abstract**

Background

In pre-clinical research, systematic reviews have the potential to mitigate translational challenges by facilitating understanding of how pre-clinical studies can inform future clinical research . Yet their conduct is encumbered by heterogeneity in the outcomes measured and reported, and those outcomes may not always relate to the most clinically important outcomes. We aimed to systematically review outcomes measured and reported in pre-clinical *in vivo* studies of pharmacological interventions to treat high blood glucose in mouse models of type 2 diabetes.

Methods

A systematic review of pre-clinical in vivo studies of pharmacological interventions aimed at addressing elevated blood glucose in mouse models of type 2 diabetes was completed. Studies were screened for eligibility and outcomes extracted from the included studies. The outcomes were recorded verbatim and classified into outcome domains using an existing outcome taxonomy. Outcomes were also compared to those identified in a systematic review of registered phase 3/4 clinical trials for glucose lowering interventions in people with type 2 diabetes

Results

Review of 280 included studies identified 532 unique outcomes across 19 domains. No single outcome, or domain, was measured in all studies and only 132 (21%) had also been measured in registered phase 3/4 clinical trials. A core outcome set, representing the minimum that should be measured and reported, developed for type 2 diabetes effectiveness clinical trials includes 18 core outcomes, of these 12 (71%) outcomes were measured and reported in one or more of the included pre-clinical studies.

Conclusions

There is heterogeneity of outcomes reported in pre-clinical research. Harmonisation of outcomes across the research pathway using a core outcome set may facilitate interpretation, evidence synthesis and translational success, and may contribute to the refinement of the use of animals in research.

Systematic review registration

The study was prospectively registered on the PROSPERO Database, registration number CRD42018106831

**Keywords:**

**Type 2 diabetes, pre-clinical research, mouse models of type 2 diabetes, core outcome set, translational research, clinical research, pre-clinical systematic review**

**Background**

Clinical trials are undertaken to evaluate the effectiveness and safety of treatments in defined populations, and use pre-defined outcome measures. However, there is often marked variability between trials in the outcomes measured and reported which contributes to research waste through the inability to compare findings and synthesise evidence from multiple trials(1). These issues can be addressed through the use of a core outcome set (COS), defined as “the minimum [set of outcomes] that should be measured and reported in all clinical trials of a specific condition” (2). Indeed, in the case of rheumatoid arthritis, a core outcome set has increased the consistency of outcome reporting, and use of the COS has increased over time (3). There are currently 337 COSs spanning 31 disease areas that have been developed for clinical research or practice or both (4). However, despite the uptake of COSs in clinical trials, little is little about their relationship to the outcomes measured at other stages of the research pathway.

In pre-clinical research there is a recognised issue in the ability of animal models to predict effectiveness in humans, with large variability in translational success rates (5-7). Pre-clinical systematic reviews have been proposed as a way to improve understanding of pre-clinical effectiveness and how this can inform clinical trials (8-10).Yet, as in clinical trials, the ability to systematically review the literature is impacted by issues of methodological rigour and further compounded by heterogeneity in the outcomes measured and reported (1, 11). Initiatives to improve the reporting of methodological details, for instance, the Animal Research: Reporting In Vivo Experiments (ARRIVE) guidelines (12), do not consider the choice of study outcome(s), that may impact not only on the ability to systematically compare and contrast the study results but also on the translatability of pre-clinical research to later phase trials. While the STAIR criteria for preclinical stroke research does address this issue, their requirements are rather broad (13). We sought to explore the issue of outcome heterogeneity in pre-clinical research and the potential application of a COS approach using type 2 diabetes research as a case study. Type 2 diabetes is a global health concern; it has been estimated that 700 million people aged 20-79 will be affected by diabetes by 2045, the majority of these cases being type 2 diabetes (14-16). There are a number of established animal models for the study of type 2 diabetes(17, 18) with mouse models offering several advantages, including ease of induction of type 2 diabetes, a relatively short breeding span, and availability of physiological and invasive testing (19). Mice are widely used in endocrine in metabolic research and, if all research areas are taken into account, represent the most widely used animal model in pre-clinical research(20). We aimed to systematically review outcomes measured in pre-clinical research for type 2 diabetes using a mouse model, and to compare these to outcomes measured in clinical trials of glucose lowering interventions in type 2 diabetes (21). Finally we examine the extent of the applicability of an existing COS(22) for type 2 diabetes in pre-clinical research.

