**A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery**

Susanna Dodd1, Mike Clarke2, Lorne Becker3, Chris Mavergames4, Rebecca Fish5, Paula R Williamson1§

1 MRC North West Hub for Trials Methodology Research, Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, L69 3GS, UK

2 Northern Ireland Methodology Hub, Queen's University Belfast, Belfast, UK

3 Department of Family Medicine, SUNY Upstate Medical University, Syracuse New York, USA

4 Informatics and Knowledge Management Department, Cochrane Central Executive, Freiburg, Germany

5 Division of Molecular and Clinical Cancer Sciences, University of Manchester, UK

§Corresponding author:

Professor Paula Williamson

Department of Biostatistics

University of Liverpool

Block F Waterhouse Building

1-5 Brownlow Street

Liverpool

L63 3GL

Phone: 0151 794 9758

Email addresses:

SD: [s.r.dodd@liverpool.ac.uk](mailto:s.r.dodd@liv.ac.uk)

MC: [m.clarke@qub.ac.uk](mailto:m.clarke@qub.ac.uk)

LB: [LBecker@cochrane.org](mailto:LBecker@cochrane.org)

CM: [CMavergames@cochrane.org](mailto:CMavergames@cochrane.org)

RF: [Rebecca.Fish-2@manchester.ac.uk](mailto:Rebecca.Fish-2@manchester.ac.uk)

PW: prw@liverpool.ac.uk

# Abstract

## Background

There is increasing recognition that insufficient attention has been paid to the choice of outcomes measured in clinical trials.

The lack of a standardised outcome classification system results in inconsistencies due to ambiguity and variation in how outcomes are described across different studies. Being able to classify by outcome would increase efficiency in searching sources such as clinical trials registries, patient registries, the Cochrane Database of Systematic Reviews and the COMET (Core Outcome Measures in Effectiveness Trials) database of core outcome sets (COS), thus aiding knowledge discovery.

## Methods

A literature review was carried out to determine existing outcome classification systems, none of which were sufficiently comprehensive or granular for classification of all potential outcomes from clinical trials. A new taxonomy for outcome classification was developed, and as proof of principle, outcomes extracted from all published COS in the COMET database, selected Cochrane reviews and clinical trial registry entries were classified using this new system.

## Results

Application of this new taxonomy to COS in the COMET database revealed that whereas 274/299 (92%) COS include at least one physiological outcome, only 177 (59%) include at least one measure of impact (global quality of life or some measure of functioning) and only 105 (35%) made reference to adverse events.

## Conclusions

This outcome taxonomy will be used to annotate outcomes included in COS within the COMET database and is currently being piloted for use in Cochrane Reviews within the Cochrane Linked Data Project. Wider implementation of this standard taxonomy in trial and systematic reviews databases and registries will further promote efficient searching, reporting and classification of trial outcomes.

# Keywords

Randomised controlled trials, outcomes, effectiveness trials, PICO, taxonomy, COMET, Cochrane, core outcome sets, systematic reviews, classification, comparative effectiveness research

# What’s new?

**Key findings**

* Existing taxonomy structures are intended as general health research vocabularies, rather than focussing on outcomes; do not provide sufficiently granular or comprehensive classification of trial outcomes; or are disease-specific or focused on patient-centred outcomes only.
* The current lack of an outcome classification system, fit for purpose, is holding back research as a result of (i) inconsistency and ambiguity in how outcomes are described across different studies, and (ii) inefficiency in searching knowledge sources including the published literature and ongoing research repositories such as clinical trials registries, which to date include outcomes as free text entries only.
* A new workable outcome taxonomy is proposed, the robustness of which has been demonstrated through application to a large number of trial registry entries in [clinicaltrials.gov](http://clinicaltrials.gov), Cochrane Reviews and core outcome sets in the COMET database.

**What this adds to what was known?**

* Core outcome set developers should give more attention to measures of life impact and adverse events when determining core outcome sets for trials of the effectiveness of health and social care interventions.

**What is the implication and what should change now?**

* An accepted taxonomy of outcomes would increase the re-use value of outcome data, just as MeSH terms have transformed the searchability of medical literature. Wider implementation of this taxonomy will help to reduce waste in research by promoting efficient searching, reporting and classification of clinical outcomes for the first time, thereby speeding up research activities including discovery science and ‘big data’ approaches to extracting knowledge from published information.

# Background

Recognition that insufficient attention has been paid to the choice of outcomes to measure in clinical trials is increasing. In the context of clinical trials, an outcome is defined to be a measurement or observation used to capture and assess the effect of treatment such as assessment of side effects (risk) or effectiveness (benefits) (1). The COMET (Core Outcome Measures in Effectiveness Trials) Initiative (web reference 1) (2) brings together people interested in the development and application of agreed standardised sets of outcomes, known as ‘core outcome sets’ (COS). These sets represent the minimum that should be measured and reported in all clinical trials of a specific condition, and are also suitable for use in clinical audit or research other than randomised trials. One of the successes of COMET has been the development of a publicly available searchable database of completed and ongoing projects in COS development (2-7). This unique resource provides information on the COS developed to date, and is currently searchable by population, intervention and condition. However, as yet the records in the COMET database have not been categorised according to outcome, the fourth of the essential elements that should be defined for a trial, according to the PICO (population, intervention, comparison, outcomes) model.

Similarly, outcomes in trials registries (including the EU Clinical Trials Register, ClinicalTrials.gov and ISRCTN registry) can be entered as free text only, hampering the ability to search for outcomes effectively because of variation and inconsistencies in how outcomes are described across different trials. Over sixty percent of queries related to requests to register a trial relate to how outcomes were described (8). Standardised terminology to describe outcomes is starting to come into use in pre-clinical research, where variations in description have impeded computational analysis of phenotypic data (9). However, there is currently no consensus on how clinical trial outcomes should be classified. Standard terminology to describe outcomes in pre-clinical and clinical research would facilitate the comparison of outcomes between pre-clinical and clinical settings, potentially providing insight into the reasons why so many late phase trials “fail” despite promising results from pre-clinical studies.

