Coordinated Research Infrastructures Building Enduring Life-science services
- CORBEL -

Deliverable D3.5

Core set of outcome measures for clinical trials for each disease condition

WP3 – Community Driven Cross-Infrastructure joint research - Medical

Lead Beneficiary: ECRIN -ERIC
WP leader: Jacques Demotes (ECRIN-ERIC)
Contributing partner(s): INFRAFRONTIER, U LIVERPOOL, VU/VUMC, IRFMN, EATRIS-ERIC

Contractual delivery date: 31 August 2017
Actual delivery date: 6 September 2017

Authors of this deliverable: Serena Battaglia, Patricia Da Silva-Buttkus, Jacques Demotes, Helmut Fuchs, Valérie Gailus-Durner, Silvio Garattini, Nicola Harman, Cecilia Prinsen, Michael Raess, Caroline Terwee, Paula Williamson

Grant agreement no. 654248
Horizon 2020
H2020-INFRADEV-1-2014
Type of action: RIA

Content

Executive Summary

Animal models play a crucial role in understanding the mechanisms of diseases and symptoms, and to test the efficacy and safety of treatments before conducting clinical trials. Selecting human diseases and developing core sets of clinical outcome measures as well as complementary sets of mouse phenotyping measures will contribute to narrow the gap between animal models and human trials.

Project objectives

With this deliverable, the project has reached/this deliverable has contributed to the following objectives:

a) Define the appropriate strategy for the successful implementation of the task
b) Selection of one model disease (type 2 diabetes mellitus)

Detailed report on the deliverable

Background

Mice are widely used in biomedical research to gain insight to the gene function in human health and diseases, to act as disease models to elucidate the involved pathways and the effects of treatments, and to support the development of (genome-based) treatments for human diseases. Developing a complementary set of mouse phenotyping assays and of clinical outcomes for specific diseases will contribute to improve the predictive value of the mouse model.

Description of Work

The overall objective of the CORBEL task 3.2 is to promote convergence between outcome measures used for clinical trials in humans and phenotyping techniques used for testing treatments in animal models.
The group composed of experts from INFRAFRONTIER – German Mouse Clinic (mouse phenotyping techniques), COMET (Core Outcome Measures in Effectiveness Trials), COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) initiatives and IRFMN (Istituto di Ricerche Farmacologiche Mario Negri) identified four issues:

1) Animal phenotyping as performed at the German Mouse Clinic focuses on the description of “symptoms” to characterize the phenotype of mutant mice (which would be comparable to the diagnostic criteria in medical practice, or to the inclusion criteria in a clinical trial), whereas the outcome measures chosen in effectiveness clinical trials are not selected to comprehensively describe the symptoms, but to measure the amplitude of the treatment-induced improvement that is relevant for the patient wellbeing
2) A significant number of outcome measurement instruments used in human trials refer to non-physiological assessments, using subjective rating scales for fatigue, quality of life, wellbeing etc

3) The German Mouse Clinic focuses on the description of phenotypes of mutant mice, measured with more than 500 quantitative and qualitative parameters. If any treatment were to be used, the effect of pharmacological agents and treatments is therefore based on changes of the above parameters and not on subjective rating scales (questionnaires, etc) as in the clinical setting
4) Another aspect that has to be considered lies in the ethical authorization obtained for mouse phenotyping at the German Mouse Clinic, as it is restricted to the battery of tests currently available. Any development of a new phenotyping technique would require additional authorization.

Accordingly, the group defined a strategy consisting of two different approaches to address these issues in a complementary and comprehensive way:

1) Identify a disease where a Core Outcome Set is currently not established

   For this first approach, the strategy plan foresees three steps:
   - Develop Core Outcome Set (COS) for the selected disease through a Delphi process, followed by a consensus meeting (task driven by U Liverpool)
   - Define instruments to measure the developed COS in humans (task driven by VU/VUmc)
   - Identify matching instruments in mice (task led by INFRAFRONTIER - GMC)

   The three activities will be performed in parallel, with the different actors participating as observers and providing information on what already exists, and what would be easy or difficult to measure.