**Methods**

**Search strategy**

Relevant pre-clinical animal studies were identified with a combined search of MEDLINE, PubMED and SCOPUS using the search terms specific to each database (Table 1). Searches were undertaken on the 16th July 2018.

Returned entries were exported to Endnote, screened for duplicates and then uploaded into the CAMARADES-NC3Rs Preclinical Systematic Review & Meta-analysis Facility (SYRF) ([www.syrf.org.uk](http://www.syrf.org.uk), accessed 3rd March-2020) for screening.

The database held on [www.preclinicaltrials.eu](http://www.preclinicaltrials.eu) was also searched for ongoing registered trials of glucose lowering interventions for diabetes, but none were identified (July 2018).

The study protocol and search strategy was prospectively registered on the PROSPERO international prospective register of systematic reviews (<https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018106831>).

**Eligibility criteria**

Publications reporting a pharmacological intervention aimed at lowering blood glucose in a mouse model for type 2 diabetes were eligible for inclusion. Eligible mouse models included dietary induced, chemically induced, monogenic or polygenic models. To be eligible for inclusion, studies must have been undertaken in the context of type 2 diabetes and not solely in other related metabolic disorders, for example, metabolic syndrome, obesity or insulin resistance. There were no restrictions on the year of publication.

Studies were excluded if they met any of the following exclusion criteria: publications reporting the use of mouse models in other related metabolic disorders but not in type 2 diabetes; publications focusing on the prevention of type 2 diabetes only; publications reporting interventions in other animal models; publications reporting *in vitro* studies only; publications reporting non-pharmacological interventions for type 2 diabetes; publications primarily focused on interventions for complications of type 2 diabetes (e.g. retinopathy, neuropathy, cardiovascular disease, gastroparesis); publications reporting interventions exclusively for type 1 diabetes or gestational diabetes; studies that are solely mechanistic. Publications were also excluded if they use a mouse model inappropriate for the study of type 2 diabetes including, but not limited to, non obese diabetic mice, Akita mice, viral induced diabetes, alloxan induced diabetes, high dose streptozotocin (100-200 mg/kg). Streptozotocin models were included if a low dose was used to induce diabetes and the study specified the use of the model in the context of type 2 diabetes.

**Assessment of study eligibility**

Abstracts were reviewed in duplicate by a team of reviewers (NH, AS-M, SP, KL and KAA). Due to the large number of included abstracts, a 25% sample (in 10-year blocks) was taken forward to full text review. At full text review, a 10% duplicate screening batch check was completed for each reviewer before proceeding with single review. Where disagreement or uncertainty about inclusion of a study was noted, the reviewers discussed the study before reaching a decision. No study required third reviewer arbitration.

**Data extraction**

Data extraction from included full texts was undertaken by NH. Data on the year of publication, region of work and the mouse model used was extracted along with the outcomes measured. Data on outcomes was extracted from the methods and results sections of papers along with figures, tables and supplementary material where available. In cases of composite outcomes, all component outcomes were extracted. Where data on a specific adverse event was collected the outcome was listed twice, once as an adverse event and once as the specific outcome.

**Outcome classification**

Each outcome was reviewed and grouped with other outcomes if they measured the same aspect albeit using a different method. Each outcome was categorised according to the COMET taxonomy(23). This taxonomy comprises 38 domains under five areas (death, physiological/clinical, life impact, resource use and adverse events). Outcome grouping and categorisation was cross checked by AS-M.