A **taxonomy** is a scheme of classification that is often used for, for example, the naming of living organisms but which can also be used as a controlled vocabulary (i.e. an authoritative list of terms for use in indexing) with a hierarchical structure (web reference 2). Taxonomies exist for many aspects of health research, such as the ICF (International Classification of Functioning, Disability and Health, web reference 3) and ICD-10 (International Statistical Classification of Diseases and Related Health Problems 10th Revision, web reference 4). The Cochrane Linked Data Project (web reference 5) use a 15-item taxonomy for high-level categorisation of interventions (for the IC components of PICO).

A standard outcome taxonomy would help to improve knowledge discovery by facilitating organised searching of trials by outcome in trial registries and databases. For example, a researcher might be interested in identifying all interventions that have been tested in a randomised trial to improve a particular outcome. Similarly, COS have the potential to reduce research waste by avoiding unnecessary duplication of efforts and allowing the results of clinical trials to be combined (2, 7), but this benefit will only be realised with COS uptake. Classification of trial outcomes will facilitate efficient assessment of COS uptake, again improving knowledge (10).

We sought to identify and further develop as necessary a taxonomy providing sufficient granularity and scope for the classification of all outcomes in the COS listed in the COMET database, which would be equally suitable for classification of outcomes included in trial registries, trial reports and systematic reviews. This taxonomy is intended for the classification of what, rather than how, outcomes are measured.

# Methods

A suitable taxonomy for clinical trial outcomes must clearly differentiate between high level outcome types, while comprehensively covering all potential outcomes from clinical trials in a sensible hierarchical structure. We carried out a literature review to identify existing outcome taxonomies that would inform the development of the one presented here. We searched PubMed for published journal articles and internet resources such as Google. Our search involved a combination of terms, including “ontology”, “taxonomy”, “classification” or “categorisation” and “health”, “health research”, “trial”, and “outcomes”.

In order to examine outcome classification systems used as part of COS development, the COS studies within the COMET database that included a systematic or literature review to identify relevant studies were reviewed to determine how they categorised their outcomes.

The lack of an existing suitable outcome taxonomy for trial outcomes led to the subsequent development of a new taxonomy to classify trial and systematic review outcomes. This was an iterative process, starting with the 15-category scale developed by Smith *et al* (11) to classify outcomes recorded in Cochrane Reviews (Table 1). A refined version of this scale, developed by two authors (PW and MC), was piloted as part of PICO classification of reviews within the Cochrane database (web reference 5). This 12-category version was then further developed to provide more detail relating to physiological, function and resource use domains, leading to a taxonomy with 38 outcome domains within five core areas. Explanations and examples of outcomes within each of these domains are found in Supplementary Table 1.

Physiological outcomes are categorised according to the underlying cause or affected body system, grouped using the MedDRA System Organ Classes (SOCs) (Supplementary Table 2) with the exception of four SOCs (Investigations, Social circumstances, Surgical and medical procedures, Product issues) which are not considered relevant within the physiological/clinical domains. For example, “endocrine outcomes” are those associated with endocrine disorders. “Outcomes related to neoplasms” include those relating to physiological function, signs and symptoms caused by benign, malignant and unspecified (including cysts and polyps) neoplasms, including solid and non-solid tumours. Examples of such outcomes include “time to recurrence”, “response rate” and “clearance of resection margins”. “General outcomes” include those affecting the whole body which cannot be attributed to a certain body system, for example, fatigue, chills, flu like symptoms, malaise, anorexia, pain (unspecified, not associated with a particular body system), fever (not attributable to infection), anthropometric measures (e.g. weight), “global” measures, “symptoms” (not associated with a particular body system), “physical health” and fitness. Laboratory parameters (for example, from blood samples) and scientific measures (for example, pharmacokinetic outcomes) should be classified within the physiological domain that captures the reason for the assessment (rather than within the “blood and lymphatic system” category, for example).

The functioning categories were extended beyond those used by Smith *et al* (11) (Activities of Daily Living and Psychosocial) to differentiate more accurately between physical, social, role, emotional and cognitive functioning.

The “delivery of care” domain contains a number of variables related to health care interventions, including compliance, withdrawal and satisfaction. These were grouped as they are all related to the appropriateness and acceptability of the intervention and may not be easily distinguishable (for example because of overlap between issues relating to compliance, satisfaction with care, withdrawal, treatment failure). Examples of outcomes in this category include patient preference; withdrawal from intervention (e.g. time to treatment failure, reason for stopping therapy); appropriateness, accessibility, quality and adequacy of intervention; patient or carer satisfaction; and process, implementation and service outcomes.

The “adverse event” domain includes outcomes broadly labelled as some form of unintended consequence of the intervention (e.g. adverse events/effects, adverse reactions, safety, harm, negative effects, toxicity, complications, sequelae). Specifically named adverse events are classified within the appropriate taxonomy domain relating to the specific event type, with an additional level of categorisation which identifies this outcome as an adverse event.

The “mortality/survival” domain includes overall (all-cause) and cause-specific survival/mortality, as well as composite survival outcomes that include death (e.g. disease-free survival). Composite outcomes should be classified in all domains relating to each of the included event types; for example, disease-free survival would be classified within the “mortality/survival” domain as well as the physiological outcome domain relating to the particular disease.

The final 38-item scale was applied to the classification of trial outcomes recorded within the 299 published COS in the COMET database that were published before 2016 and to outcomes from 3515 Cochrane reviews as part of the pilot phase of the Cochrane Linked Data Project. To further illustrate its applicability, the taxonomy has been applied to outcomes listed in 30 studies identified from a search of the US National Institutes of Health clinical trials registry (www.clinicaltrials.gov). Furthermore, two case studies are presented to demonstrate how the taxonomy can provide standard classification of outcomes across different research settings linked to particular clinical areas. One of the authors (SD) assessed all of the outcomes in the COS database and NIH clinical trial registry. In cases of any doubt or ambiguity, a second opinion (PW) was sought. Cochrane review outcomes were classified by Cochrane reviewers.