   This study represents the opportunity to test a new methodology procedure: indeed, the current development of COS and measurement techniques is sequential (first "what" to measure, then "how" to measure it); introducing an input from downstream partners may improve the implementation of the "how to measure" in humans and the equivalent parameters in mice.

   Type 2 diabetes mellitus was selected as a model disease. A search of the COMET database confirmed that no COS for this condition had been developed or was in development. In addition, type 2 diabetes mouse models are well described in the literature (Cefalu, 2006; King, 2012).

   A search of the clinicaltrials.gov registry identified 138 eligible trials from which 1444 individual outcomes were identified and categorized according to the Williamson-Clarke taxonomy (Appendix 1). This work confirms that there is no single outcome measured across all registered trials. Whilst the ICF (International Classification of Functioning, Disability and Health) recommends the use of their brief core set in clinical trials this is too large and is not used by current, open registered trials highlighting the need to develop a COS for the treatment of hyper glycaemia in type 2 diabetes.

   A paper describing in details the systematic review will shortly be submitted for publication.

2) Identify mouse phenotyping animal tests that approximate the subjective rating scales used to assess fatigue, quality of life or wellbeing in humans

   The instruments to assess fatigue and wellbeing in animals will be compared to the measurement instruments for humans.

As a first step the German Mouse Clinic and VU/VUmc prepared inventories respectively of proxy measures to analyse quality of life, fatigue and wellbeing in mice (Appendix 2) and the already available measurement instruments in humans (Appendix 3).

Next steps

Approach 1

- In order to develop a COS for Type 2 diabetes mellitus, a group of experts (health care professionals and patients) will be convened to advise on the next stages of the COS development, namely an online Delphi (autumn 2017, two rounds), followed by a face to face consensus meeting (Spring 2018). This work will determine "what" is important to measure in clinical effectiveness studies. As mentioned above, COSMIN and INFRAFRONTIER – GMC will participate as observers
- Meanwhile VU/VUmc will start a systematic review of the Patient Reported Outcome Measures (PROMs) in order to provide input after the Delphi process in relation to "how" to measure particular outcomes in humans; again, INFRAFRONTIER – GMC will participate as an observer
- Based on the interim results of the Delphi process taken from the first round, INFRAFRONTIER – GMC will start the comparison with the measurements in mice, in particular for the life impact categories
- A paper describing the innovative methodology procedure involving in parallel these different areas of expertise will be prepared.

Approach 2

- The measurement instruments in humans and mice should be compared to identify similar tests. According to the results of the comparison, we will consider to translate the existing questionnaires to assess wellbeing, fatigue and quality of life in humans into a new observation tool for mice. In this process, we will incorporate the experience of the IMPC (International Mouse Phenotyping Consortium) for assessing wellbeing in aging mutant lines.

References

Cefalu WT (2006). Animal models of type 2 diabetes: clinical presentation and pathophysiological relevance to the human condition. ILAR J: 186-98

King AJ (2012). The use of animal models in diabetes research. Br J Pharmacol. 166: 877-94

Abbreviations

COMET	Core Outcome Measures in Effectiveness Trials
COS	Core Outcome Set
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
ICF	International Classification of Functioning, Disability and Health
GMC	German Mouse Clinic
SYRCLE	Systematic Review Center for Laboratory animal Experimentation

Delivery and schedule

The delivery is delayed: NO

Adjustments made

No adjustments made

Appendices

Appendix 1 "Summary of tables from SR"