**Comparison with clinical trials**

Outcomes were then compared to those identified from a previous review of phase 3/4 registered clinical trials(21) and to those included in the core outcome set for type 2 diabetes(22).

**Results**

**Characteristics of included studies.**

A systematic review was performed, to identify relevant pre-clinical *in vivo* studies using a mouse model of type 2 diabetes. A sample of 25% of included abstracts was assessed for eligibility at the full text stage and outcomes extracted from 280 eligible studies (Figure 1). All studies used a mouse model of type 2 diabetes with the majority (63%) using a genetic model, for example, KK-Ay or Lepr db/dbmice. The characteristics of the included studies are described in Table 2.

**Outcomes measured in pre-clinical studies**

A total of 2874 individual outcomes were extracted with a median of 8 outcomes per trial (range 1-46). Each outcome was reviewed and categorised using the COMET taxonomy(23) (Table 3). Outcomes were also tagged if they had also been measured in phase 3/4 clinical trials in Type 2 diabetes identified in a previous systematic review (21).

The 2874 outcomes represented 532 unique outcomes across 19 domains. Of the unique outcomes, 205 (39%) represented outcomes relevant to the mechanism of drug action rather than safety or efficacy. No single outcome was measured in all studies. The most frequently represented domain was “metabolism and nutrition” with all but one (279/280) measuring one or more outcomes within the domain. Within this domain, 90% (253/279) measured blood or plasma glucose or both; and 99% (277/279) either blood/plasma glucose, tissue glucose, glycaemic control, glucose tolerance, hypoglycaemia or urinary glucose. 171 of 280 studies (61%) reported “general outcomes” (not attributed to a certain body system), for example, outcomes relating to body weight or composition (166 of 280, 59%).164 of 280 studies (59%) included an outcome in the “endocrine outcomes” domain and of these 151 (92%) included an outcome relating to insulin, c-peptide or glucagon. Adverse events or effects were less frequently reported with only 7% of trials including one or more outcomes in this domain.

**Comparison with outcomes measured in later phase clinical trials**

All domains measured in pre-clinical *in vivo* studies had also been measured in phase 3/4 clinical trials. The distribution of outcomes across the COMET domains was similar between pre-clinical and clinical studies with the exception of “vascular”, “cardiac”, “adverse events” and “delivery of care” outcomes that were more prevalent in phase 3/4 trials; and “endocrine outcomes”, that were more frequently measured pre-clinically. The clinical trials also included outcomes in an additional 11 domains (Table 3). Of these additional domains, “economic”, “hospital”, “role functioning” and “perceived health status” could only relate to human intervention studies.

Importantly, of the 532 unique outcomes reported pre-clinically, only 21% had also been measured in type 2 diabetes clinical trials. This may reflect the prevalence of mechanistic outcomes in pre-clinical studies, or greater feasibility for measuring certain outcomes in the pre-clinical setting compared with clinical trials.

**Comparison with an existing core outcome set**

Core outcome sets (COS) represent the minimum set of outcomes that should be measured and reported in every clinical trial of a specific area of health(24). Their purpose is to reduce the heterogeneity in outcomes measured in clinical trials of a particular condition, facilitate evidence synthesis, and promote the measurement of outcomes relevant to all stakeholders. A COS for glucose lowering interventions for type 2 diabetes has been developed(22) and the outcomes measured pre-clinically were compared to this. The core outcome set includes 18 outcomes, and of these 17 could potentially be measured in a mouse model. Twelve (71%) of these core outcomes were represented to some extent in the outcomes measured in pre-clinical studies (Table 4). However, studies typically measured only 1 or 2 outcomes and no single study reported more than seven outcomes in the COS (figure 2). Similar patterns were observed in clinical trials, registered prior to the publication of the COS, although a larger proportion of trials measured multiple core outcomes. In pre-clinical studies there were multiple outcomes reported that could be used to measure a core outcome (table 4). Furthermore within these there were multiple methods of assessment. For example, “glycaemic control” was reported in 40 studies using four different outcomes of which glycated haemoglobin (HbA1c) was the most frequently measured (35/40 studies). There are four, commonly used, methods for the measurement of HbA1c(25), details of the method used were reported in 29/35 papers. Each of the four methods was used at least once with immunoassay used most frequently (n=18) followed by, ion-exchange high-performance liquid chromatography (HPLC) (n=4), boronate affinity HPLC (n=6), and enzymatic assays (n=1), highlighting the variability in “how” outcomes are measured.