|  |  |  |  |
| --- | --- | --- | --- |
| **Core area** | **Smith** | **Williamson/Clarke (initial)** | **Williamson/Clarke (revised)** |
| **Death** | 1: Mortality/survival | 1: Mortality/survival | 1: Mortality/survival |
| **Physiological or clinical** | 2: Physiological/ clinical | 2: Physiological/clinical | 2-24: **Physiological/clinical (see below)** |
| 3: Infection | 3: Infection |
| 4: Pain | 4: Pain |
| **Life impact** | 5: Activities of Daily Living | 5: Function | **Functioning**  25: Physical functioning  26: Social functioning  27: Role functioning  28: Emotional functioning/wellbeing  29: Cognitive functioning |
|  | - Physical |
| - Social |
| - Role |
| 6: Psychosocial | 6: Psychosocial |
| 7: Mental Health |
|  | 7: QoL | 8: HRQL | 30: Global quality of life |
|  |  |  | 31: Perceived health status |
|  | 8: Compliance | 9: Compliance (including withdrawal from treatment) | 32: Delivery of care, including  - Satisfaction / Patient preference  - Acceptability and availability  - Adherence/ Compliance  - Withdrawal from treatment  - Appropriateness of treatment  - Process, implementation and service outcomes |
|  | 9: Withdrawal from treatment/ study |
|  | 10: Satisfaction (patient, carer, health care provider) | 10: Satisfaction |
|  |  | 33: Personal circumstances |
| **Resource use** | 11: Medication | 11: Resource use  - Economic  - Hospital  - Operative  - Medication | **Resource use**  34: Economic  35: Hospital  36: Need for further intervention  37: Societal/carer burden |
|  | 12: Economic |
|  | 13: Hospital |
|  | 14: Operative |
|  |  |
| **Adverse events** | 15: Adverse events/effects | 12: Adverse events/effects | 38: Adverse events/effects |

Table 1 Development of 38-category scale

**Physiological/clinical domains**

1. Blood and lymphatic system outcomes
2. Cardiac outcomes
3. Congenital, familial and genetic outcomes
4. Endocrine outcomes
5. Ear and labyrinth outcomes
6. Eye outcomes
7. Gastrointestinal outcomes
8. General outcomes
9. Hepatobiliary outcomes
10. Immune system outcomes
11. Infection and infestation outcomes
12. Injury and poisoning outcomes
13. Metabolism and nutrition outcomes
14. Musculoskeletal and connective tissue outcomes
15. Outcomes relating to neoplasms: benign, malignant and unspecified (including cysts and polyps)
16. Nervous system outcomes
17. Pregnancy, puerperium and perinatal outcomes
18. Renal and urinary outcomes
19. Reproductive system and breast outcomes
20. Psychiatric outcomes
21. Respiratory, thoracic and mediastinal outcomes
22. Skin and subcutaneous tissue outcomes
23. Vascular outcomes

# Results

## 3.1 Literature review

A review of the literature identified several vocabularies (such as MedDRA, web reference 6, and SNOMED CT, web reference 7) which exist to organise and classify text relating to health research, many of which are included within the Unified Medical Language System (UMLS, web reference 8). However, few relate specifically to outcome classification. For example, the International Classification of Functioning, Disability and Health (ICF) provides a conceptual framework for understanding and describing health and disability, accounting for both patient and contextual factors, rather than an explicit classification of trial outcomes (web reference 3). Medical Subject Headings (MeSH, the National Library of Medicine's controlled vocabulary thesaurus, web reference 9) categories extend beyond health outcomes, covering not only anatomy, diseases and health care, but also technology, occupations, information science, geographicals etc. Subclasses within their “Diseases” category are similar to our physiological/clinical domains but with additional levels of differentiation, for example, between bacterial/virus/parasitic diseases and occupational diseases/disorders of environmental origin/chemically-induced disorders.

Of those vocabularies that can be applied to outcomes, none are suitable for the classification of all potential outcomes from clinical trials, as they provide only a partial perspective or are relevant only for specific diagnoses or fields of research. For example, physiological domains alone are categorised in some of these vocabularies, such as ICD-10 (web reference 4), NICE (web reference 10), UK Clinical Research Collaboration (UKCRC) Health Research Classification System (HRCS) Health Categories (web reference 11), PCORI (Patient-Centered Outcomes Research Institute, web reference 12), as well as in other ontologies relating to genetic research (for example the Human Phenotypic Ontology,web reference 13, see Supplementary Table 2). The Diagnostic and Statistical Manual of Mental Disorders (DSM, web reference 14) provides a comprehensive classification system of mental disorders only. The Grid-Enabled Measures (GEM) Database (web reference 15), an online tool for the organisation of scientific measures used in behavioural, social science and other scientific research areas, classifies measures according to physiological and methodological research areas. However, these categories do not cover the full range of potential trial outcomes, and thus fail to provide a comprehensive structure for outcome classification.

The Agency for Healthcare Research and Quality (AHRQ) commissioned a project to determine existing methods to standardise outcome measure definitions, in order to inform the development of its Outcome Measures Framework (web reference 16). Their literature review identified few existing methods to categorise outcome measures, none of which are entirely relevant for our purposes. For example, the National Quality Form (NQF) Quality Positioning System (QPS) provides a search facility for quality measures, with search categories that extend beyond the remit of classifying trial outcomes (see Supplementary Table 2). The OMF provides a means to describe the context relating to outcome measures within patient registry entries rather than providing a comprehensive taxonomy structure for outcome classification. This system provides a conceptual framework to categorise data elements according to the characteristics of the study (in particular, of the participant, disease and treatment provider) and treatment (type and intent), as well as of the outcome (categorised as survival, disease response, events of interest, patient reported outcomes and health system utilisation). These five main outcome classification categories do not provide a comprehensive system for classifying all trial outcomes.