Summary of outcomes categorised into the Williamson/Clarke core outcome domains.				
Core area	**Core domains**	Number of trials including one or more outcome in core domain (%)	Number of outcomes included in core domain (%)	Number of trials including as a primary outcome[a]
Death	Mortality/survival	3 (2.2)	3 (0.2)	0
Physiological/ clinical	Blood and lymphatic system outcomes	9 (6.5)	19 (1.3)	1
	Cardiac outcomes	20 (14.5)	56 (3.9)	9
	Congenital, familial and genetic outcomes	0(0)	0 (0)	
	Endocrine outcomes	31(22.5)	50 (3.5)	7
	Ear and labyrinth outcomes	0 (0)	0 (0)	0
	Eye outcomes	2 (1.4)	2 (0.1)	0
	Gastrointestinal outcomes	5 (3.6)	20 (1.4)	2
	General outcomes	65 (47.1)	146 (10.1)	3
	Hepatobiliary outcomes	12 (8.7)	25 (1.7)	3
	Immune system outcomes	28 (20.3)	73 (5.1)	4
	Infection and infestation outcomes	4 (2.9)	8 (0.6)	0
	Injury and poisoning outcomes	0 (0)	0 (0)	0
	Metabolism and nutrition outcomes	121 (87.7)	582 (40.3)	92
	Musculoskeletal and connective tissue outcomes	2 (1.4)	2 (0.1)	1
	Outcomes relating to neoplasms: benign, malignant and unspecified (including cysts and polyps)	0 (0)	0 (0)	0
	Nervous system outcomes	6 (4.3)	16 (1.1)	2
	Pregnancy, puerperium and perinatal outcomes	0 (0)	0 (0)	0
	Renal and urinary outcomes	27 (19.6)	76 (5.3)	5
	Reproductive system and breast outcomes	0 (0)	0 (0)	0
	Psychiatric outcomes	2 (1.4)	2 (0.1)	0
	Respiratory, thoracic and mediastinal outcomes	3 (2.2)	11 (0.8)	1
	Skin and subcutaneous tissue outcomes	1 (0.7)	1 (0.1)	0
	Vascular outcomes	51 (37)	134 (9.3)	13
	Physical functioning	5 (3.6)	7 (0.5)	0

Summary of outcomes categorised into the Williamson/Clarke core outcome domains.				
Core area	**Core domains**	Number of trials including one or more outcome in core domain (%)	Number of outcomes included in core domain (%)	Number of trials including as a primary outcome[a]
Life impact	Social functioning	5 (3.6)	6 (0.4)	0
	Role functioning	3 (2.2)	6 (0.4)	0
	Emotional functioning/wellbeing	8 (5.8)	28 (1.9)	0
	Cognitive functioning	2 (1.4)	22 (1.5)	0
	Global quality of life	4 (2.9)	5 (0.3)	0
	Perceived health status	4 (2.9)	4 (0.3)	0
	Delivery of care	30 (21.7)	60 (4.2)	4
	Personal circumstance	0 (0)	0 (0)	0
Resource use	Economic	4 (4)	6 (0.4)	0
	Hospital	3 (2.2)	4 (0.3)	0
	Need for intervention	16 (11.6)	24 (1.7)	1
	Societal/carer burden	0 (0)	0 (0)	0
Adverse Events	Adverse events/effects	33 (23.9)	46 (3.2)	5

Appendix 2 "Proxy measures to analyse QL, fatigue and WB in mice"

Image by Cynthia McKelvey

Quality of life

Health-related quality of life research measures are examined "*at a descriptive level, providing data concerning the* ***impact of disease and treatment on the physical, functional, psychologic, and social health*** *of human populations*" (Aaronson, 1989). A corresponding assessment of quality of life in mice should encompass as many measurements of changes in the above health characteristics as possible. There are at least two considerations in mice phenotyping that may limit a clear mapping of clinical outcome measures and phenotyping measures that can be examined in mice: a) the absence of self-reporting, b) the use of provocative tests in which an external stimulus is provided to the mice, failing to study spontaneous behaviours (e.g. pain). Nevertheless, according to the vast experience at the German Mouse Clinic (GMC) and to literature researched in different mouse models, we can assume that performance in specific tests will correlate with changes in emotional, cognitive and physiological states and translate into measures of quality of life.
Assessment of quality of life in mice - ability to assess changes in:

1) **Open field** – assessment of locomotor and exploratory activities and reactivity to novelty
2) **Rotarod** – assessment of balance, grip strength and motor coordination
3) **Virtual Optokinetic Drum test** – assessment of eye functionality
4) **Light/Dark Box** - assessment of exploratory activity based on the innate aversion of mice to brightly illuminated spaces
5) **Elevated Plus Maze** – assessment of anxiety-related behaviour
6) **Social Interaction** – assessment of social anxiety
7) **Swimming test** – assessment of depression
8) **Y-Maze** – assessment of exploration of a novel environment and spatial working memory
9) **Object recognition** - assessment of exploration of an unfamiliar object
10) **IntelliCage** – automated assessment of cognitive function (requires subcutaneous implantation of passive transponders)

While the first six tests address the emotional experience of the mice such as anxiety-related behaviours (see review Hölter *et al*., 2015a), the last three tests assess changes in information processing and learning in mice (see review Hölter *et al*., 2015b).
The first three tests belong to the primary phenotyping pipeline of the GMC (meaning that they are routinely measured in all mouse lines in the GMC, Figure 1), whereas all other tests are part of secondary pipelines (additional investigations) established in the GMC and approved by the responsible animal welfare authority of the district government of Upper Bavaria, Germany. It is important to note that tests 4 to 10 are part of different secondary screen pipelines constructed to specifically examine the emotional experience (pipeline B in Figure 2), memory impairment (pipeline C in Figure 2) or motor deficits (pipeline D in Figure 2) in mice. Presently, there is no secondary pipeline that comprises all measurements described for quality of life assessment in mice.

Fatigue

Fatigue is a complex multidimensional symptom that is characterised by seven primary characteristics in humans (Barsevick *et al.* 2010): *"it is **subjective** (as assessed by self-reporting) and **unusual** (not proportional to prior activity and not relieved by rest); **physical sensations range from lassitude to exhaustion**, and the fatigue has a **negative impact on function** (decreased capacity for work, poor sleep quality, withdrawal from activities); there is **decreased cognitive ability** and an **unpredictable temporal course** (it can be either chronic or acute); the **negative emotions** associated with fatigue include helplessness, vulnerability, impatience, anxiety, and emotional numbness".*
Taking this into account, we identify below quantifiable measures inferred from mice behaviour that can be used for objectively testing fatigue in mice and that may have a translation value to human populations. One of the most widely used test for fatigue is the voluntary physical activity (Harrington, 2012) but other parameters described below, in combination with specific disease markers, may provide a comprehensive assessment of fatigue levels in mice.
Assessment of fatigue in mice, ability to assess changes in:

1) **Spontaneous wheel-running activity** – assessment of voluntary physical exercise
2) **Activity on cage floor** (burrowing) – representing activities of daily life
3) **Open field** - assessment of locomotor and exploratory activities
4) **Modified-SHIRPA** – assessment of general health, posture, reflexes and behavioral aspects (e.g. as vocalisation, tremor)
5) **Rotarod** - assessment of coordination and balance of poor running-wheel performance
6) **Grip strength** - measurement of muscle strength of fore limbs and combined fore and hind limbs
7) **Vertical pole** – assessment of motor co-ordination during turning behavior in mice
8) **Food and water consumption, body weight and sleep patterns**
9) **Blood cell count** (red blood cell (RBC) count for anemia, white blood cell count (WBC) for infection) and **cytokine levels** (interleukin 6 and tumor necrosis factor (TNF)).
10) **Body surface temperature**

The tests above are integrated in the primary phenotyping pipeline in the GMC except for the voluntary running wheel exercise, vertical pole and sleep pattern assessments. In addition, the activity on the cage floor and food and water intakes are quantified in the calorimetric cage system in the secondary pipeline H (not shown).

To date the GMC has not embarked on objectively developing a secondary pipeline to assess fatigue in mice.