Outcomes in the COS that were not reported, or infrequently reported, represented longer term outcomes associated with morbidities (nephropathy, neuropathy, retinopathy, cardiovascular and cerebrovascular disease) resulting from long term insulin resistance(26). Measuring such complications in otherwise healthy, usually adolescent laboratory mice where the rates of spontaneous development of these complications of diabetes is low would bring some challenges, including the ethical and resource costs of the longer duration of experiments which would be required.

**Discussion**

Addressing methodological issues in pre-clinical research will help researchers take one step closer to achieving successful translation of safe and effective treatments (12, 27, 28). However, the issue of outcome heterogeneity in pre-clinical studies, demonstrated here for type 2 diabetes, has been overlooked, impacting on the ability to synthesise evidence, contributing to research waste and widening the translational gap.

Initiatives to harmonise outcomes have focused on later phase effectiveness trials (2, 24) yet there is potential to apply COS across the research pathway. In the case of type 2 diabetes, over 70% of the existing COS was measured, to some extent, in publications using pre-clinical mouse models. Discordance was observed for outcomes relating to long term complications but these too were infrequently measured in clinical trials with “gangrene and amputation of the leg, foot or toe”, and “cerebrovascular disease” not measured at all and “deterioration of vision” and “myocardial infarction” each measured in a single clinical trial (21). There were also limits of the pre-clinical search strategy which excluded studies that used specific mouse models of a long term diabetes complication. Mice display a different rate of development/aging to humans that cannot easily be converted between species (29, 30) ; consequently, disease progression and time to the onset of long term complications may not be feasible to assess unless a specific animal model, pre-disposed to the development of such conditions, is used.

Death is reported in clinical trials, either as a specific outcome or in the collection of serious adverse events, and is routinely recorded in clinical practice but in the pre-clinical setting it is widely accepted that “death as an endpoint to a procedure should be avoided as far as possible and replaced by earlier, humane endpoints”(31). Instead, alternative surrogate outcomes for death may be more appropriate and alleviate terminal distress in mice whilst also capturing the core outcome.(32, 33)

In the present study we have applied surrogate markers of quality of life including “food and water intake”, measured in 39% of studies. Yet in these studies the reason for measurement was not specified and, in the case of some diabetes treatments, may indicate assessment of a side effect of treatment (weight gain) or polydipsia (a symptom of elevated blood glucose). Animal welfare encompasses an animal’s overall quality of life, taking into consideration its physical and psychological health along with the suitability of living conditions that give the animal opportunity to exhibit natural behaviours. Assessment of welfare is a critical component of research involving animals but this may go unreported in study publications. This under-reporting is further compounded by multiple methods of assessment and clarity is needed on how quality of life should be measured(34).

A COS has the potential to reduce the risk of outcome reporting bias, an issue that has been highlighted in both pre-clinical and clinical research (35, 36) . For a COS to contribute to reducing the risk of reporting bias there is an expectation that it is used in its entirety or that clear justification is made for why some outcomes have not been measured. It is important to recognise that it may not be practical, or indeed ethical, to measure the full set of core outcomes in every pre-clinical study and instead data on the core outcomes may be collected by the triangulation of data from multiple pre-clinical studies. Using glycaemic control as an example in the present study it is clear that not only are there different ways to define the outcome but, in the case of HbA1c, multiple methods of measurement. For pre-clinical studies to apply the COS there needs to be clear reporting on which of the outcomes will be measured, including reasons why outcomes are not assessed, together with consensus on “how” each of the core outcome should be measured and further work is warranted in this area.