The literature review identified several outcome classification systems; however, none of these provide a hierarchical structure of sufficient scope or granularity to be usefully applied to all potential trial outcomes:

1. Wilson and Cleary (12) developed a health-related quality of life (HRQL) conceptual model rather than providing a detailed outcome taxonomy structure, and excluded outcomes such as resource use or adverse events.
2. PROMIS (Patient-Reported Outcomes Measurement Information System, web reference 17) provides a structure for classifying patient-reported measures only; outcomes collected by health care providers, and those affecting wider society, are therefore not included.
3. Similarly, the Nursing Outcomes Classification (NOC, web reference 18) only covers outcomes relevant to nursing, thus excluding outcome domains with wider relevance, such as resource use and adverse events.
4. Various disease-specific classification structures provide outcome taxonomies relevant to a specific disease or condition only (NIH Toolbox, DOMS, Neuro-QoL, ASCQ-Me, web reference 19).
5. Outcome Measures in Rheumatology (OMERACT) provides a useful structure of outcome “core areas”, with examples of domains to be included within each of these “core areas”; however, this structure is not sufficiently detailed to provide standardised classification of outcomes beyond the top “core area” level (13).
6. Davey *et al* (14) used a data-driven approach to categorise outcomes from Cochrane Reviews into 11 categories. The disadvantage of a data-driven approach is that it potentially will not be fully comprehensive, as it may not extend beyond the collected outcomes to cover all possible trial outcomes. This structure provided a useful starting point for classification of trial outcomes but lacked a hierarchical structure; some categories are overly broad while others are too specific for classification purposes. Similarly, Smith *et al* (11)grouped outcomes from Cochrane Reviews into 15 categories; however, this classification system also failed to systematically differentiate between higher level outcome types with a structured hierarchy.

## 3.2 Outcome classification systems used in COS studies

One third (99/299) of published COS studies involved a systematic or literature review to identify relevant outcomes. Of these, 21 applied their own data-driven approach to outcome classification. Six applied an existing classification system: four studies used ICF terms, one study used a simplified version of the Wilson and Cleary model and one study used outcome categories defined by previous authors (specifically for stroke outcomes) (15).

## 3.3 Categorisation of COS outcomes

The newly proposed 38-item classification system was applied to the 299 published COS in the COMET database, where the median (range) number of outcomes per COS is 5 (1, 46). Table 2 displays the number of COS that include at least one outcome from each of the categories. Whereas 92% (274 COS) include at least one physiological outcome, only 59% (177 COS) include at least one measure of impact (HRQL or some measure of functioning). Only one third (105, 35%) of COS explicitly call for adverse events/effects to be recorded. At least one resource use outcome was included in only 84 (28%) of COS. As expected, the breakdown according to physiological/clinical domains largely reflects the profile of diseases and conditions for which COS have been developed (5-7).

|  |  |  |
| --- | --- | --- |
| **Outcome area** | **Outcome domain** | **Number of COS (% of 299)** |
| **Mortality/survival** | **Mortality/survival** | **99 (33)** |
| **Physiological/clinical** | **Physiological/clinical (≥1)** | **274 (92)** |
|  |  |  |
|  | Blood and lymphatic system outcomes | 9 (3) |
|  | Cardiac outcomes | 24 (8) |
|  | Congenital, familial and genetic outcomes | 1 (0.3) |
|  | Endocrine outcomes | 3 (1) |
|  | Ear and labyrinth outcomes | 3 (1) |
|  | Eye outcomes | 6 (2) |
|  | Gastrointestinal outcomes | 43 (14) |
|  | General outcomes | 57 (19) |
|  | Hepatobiliary outcomes | 6 (2) |
|  | Immune system outcomes | 6 (1) |
|  | Infection and infestation outcomes | 18 (6) |
|  | Injury and poisoning outcomes | 7 (2) |
|  | Metabolism and nutrition outcomes | 1 (0.3) |
|  | Musculoskeletal and connective tissue outcomes | 58 (19) |
|  | Outcomes relating to neoplasms: benign, malignant and unspecified (including cysts and polyps) | 33 (11) |
|  | Nervous system outcomes | 48 (17) |
|  | Pregnancy, puerperium and perinatal outcomes | 8 (3) |
|  | Renal and urinary outcomes | 13 (4) |
|  | Reproductive system and breast outcomes | 8 (3) |
|  | Psychiatric outcomes | 23 (8) |
|  | Respiratory, thoracic and mediastinal outcomes | 32 (11) |
|  | Skin and subcutaneous tissue outcomes | 12 (4) |
|  | Vascular outcomes | 31 (10) |
|  |  |  |
| **Life impact** | **Functioning (≥1)** | **128 (43)** |
|  | Physical | 111 (37) |
|  | Social | 25 (8) |
|  | Role | 11 (4) |
|  | Emotional/wellbeing | 29 (10) |
|  | Cognitive | 21 (7) |
|  | **Global quality of life** | **121 (40)** |
|  | **Perceived health status** | **0 (0)** |
|  | **Delivery of care** | **52 (17)** |
|  | **Personal circumstances** | **0 (0)** |
| **Resource use** | **Resource use (≥1)** | **84 (28)** |
|  | Economic | 37 (12) |
|  | Hospital | 24 (8) |
|  | Need for further intervention | 44 (15) |
|  | Societal/carer burden | 5 (2) |
| **Adverse events/effects** | **Adverse events/effects** | **105 (35)** |

Table 2 Breakdown of outcomes within 299 COS in COMET database

## 3.4 Categorisation of systematic review outcomes

A total of 16525 outcomes from 3515 Cochrane Reviews have been classified according to our taxonomy to date as part of the pilot phase of the Cochrane Linked Data Project (Table 3). The majority of the annotated reviews came from the Cochrane Pregnancy and Childbirth and the Neonatal groups; a smaller set came from the Cochrane Developmental, Psychosocial and Learning Problems group. In these selected Cochrane reviews, outcomes were less commonly reported within each of the overarching outcome areas than for COS, with the exception of resource use. Less than one quarter (831, 24%) of reviews include a measure of impact (function or quality of life, QoL) while physiological outcomes dominate, being present in 83% (2915) of reviews annotated to date.

|  |  |  |
| --- | --- | --- |
|  | **Number (%) of 3515 Cochrane reviews** | **Number (%) of 16525 outcome classifications** |
| Adverse events | 596 (17) | 951 (6) |
| Mortality | 857 (24) | 1246 (8) |
| Physiological | 2915 (83) | 9820 (59) |
| Function/QoL | 831 (24) | 1844 (11) |
| Delivery of care | 419 (12) | 493 (3) |
| Resource use | 1117 (32) | 2171 (13) |

Table 3 Cochrane Linked Data Project pilot phase outcome classifications

## 3.5 Categorisation of trial outcomes

The outcomes listed in 30 studies identified from a search for randomised, phase 3 and 4 interventional studies currently recruiting participants and received by the US National Institutes of Health clinical trials registry, clinicaltrials.gov, during the first 20 days of 2017 (https://clinicaltrials.gov search terms “Randomized”, “Phase 3,4”, “Recruiting”, “Interventional Studies”, “Received from 01/01/2017 to 01/20/2017”) have been categorised in Supplementary Table 3, demonstrating the general applicability of our ontology to trials in a trials registry.