Wellbeing

Nest building (non-maternal) and burrowing are two **spontaneous behaviours** in mice that represent activities of daily life and are a reliable assessment of wellbeing in mice (Jirkof, 2014).
Despite the use of scoring systems to examine welfare, or perhaps better its decline, of experimental mice in the GMC, we do not assess these two spontaneous behaviours in mice.

Summary

All tests described above were compiled from a literature search in various mouse models of human disease and make an attempt to evaluate the subjective components associated with quality of life, fatigue and wellbeing in mice.
Currently, despite the technical resources and scientific expertise, there is no specific secondary screen pipeline in the GMC that contains all potential proxy indicators for quality of life, fatigue and wellbeing in mice.

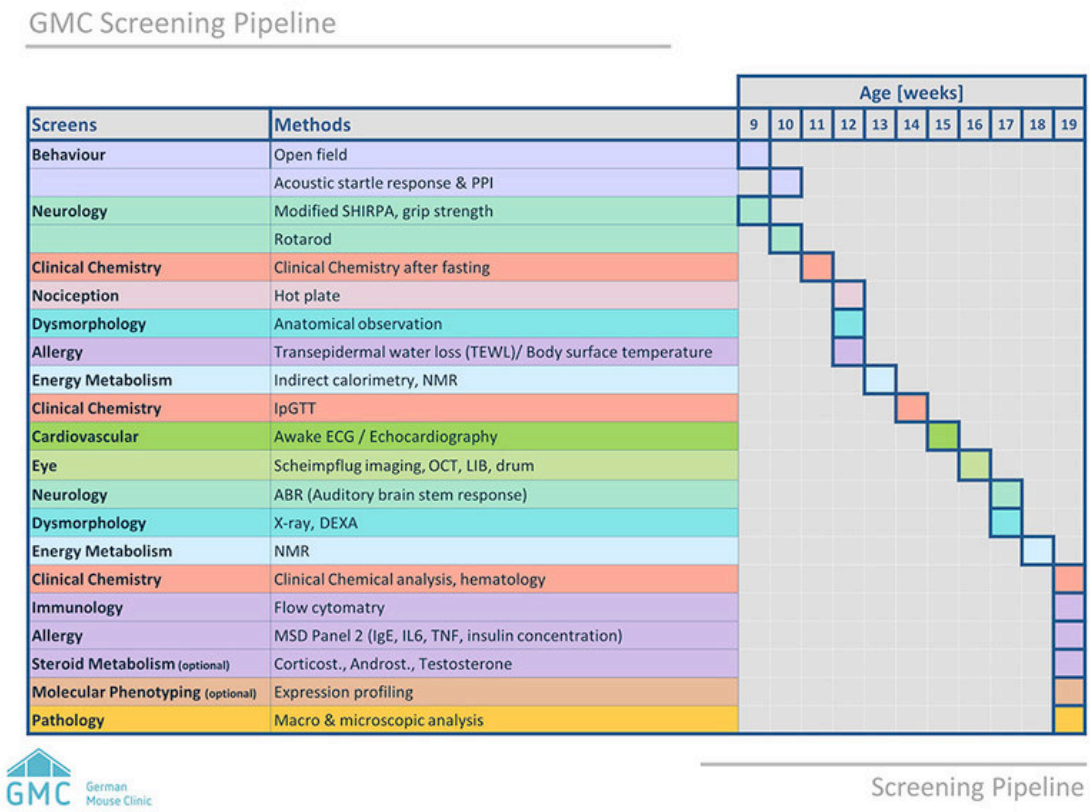

Figure 1. Primary screen flow in the GMC

B - Pipeline "Emotionality"

C - Pipeline "Memory impairment"

D - Pipeline "Motor disorders"

E - Pipeline "Sensory disorders"

Figure 2. Secondary screen pipelines in the GMC (tests highlighted in red are listed in Table 1)