**Conclusion**

The COS developed for type 2 diabetes shows a large overlap with outcomes already measured and reported in pre-clinical research using a mouse model of type 2 diabetes. Whilst mouse models are the most frequently used pre-clinical animal model, review of large animal and non-human primate models may identify further overlap of outcomes resulting from the ability to perform particular physiological assessments in these larger animals. Application of the COS, using agreed methods, in both pre-clinical research and clinical trials will mean that the same outcomes are measured and reported as a minimum, across the research pathway, facilitating evidence synthesis that has the potential to identify the most promising treatments.

**List of abbreviations**

COS, core outcome set

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

**Availability of data and materials**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. A full list of included studies used to identify outcomes is provided in supplementary file 1.

**Competing interests**

PRW chairs the COMET Management Group.

**Funding**

This work has received funding from the European Union's Horizon 2020 research and innovation programme (CORBEL, under grant agreement n° 654248). The study funder was not involved in the design of the study; the collection, analysis, and interpretation of data; writing the report; and did not impose any restrictions regarding the publication of the report.

**Authors' contributions**

### PRW and NH conceived and designed the study and wrote the study protocol. KAA, AS-M, KL,SP and MM reviewed the protocol. PRW, NH, MM developed the search strategy. NH, KAA, AS-M, KL, SP, NH screened abstracts and full text reports. NH extracted and categorised outcomes. All authors contributed to the interpretation of the data. NH drafted the manuscript; all authors critically reviewed and approved the final manuscript.

**Acknowledgements**

We acknowledge themembers of the wider CORBEL 3.2 work package study team: Serena Battaglia, Jacques Demotes-Mainard, Valerie Gailus-Durner, Silvio Garattini, Cecilia AC Prinsen, Michael Raess, Patricia da Silva-Buttkus, Caroline B Terwee.

**References**

1. Clarke M, Williamson PR. Core outcome sets and systematic reviews. Systematic Reviews. 2016;5(1):11.

2. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: version 1.0. Trials. 2017;18(3):280.

3. Kirkham JJ, Clarke M, Williamson PR. A methodological approach for assessing the uptake of core outcome sets using ClinicalTrials.gov: findings from a review of randomised controlled trials of rheumatoid arthritis. BMJ. 2017;357:j2262.

4. Gargon E, Gorst SL, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 5th annual update to a systematic review of core outcome sets for research. PLOS ONE. 2019;14(12):e0225980.

5. Leenaars CHC, Kouwenaar C, Stafleu FR, Bleich A, Ritskes-Hoitinga M, De Vries RBM, et al. Animal to human translation: A systematic scoping review of reported concordance rates. Journal of translational medicine. 2019;17(1).

6. Hackam DG, Redelmeier DA. Translation of Research Evidence From Animals to Humans. JAMA. 2006;296(14):1727-32.

7. Sena ES, van der Worp HB, Bath PMW, Howells DW, Macleod MR. Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy. PLOS Biology. 2010;8(3):e1000344.

8. Hooijmans CR, Ritskes-Hoitinga M. Progress in Using Systematic Reviews of Animal Studies to Improve Translational Research. PLOS Medicine. 2013;10(7):e1001482.

9. Pound P, Ritskes-Hoitinga M. Can prospective systematic reviews of animal studies improve clinical translation? Journal of translational medicine. 2020;18(1):15.

10. Sandercock P, Roberts I. Systematic reviews of animal experiments. Lancet (London, England). 2002;360(9333):586.

11. Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I. Where is the evidence that animal research benefits humans? British Medical Journal. 2004;328(7438):514-7.

12. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLOS Biology. 2010;8(6):e1000412.

13. Recommendations for Standards Regarding Preclinical Neuroprotective and Restorative Drug Development. Stroke. 1999;30(12):2752-8.

14. Holman N, Young B, Gadsby R. Current prevalence of Type 1 and Type 2 diabetes in adults and children in the UK. Diabetic medicine : a journal of the British Diabetic Association. 2015;32(9):1119-20.