## 3.6 Case studies

### Eczema

The taxonomy has been applied to the COS for eczema (16). In addition, the outcomes listed in 8 eczema studies identified from a search of clinicaltrials.gov (search terms “Randomized”, “Phase 3,4”, “Recruiting”, “Interventional Studies”, "Eczema”, no date restrictions) have been categorised in Supplementary Table 4, demonstrating the general applicability of our ontology to eczema trials.

### Rheumatoid arthritis

The taxonomy has been applied to the rheumatoid arthritis (RA) COS (17). The outcomes listed in 10 RA studies identified from a search of clinicaltrials.gov (search terms “Randomized”, “Phase 3,4”, “Recruiting”, “Interventional Studies”, "Rheumatoid Arthritis", “Received from 01/01/2017 to 01/20/2017”) have been categorised in Supplementary Table 5, again demonstrating the general applicability of our ontology to a particular clinical area.

# Discussion

A literature review identified several health research vocabularies which extend beyond the remit of outcome classification, as well as a number of outcome classification systems. However, none ofthese are sufficiently comprehensive or granular for the specific purpose of classifying all potential outcomes from clinical trials with structured hierarchical differentiation between high level outcome types. We have therefore described the development of a new taxonomy that can be used for the classification of outcomes included in all trials, COS, systematic reviews and trial registries. This classification system is based on similar top level “core areas” common to other outcome hierarchies (12, 13) but provides a more detailed taxonomy appropriate for all potential outcomes, in particular relating to physiological, functioning and resource use domains.

Health related quality of life (HRQL) measurement tools typically cover multiple domains (such as functioning, resource use, general physiological health and global quality of life) and should therefore be classified within each of these domains, even when overall summary measures are reported, as we would recommend for any composite outcome. For example, see Supplementary Table 6 for the mapping between our taxonomy and the facets included in the WHOQOL-100 tool (web reference 20).

The “global quality of life” domain in our taxonomy is reserved for specific individual questions or tools which measure the implicit composite outcome of global QoL (for example, “How would you rate your overall quality of life?”), rather than for overall summary measures from HRQL tools covering multiple domains. Comparison of our taxonomy with the individual HRQL measures listed in Macefield (18) demonstrate that the vast majority of questions or components included in HRQL tools should be classified in domains other than global QoL. To further promote this transparency relating to the content of HRQL measures, we support the advice given by Macefield (18) that HRQL tools should be split into their individual components. For example, the Diabetes Therapy Related Quality of Life Questionnaire can be split into various factors assessing burden on social activities and daily activities; anxiety and dissatisfaction; hypoglycaemia; and treatment satisfaction (19).

The PROMIS website (web reference 17) groups its adult and paediatric measures into profile domains in three core areas (physical, mental and social health), along with specific domains which act as search terms to identify relevant measures. These patient-reported outcome domains can all be categorised within various physiological and functioning domains in our taxonomy, demonstrating the applicability of our taxonomy to another commonly used trial outcome resource.

Any specifically named adverse events (for example, fatigue or pain) should be categorised under the appropriate taxonomy domain, rather than within the adverse event domain. In such cases, we would add an additional level of categorisation which specifies that this outcome was reported as an adverse event. Thus we suggest that our outcome classification system should be implemented as a two-component taxonomy, the first defining the outcome structure (as we have specified in the 38-item scale) and the second specifying whether or not the outcome is being measured as a benefit or a harm outcome. For example, the COS for colorectal cancer surgery (20) includes faecal urgency, which is a potential adverse effect of the surgery. In our system, this would be classified as a physiological outcome, under the gastrointestinal category, but a second component would identify it as an adverse outcome. In a particular example of the detailed classification of adverse events relating to total ankle arthroplasty (21), the adverse events listed can be classified within existing physiological categories, predominantly musculoskeletal and connective tissue, and infection domains.

In contrast, the adverse event domain only includes outcomes explicitly labelled as some form of unintended consequence of the intervention, such as “adverse events”, “adverse effects”, “adverse reactions”, “complications”, “toxicity” or “sequelae”. This domain, which is not intended to include any specifically named adverse events, is important as it indicates whether or not trialists or researchers considered the need to record events that may not necessarily be prespecified ahead of time. Unless the adverse event profile is very well established for a given intervention, it is important that the incidence of all adverse events, expected or otherwise, is reported. Similarly, COS or systematic reviews that cover multiple intervention types should address the potential for unspecified adverse events.

The economic outcome domains in our taxonomy map well to those identified by Thorn et al as key health economic items to be collected as part of clinical trials (22). The 10 items in their final core set are classified under different types of care, all of which can be classified within our economic outcome domains: “hospital care” or “emergency care” fit within our “hospital” domain; “care at a GP surgery, health clinic or other community setting” and “health care at home” belong to our “societal/carer burden domain”; and “medication” fits within our “need for further intervention” domain.

We are confident that our taxonomy provides a sufficiently comprehensive basis for the categorisation of outcomes included in clinical trials in general. However, we would welcome feedback from researchers applying the taxonomy in their clinical settings in order to demonstrate further validation of the taxonomy or to highlight any necessary changes. Note that we are not suggesting that trials or reviews should necessarily include outcomes from each of the core areas in this taxonomy. Note also that this taxonomy relates to outcomes measured at an individual-patient level (including those relating to the direct impact of the individual patient’s treatment or condition on wider society, for example resource use or carer burden) but is not intended to cover outcomes relating to the health or functioning of wider society (for example, family or community health). Therefore, health promotion or public health outcomes from trials of family- or community-based interventions can be classified using our taxonomy if they relate to an individual’s condition or care, but not if they are measured at the family or community level.