Test/Measurement	Quality of life	Fatigue	Wellbeing
Open field	√	√	
Modified-SHIRPA		√	
Grip strength		√	
Rotarod	√	√	
Vision test	√		
Blood sampling		√	
Surface body temperature		√	
Body weight		√	
Light/dark box	√		
Elevated plus maze	√		
Social interaction	√		
Swimming test	√		
Y-Maze	√		
Object recognition	√		
Intellicage	√		
Wheel-running activity		√	
Vertical pole		√	
Food and water consumption		√	
Sleep pattern		√	
Nest building			Not assessed in GMC
Burrowing			Not assessed in GMC

Table 1- Overview of tests and measurements performed in mice in the GMC (grey- tests used in the primary screen workflow; blue - tests chosen from different pre-designed secondary pipelines)

References

Aaronson NK (1989) Quality of life assessment in clinical trials: methodologic issues. Control Clin Trials 10: 195S-208S

Barsevick, AM, Cleeland, CS, Manning, DC, O'Mara, AM, Reeve, BB, Scott, JA and Sloan, JA (2010) ASCPRO Recommendations for the Assessment of Fatigue as an Outcome in Clinical Trials J Pain Symptom Manage 39: 1086–1099.

Harrington ME (2012) Neurobiological studies of fatigue. Prog Neurobiol. 99:93-105

Hölter SM, Einicke J, Sperling B, Zimprich A, Garrett L, Fuchs H, Gailus-Durner V, Hrabé de Angelis M & Wurst W (2015a). Tests for Anxiety-Related Behavior in Mice. Curr Protoc Mouse Biol. 5:291-309.

Hölter SM, Garrett L, Einicke J, Sperling B, Dirscherl P, Zimprich A, Fuchs H, Gailus-Durner V, Hrabě de Angelis M, Wurst W (2015b). Assessing Cognition in Mice. Curr Protoc Mouse Biol. 5:331-58

Jirkof P (2014) Burrowing and nest building behavior as indicators of well-being in mice. J Neurosci Methods. 234:139-46

Appendix 3 "Mapping outcomes in humans"

Levels of health outcomes (1)	Health outcomes	Type of outcome measurement instrument	Outcome measurement instruments used in humans
Biological and Physiological variables	Blood sampling	Laboratory assessments	Not further specified
	Body surface temperature	Medical device	Thermometer
	Body weight	Medical device	Scale
	Grip strength	Medical device	Handgrip dynamometer
Symptom status	Pain	Generic instruments	PROMIS
		Disease-specific instruments	KOOS-pain
		Observer-reported outcome measures	Critical-Care Pain Observation Tool (CPOT)
	Fatigue	Generic instruments	PROMIS
		Disease-specific instruments	Parkinson's Disease Fatigue Scale (PFS-16)
		Proxy instruments	PROMIS Partent-Proxy Fatigue
		Performance-based tests	
Functional status	Physical functioning	Generic instruments	PROMIS Physical Functioning Short Forms
			SF-36 Physical Functioning
		Disease-specific instruments	KOOS-ADL
		Performance-based tests	6-minute walk test
			Stair climbing test (SCT)
			Timed up and go test (TUG)
			30-second chair stand test (CST)
	Social functioning	Generic instruments	SF-36 Social Functioning
			PROMIS Ability to perform social roles and activities
		Observer-reported outcome measures	
	Mental functioning – Anxiety	Generic instruments	PROMIS Anxiety
			Hospital Anxiety and Depression Scale (HADS)
		Observer-reported outcome measures	
	Mental functioning - Depression	Generic instruments	PROMIS Pediatric Depressive Symptoms
			Hospital Anxiety and Depression Scale (HADS)
		Performance-based tests	
General health perceptions	Self-rated health	Generic instruments	PROMIS
			SF-36, item 1

		Observer-reported outcome measures	
Overall quality of life	Overall quality of life, well-being	Generic instruments	PROMIS
		Disease-specific observation instruments	Qualidem questionnaire for dementia
		Performance-based tests	
		Observer-reported outcome measures	
		Medical device	Thermometer

(1) Wilson & Cleary model; JAMA, 1995