15. Unnikrishnan R, Pradeepa R, Joshi SR, Mohan V. Type 2 Diabetes: Demystifying the Global Epidemic. Diabetes. 2017;66(6):1432-42.

16. Federation ID. Diabetes Atlas Ninth Edition 2019 2019 [6th April 2020]. Available from: https://diabetesatlas.org/en/resources/.

17. Kleinert M, Clemmensen C, Hofmann SM, Moore MC, Renner S, Woods SC, et al. Animal models of obesity and diabetes mellitus. Nature reviews Endocrinology. 2018;14(3):140-62.

18. Neubauer N, Kulkarni RN. Molecular Approaches to Study Control of Glucose Homeostasis. ILAR Journal. 2006;47(3):199-211.

19. Cefalu WT. Animal models of type 2 diabetes: clinical presentation and pathophysiological relevance to the human condition. Ilar j. 2006;47(3):186-98.

20. Report from the Comission to the European Parliament and the Council . 2019 report on the statistics on the use of animals for scientific purposes in the Member States of the European Union in 2015-2017. COM/2020/16 fina. European Comission 2019.

21. Harman NL, James R, Wilding J, Williamson PR. SCORE-IT (Selecting Core Outcomes for Randomised Effectiveness trials In Type 2 diabetes): a systematic review of registered trials. Trials. 2017;18(1):597.

22. Harman NL, Wilding JPH, Curry D, Harris J, Logue J, Pemberton RJ, et al. 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 Research &amp; Care. 2019;7(1):e000700.

23. Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. Journal of Clinical Epidemiology. 2018;96:84-92.

24. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13:132.

25. Little RR, Roberts WL. A review of variant hemoglobins interfering with hemoglobin A1c measurement. J Diabetes Sci Technol. 2009;3(3):446-51.

26. Nathan DM. Long-term complications of diabetes mellitus. The New England journal of medicine. 1993;328(23):1676-85.

27. Ali Z, Chandrasekera PC, Pippin JJ. Animal research for type 2 diabetes mellitus, its limited translation for clinical benefit, and the way forward. Alternatives to laboratory animals : ATLA. 2018;46(1):13-22.

28. Pound P, Ritskes-Hoitinga M. Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. Journal of translational medicine. 2018;16(1):304.

29. Dutta S, Sengupta P. Men and mice: Relating their ages. Life sciences. 2016;152:244-8.

30. Agoston DV. How to Translate Time? The Temporal Aspect of Human and Rodent Biology. Frontiers in Neurology. 2017;8(92).

31. Directive', C., 2010-63-EU, art. 13.3. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes Text with EEA relevance, (2010).

32. Ray MA, Johnston NA, Verhulst S, Trammell RA, Toth LA. Identification of markers for imminent death in mice used in longevity and aging research. Journal of the American Association for Laboratory Animal Science : JAALAS. 2010;49(3):282-8.

33. Littin K, Acevedo A, Browne W, Edgar J, Mendl M, Owen D, et al. Towards humane end points: behavioural changes precede clinical signs of disease in a Huntington's disease model. Proceedings Biological sciences. 2008;275(1645):1865-74.

34. Adrián Sanz-Moreno PdS-B, Cecilia Prinsen, Caroline Terwee, Michael Raess, Valérie Gailus-Durner, Helmut Fuchs, Martin Hrabe de Angelis. Assessing quality of life, fatigue and wellbeing in mouse models of disease - A systematic review 2018 [9-4-2020]. Available from: https://[www.crd.york.ac.uk/PROSPERO/display\_record.php?RecordID=103507](http://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=103507).

35. Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E, Howells DW, et al. Evaluation of Excess Significance Bias in Animal Studies of Neurological Diseases. PLOS Biology. 2013;11(7):e1001609.

36. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One. 2008;3(8):e3081.

Figure legends

**Figure 1. PRISMA flow diagram**

**Figure 2: Proportion of pre-clinical and clinical studies measuring one or more of the core outcomes**