Outcome categories within our taxonomy may be classified in even greater detail in relation to particular interventions (for example, the classification of outcomes for childhood vaccination communication interventions (23)). Indeed, we would encourage further subdivision of each outcome domain by researchers specialising in relevant clinical or methodological areas. There may be existing taxonomies that could be used to provide finer classification within our high-level taxonomy domains; for example, the DSM could be used to classify mental disorders within the psychiatry domain.

Adoption of this classification system will facilitate literature searches; for example, if clinical trial outcomes were routinely classified according to this taxonomy, researchers would easily be able to identify clinical trials that included outcomes domains from a particular COS. A readily available taxonomy will also assist COS developers who need to categorise outcomes, for example as part of their Delphi survey, thereby speeding up the development of COS and expediting their completion and availability for use by trialists and other researchers. Application of this classification system to COS contained within the COMET database has highlighted key points to note, including that, although the COMET database relates to COS recommended for effectiveness trials, far fewer of the COS contain measures of impact (58%) than physiological outcomes (92%). Furthermore, only one third of COS reports highlight the need to record unintended adverse consequences, and even fewer COS (29%) include any economic outcomes.

The lack of a standard taxonomy relating to trial outcomes impedes the ability to efficiently and effectively search the literature. An accepted taxonomy of outcomes would increase the re-use value of outcome data, just as MeSH terms have transformed the searchability of medical literature. The taxonomy would firstly help drive to push for consistency of clinical outcome terms between clinical trials, which has been a major focus of the COMET initiative. More importantly it will allow efficient searching, reporting and classification of clinical outcomes for the first time, thereby speeding up research activities including discovery science and ‘big data’ approaches to extracting knowledge from published information.

In summary, the applicability of this new taxonomy has been demonstrated for the categorisation of outcomes from core outcome sets, systematic reviews and trials recorded within a clinical trial registry. Similarly, two case studies demonstrate the relevance of standardising outcome classification to link COS, Cochrane Reviews, and trial registry entries within particular clinical areas. This taxonomy has been designed with the purpose of providing high-level differentiation between outcome domains in order to facilitate uniformity of outcome classification in electronic databases. We would welcome further testing of this taxonomy, and further development of subcategories to provide finer classification within each of the outcome domains is encouraged. Ongoing COS studies have used this taxonomy to classify outcomes for their initial list for a Delphi (web references 21-23). We will monitor use of the taxonomy and collate feedback, to be subsequently reported.

| **Core area** | **Outcome domain** | **Explanation** |
| --- | --- | --- |
| **Death** | 1. Mortality/survival | Includes overall (all-cause) survival/mortality and cause-specific survival/mortality, as well as composite survival outcomes that include death (e.g. disease-free survival, progression-free survival, amputation-free survival) |
|  |  |
| **Physiological/ clinical** | **Physiological/clinical**   1. Blood and lymphatic system outcomes 2. Cardiac outcomes 3. Congenital, familial and genetic outcomes 4. Endocrine outcomes 5. Ear and labyrinth outcomes 6. Eye outcomes 7. Gastrointestinal outcomes 8. General outcomes 9. Hepatobiliary outcomes 10. Immune system outcomes 11. Infection and infestation outcomes 12. Injury and poisoning outcomes 13. Metabolism and nutrition outcomes 14. Musculoskeletal and connective tissue outcomes 15. Outcomes relating to neoplasms: benign, malignant and unspecified (including cysts and polyps) 16. Nervous system outcomes 17. Pregnancy, puerperium and perinatal outcomes 18. Renal and urinary outcomes 19. Reproductive system and breast outcomes 20. Psychiatric outcomes 21. Respiratory, thoracic and mediastinal outcomes 22. Skin and subcutaneous tissue outcomes 23. Vascular outcomes | Physiological/clinical outcomes include measures of physiological function, signs and and symptoms, as well as laboratory (and other scientific) measures relating to physiology, and are categorised according to the underlying cause/body system.  “General disorders” includes those affecting the whole body and cannot be attributed to a certain body system e.g. fatigue, chills, flu like symptoms, malaise, anorexia, pain (unspecified, not associated with a particular body system), fever (not attributable to infection), anthropometric measures (e.g. weight), “global” measures, “symptoms” (not associated with a particular body system), “physical health”, fitness.  Pain outcomes are categorised according to underlying cause or body system or within the “General symptoms” category (if non-specific).  Laboratory parameters (for example, from blood samples) and scientific measures (for example, pharmacokinetic outcomes) should be classified within the physiological domain that captures the reason for the assessment (rather than within the “blood and lymphatic system” category, for example).  Psychiatric outcomes include all those relating to mental health conditions and associated behaviours (e.g. addictions and behavioural problems).  Pregnancy, puerperium and perinatal domain extends to outcomes relating to breastfeeding and weaning.  Outcomes relating to neoplasms include those related to non-solid and solid tumours. |
| **Life impact** | **Functioning**   1. Physical functioning 2. Social functioning 3. Role functioning 4. Emotional functioning/wellbeing 5. Cognitive functioning | **Impact outcomes**  Physical functioning: impact of disease/condition on physical activities of daily living (for example, ability to walk, independence, self-care, performance status, disability index, measures relating to sleep, motor skills, sexual dysfunction. health behaviour and management)  Social functioning: impact of disease/condition on social functioning (e.g. ability to socialise, behaviour within society, communication, companionship, psychosocial development, aggression, recidivism, participation)  Role functioning: impact of disease/condition on role (e.g. ability to care for children, work status)  Emotional functioning/wellbeing: impact of disease/condition on emotions or overall wellbeing (e.g. ability to cope, worry, frustration, confidence, perceptions regarding body image and appearance, psychological status, stigma, life satisfaction, meaning and purpose, positive affect, self-esteem, self-perception and self-efficacy)  Cognitive functioning: impact of disease/condition on cognitive function (e.g. memory lapse, lack of concentration, attention); outcomes relating to knowledge, attitudes and beliefs (e.g. learning and applying knowledge, spiritual beliefs, health beliefs/knowledge) |
| 1. Global quality of life | Includes only implicit composite outcomes measuring global quality of life |
| 1. Perceived health status | Subjective ratings by the affected individual of their relative level of health |
| 1. Delivery of care | Includes outcomes relating to the delivery of care, including   * adherence/compliance * patient preference * tolerability/acceptability of intervention * withdrawal from intervention (e.g. time to treatment failure, reason for stopping therapy) * appropriateness of intervention * accessibility, quality and adequacy of intervention * patient/carer satisfaction (emotional rather than financial burden) * process, implementation and service outcomes (e.g. overall health system performance and the impact of service provision on the users of services) |
|  | 1. Personal circumstances | Outcomes relating to patient’s finances, home and environment |
| **Resource use** | **Resource use**   1. Economic 2. Hospital 3. Need for further intervention 4. Societal/carer burden | Economic: general outcomes (e.g. cost, resource use) not captured within other specific resource use domains  Hospital: outcomes relating to inpatient or day case hospital care (e.g. duration of hospital stay, admission to ICU)  Need for further intervention: outcomes relating to medication (e.g. concomitant medications, pain relief), surgery (e.g. caesarean delivery, time to transplantation) and other procedures (e.g. dialysis-free survival, mode of delivery)  Societal/carer burden: outcomes relating to financial or time implications on carer or society as a whole (e.g. need for home help, entry to institutional care, effect on family income) |
|  |
|  |
|  |
|  |
|  |
|  |
| **Adverse events** | 1. Adverse events/effects | Includes outcomes broadly labelled as some form of unintended consequence of the intervention (e.g. adverse events/effects, adverse reactions, safety, harm, negative effects, toxicity, complications, sequelae). Specifically named adverse events should be classified within the appropriate taxonomy domain above with an additional level of categorisation which identifies that this outcome is being considered as an adverse event. |

Supplementary Table 1 Explanation and examples of outcomes within each outcome domain

|  |  |  |  |
| --- | --- | --- | --- |
| **MedDRA** | **NICE** | **NQF: QPS1** | **HPO “Phenotypic abnormality” superclass** |
| Blood and lymphatic system disorders  Immune system disorders | Blood and immune system conditions |  | Abnormality of the blood and blood-forming tissues  Abnormality of the immune system |
| Neoplasms benign, malignant and unspecified (including cysts and polyps) | Cancer | Cancer | Neoplasm |
| Cardiac disorders  Vascular disorders | Cardiovascular conditions | Cardiovascular | Abnormality of the cardiovascular system |
| Endocrine disorders  Metabolism and nutrition disorders | Diabetes and other endocrinal, nutritional and metabolic conditions | Endocrine | Abnormality of the endocrine system  Abnormality of metabolism/homeostasis  Growth abnormality |
| Gastrointestinal disorders | Digestive tract conditions  Oral and dental health | Gastrointestinal  Dental | Abnormality of the abdomen |
| Ear and labyrinth disorders | Ear, nose and throat conditions | Ears, Nose, Throat (ENT) | Abnormality of the ear  Abnormality of the voice  Abnormality of the head or neck (partial mapping) |
| Eye disorders | Eye conditions | Eye Care | Abnormality of the eye |
| Pregnancy, puerperium and perinatal conditions | Fertility, pregnancy and childbirth | Perinatal Health | Abnormality of prenatal development or birth |
| Congenital, familial and genetic disorders | Genetic conditions |  |  |
| Reproductive system and breast disorders | Gynaecological conditions  Urological conditions | Gynaecology  Reproductive Health | Abnormality of the breast |
| Infections and infestations | Infections | Infectious Diseases |  |
| Injury, poisoning and procedural complications | Injuries, accidents and wounds |  |  |
| Renal and urinary disorders | Kidney conditions  Urological conditions | Renal | Abnormality of the genitourinary system |
| Hepatobiliary disorders | Liver conditions | Liver |  |
| Psychiatric disorders | Mental health and behavioural conditions | Behavioural Health |  |
| Musculoskeletal and connective tissue disorders | Musculoskeletal conditions | Musculoskeletal | Abnormality of connective tissue  Abnormality of the musculature  Abnormality of the skeletal system  Abnormality of limbs |
| Nervous system disorders | Neurological conditions | Neurology | Abnormality of the nervous system |
| Respiratory, thoracic and mediastinal disorders | Respiratory conditions | Respiratory | Abnormality of the thoracic cavity  Abnormality of the respiratory system |
| Skin and subcutaneous tissue disorders | Skin conditions |  | Abnormality of the integument |
| General disorders and administration site conditions |  |  |  |
| 2 Investigations, Social circumstances, Surgical and medical procedures, Product issues | 2 Chronic fatigue syndrome, Multiple long-term conditions | 2 Critical Care, Palliative Care and End-of-Life Care, Surgery |  |
|  | | | |
| 1 Search categories for “Clinical Condition/Topic Area” only  2 Extra categories (not directly mapped to other classification systems) | | | |

Supplementary Table 2 Mapping between MedDRA, NICE, NQF QPS and HPO categorisation of physiological conditions

|  |  |  |
| --- | --- | --- |
| **WHOQOL-100 domain** | **Facet within WHOQOL-100 domain** | **Outcome taxonomy domain** |
|  | Overall QoL and General Health | Global quality of life |
| Physical health |  |  |
|  | Energy and fatigue | Physical functioning |
|  | Pain and discomfort | Physical functioning |
|  | Sleep and rest | Physical functioning |
| Psychological |  |  |
|  | Bodily image and appearance | Emotional functioning (emotions) |
|  | Negative feelings | Emotional functioning |
|  | Positive feelings | Emotional functioning |
|  | Self-esteem | Emotional functioning |
|  | Thinking, learning, memory and concentration | Cognitive functioning |
| Level of independence |  |  |
|  | Mobility | Physical functioning |
|  | Activities of daily living | Physical functioning |
|  | Dependence on medicinal substances and medical aids | Need for further intervention |
|  | Work capacity | Role functioning |
| Social relationships |  |  |
|  | Personal relationships | Social functioning |
|  | Social support | Societal/carer burden |
|  | Sexual activity | Social functioning |
| Environment |  |  |
|  | Financial resources | Personal circumstances |
|  | Freedom, physical safety and security | Emotional functioning (feelings); Personal circumstances (environment) |
|  | Health and social care; accessibility and quality | Delivery of care |
|  | Opportunities for acquiring new information and skills | Personal circumstances |
|  | Participation in and opportunities for recreation/leisure activities | Social functioning (participation); Personal circumstances (opportunities) |
|  | Physical environment (pollution/noise/traffic/climate) | Personal circumstances |
|  | Transport | Personal circumstances |
| Spirituality/Religion/Personal beliefs |  |  |
|  | Spirituality/Religion/Personal beliefs | Cognitive functioning |

Supplementary Table 6 Mapping between WHOQOL-100 and outcome taxonomy domains

# References

1. 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.

2. Gargon E. The COMET Initiative database: progress and activities from 2011 to 2013. Trials. 2014;15.

3. Gargon E. The COMET Initiative database: progress and activities update (2014). Trials. 2015;16.

4. Gargon E, Williamson PR, Altman DG, Blazeby JM, Tunis S, Clarke M. The COMET Initiative database: progress and activities update (2015). Trials. 2017;18(1):54.

5. Gargon E. Choosing important health outcomes for comparative effectiveness research: a systematic review. PLoS One. 2014;9.

6. Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey. PLOS ONE. 2016;11(1):e0146444.

7. Gorst SL, Gargon E, Clarke M, Smith V, Williamson PR. Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and Identification of Gaps. PLOS ONE. 2016;11(12):e0168403.

8. Cuff A. Improving Outcome Measures in the ISRCTN Clinical Trial Registry. 2016.

9. Robinson PN, Kohler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. American journal of human genetics. 2008;83(5):610-5.

10. 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.

11. Smith V, Clarke M, Williamson P, Gargon E. Survey of new 2007 and 2011 Cochrane reviews found 37 % of prespecified outcomes not reported. J Clin Epidemiol. 2015;68.

12. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. Jama. 1995;273(1):59-65.

13. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol. 2014;67(7):745-53.

14. Davey J, Turner RM, Clarke MJ, Higgins JP. Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis. BMC Medical Research Methodology. 2011;11(1):160.

15. Roberts L, Counsell C. Assessment of clinical outcomes in acute stroke trials. Stroke; a journal of cerebral circulation. 1998;29(5):986-91.

16. Schmitt J, Langan S, Williams HC. What are the best outcome measurements for atopic eczema? A systematic review. Journal of Allergy and Clinical Immunology. 2007;120(6):1389-98.

17. Boers M, Tugwell P, Felson DT, van Riel PL, Kirwan JR, Edmonds JP, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. The Journal of rheumatology Supplement. 1994;41:86-9.

18. Macefield RC, Jacobs M, Korfage IJ, Nicklin J, Whistance RN, Brookes ST, et al. Developing core outcomes sets: methods for identifying and including patient-reported outcomes (PROs). Trials. 2014;15(1):1-12.

19. Ishii H. Development and psychometric validation of the Diabetes Therapy-Related QOL (DTR-QOL) questionnaire. Journal of medical economics. 2012;15(3):556-63.

20. McNair AGK, Whistance RN, Forsythe RO, Macefield R, Rees J, Pullyblank AM, et al. Core Outcomes for Colorectal Cancer Surgery: A Consensus Study. PLOS Medicine. 2016;13(8):e1002071.

21. Mercer J, Penner M, Wing K, Younger ASE. Inconsistency in the Reporting of Adverse Events in Total Ankle Arthroplasty. Foot & Ankle International. 2016;37(2):127-36.

22. Thorn JC, Brookes ST, Ridyard C, Riley R, Hughes DA, Wordsworth S, et al. Core Items for a Standardized Resource Use Measure (ISRUM): Expert Delphi Consensus Survey. Value in Health. 2017.

23. Kaufman J, Ryan R, Glenton C, Lewin S, Bosch-Capblanch X, Cartier Y, et al. Childhood vaccination communication outcomes unpacked and organized in a taxonomy to facilitate core outcome establishment. J Clin Epidemiol. 2017;84:173-84.

# Web References

All web references were last accessed on October 12, 2017.

1. <http://www.comet-initiative.org>
2. <http://www.taxonomies-sig.org/about.htm>
3. <http://www.who.int/classifications/icf/en/>
4. <http://apps.who.int/classifications/icd10/browse/2016/en>
5. <http://linkeddata.cochrane.org/linked-data-project/metadata-and-vocabularies>
6. <http://www.meddra.org/>
7. <http://www.ihtsdo.org/>
8. <https://www.ncbi.nlm.nih.gov/books/NBK9675/>
9. <https://www.nlm.nih.gov/mesh>
10. <https://www.nice.org.uk/guidance/conditions-and-diseases>
11. <http://www.hrcsonline.net/hrcs_overview>
12. <http://www.pcori.org/research-results-home>
13. <http://human-phenotype-ontology.github.io/>
14. <https://www.psychiatry.org/psychiatrists/practice/dsm>
15. <https://www.gem-beta.org/Public/Home.aspx>
16. <https://www.effectivehealthcare.ahrq.gov/ehc/products/311/2310/registry-of-patient-registries-OMF-report-160912.pdf>
17. <http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis>
18. <https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NOC/>
19. <http://www.healthmeasures.net/explore-measurement-systems/overview>
20. <http://depts.washington.edu/seaqol/docs/WHOQOL_Info.pdf>
21. <http://www.comet-initiative.org/studies/details/764?result=true>
22. <http://www.comet-initiative.org/studies/details/956?result=true>
23. <http://www.comet-initiative.org/studies/details/1030?result=true>

# Competing interests

SD and PW are funded by MRC North West Hub for Trials Methodology Research MR/K025635/1. PW and MC are members of the COMET Management Group.

# Authors’ contribution

PW and MC jointly conceived the study, SD performed the analysis and wrote the paper, and LB and CM provided Cochrane Review data. SD, RF, MC and PW contributed to development of the taxonomy, and all authors commented on the manuscript and approved the final version.