REVIEW

Mobility endpoints in marketing authorisation of drugs: what gets the European medicines agency moving?

SIMON U. JAEGER[1,2,3], MARTIN WOHLRAB[1,2], DANIEL SCHOENE[4], ROMAN TREMMEL[1,2], MICHAEL CHAMBERS[5], LETIZIA LEOCANI[6], SOLANGE CORRIOL-ROHOU[7], JOCHEN KLENK[8,9], BASIL SHARRACK[10], JUDITH GARCIA-AYMERICH[11,12,13], LYNN ROCHESTER[14], WALTER MAETZLER[15], MILO PUHAN[16], MATTHIAS SCHWAB[1,2,3,17,†], CLEMENS BECKER[8,†]

[1]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany
[2]University of Tuebingen, Tuebingen, Germany
[3]Department of Clinical Pharmacology, University Hospital, University of Tuebingen, Tuebingen, Germany
[4]Institute of Medical Physics, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany and Institute of Exercise and Public Health, University of Leipzig, Leipzig, Germany
[5]MC Healthcare Evaluation, London, UK
[6]Experimental Neurophysiology Unit, Institute of Experimental Neurology (INSPE), Scientific Institute Hospital San Raffaele, University Vita-Salute San Raffaele, Milan, Italy
[7]AstraZeneca, GMD, Global Regulatory Excellence, Paris, France
[8]Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany
[9]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany
[10]Department of Neuroscience and NIHR Neurosciences Biomedical Research, Sheffield Teaching Hospitals NHS Foundation Trust, University of Sheffield, Sheffield, UK
[11]ISGlobal, Barcelona, Spain
[12]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[13]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[14]Translational and Clinical Research Institute, Newcastle University, Newcastle, UK
[15]Department of Neurology, University Hospital Schleswig-Holstein, Kiel, Germany
[16]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland
[17]Depts. of Clinical Pharmacology, and of Biochemistry and Pharmacy, University of Tuebingen, Tuebingen, Germany

Address correspondence to: Simon Jaeger, Department of Clinical Pharmacology, University Hospital Tuebingen University of Tuebingen and Dr Margarete Fischer-Bosch-Institute of Clinical Pharmacology. Email: simon.jaeger@ikp-stuttgart.de
[†]These authors have contributed equally to this work.

Abstract

Background: Mobility is defined as the ability to independently move around the environment and is a key contributor to quality of life, especially in older age. The aim of this study was to evaluate the use of mobility as a decisive outcome for the marketing authorisation of drugs by the European Medicines Agency (EMA).
Methods: Fifteen therapeutic areas which commonly lead to relevant mobility impairments and alter the quantity and/or the quality of walking were selected: two systemic neurological diseases, four conditions primarily affecting exercise capacity, seven musculoskeletal diseases and two conditions representing sensory impairments. European Public Assessment Reports (EPARs) published by the EMA until September 2020 were examined for mobility endpoints included in their 'main studies'. Clinical study registries and primary scientific publications for these studies were also reviewed.
Results: Four hundred and eighty-four EPARs yielded 186 relevant documents with 402 'main studies'. The EPARs reported 153 primary and 584 secondary endpoints which considered mobility; 70 different assessment tools (38 patient-reported

outcomes, 13 clinician-reported outcomes, 8 performance outcomes and 13 composite endpoints) were used. Only 15.7% of those tools distinctly informed on patients' mobility status. Out of 402, 105 (26.1%) of the 'main studies' did not have any mobility assessment. Furthermore, none of these studies included a digital mobility outcome.

Conclusions: For conditions with a high impact on mobility, mobility assessment was given little consideration in the marketing authorisation of drugs by the EMA. Where mobility impairment was considered to be a relevant outcome, questionnaires or composite scores susceptible to reporting biases were predominantly used.

Introduction

Physical mobility can be defined as the ability to move by changing body position or location, or by transferring from one place to another [1]. Mobility and physical activity are known to prevent chronic diseases or mitigate their consequences [2, 3] and higher levels of total physical activity and less time spent sedentary correlate with a reduced risk for premature mortality [4]. Vice versa, sedentariness and physical inactivity have been identified as a leading cause of death worldwide, particularly in high-income countries [5, 6] and mobility characteristics such as walking speed are associated with the risk of death [7]. Physical mobility is not only a core component of quality of life but also a crucial indicator for general health. Although this is applicable to the general population, it is particularly relevant for older persons who suffer from multiple health problems.

In clinical trials, mobility outcomes can be assessed via clinician-reported outcomes (ClinRO, via investigators with specific professional training), observer-reported outcomes (ObsRO, via observers without specific professional training), patient-reported outcomes (PRO) and performance-based outcomes (PerfO, e.g. habitual performance assessments such as supervised gait speed) [8]. Composite endpoints are combinations of those outcomes. As recall and reporting bias inherent to PROs [9–11] or 'over-performing' during artificial and laboratory-based assessments of mobility in research and clinical settings [12, 13] are well-known limitations of present outcomes, objective measures in real-world environments should be considered more intensively [14]. Since human locomotion is complex, instrumented analysis by digital recording through body-worn sensors is mandatory for a better understanding of underlying pathologies [15]. Digital mobility outcomes (DMOs) have been used in clinical research for more than 20 years, and may contribute to objective measurement of mobility performace in real-world contexts, relevant to patients' daily life activities. DMOs include gait parameters and measurements of physical activity, and these are being developed to monitor the efficacy and safety, of treatments, predict future events, evaluate interventions and to stratify patients for prospective studies [16].

The relevance of mobility characteristics for clinical outcome assessments has been recognised in the IMI2 Mobilise-D initiative *Linking digital assessment of mobility to clinical endpoints to drive regulatory acceptance and clinical practice* (www.mobilise-d.eu). The Mobilise-D consortium investigates chronic obstructive pulmonary disease (COPD), multiple sclerosis (MS), Parkinson's disease (PD) and proximal femoral fracture (PFF) as common conditions which substantially contribute to the burden of disease in high-income countries. These conditions share a profound impact on mobility and a wide range of mobility impairments, including reduced gait quantity, gait asymmetry, increased gait variability and/or slow gait speed. In the evaluation of new drugs as well as medical products targeting these and other entities with high impact on mobility, the new intervention's ability to affect mobility positively or negatively should be an important criterion for decision-making. Thus, positive or negative impact on mobility need to be considered if decisions on marketing authorisation (MA) are made by regulatory bodies such as the European Medicines Agency (EMA).

To the best of our knowledge, there is no study available elucidating mobility-related outcome measurements in the MA decision process of the EMA. Therefore, we investigated mobility outcomes used in studies reported in the European Public Assessment Reports (EPARs). EPARs are scientific reports summarising the relevant information supporting a decision for or against MA. Because there is evidence for reporting bias when comparing data published in journals and data submitted for regulatory approval, we also extracted pertinent information on mobility outcomes from study registries and primary scientific publications [17, 18].

Methods

Methodological recommendations and standards outlined in the 'preferred reporting items for systematic reviews and meta-analysis statement' [19] were followed throughout this study if applicable to the EPAR documents. Data extraction from all sources (EPAR, trial databases, scientific publications) was performed independently by at least two reviewers for each condition. Consensus on inclusion of studies and mobility endpoints between the two reviewers was reached; upon dissent, a third party arbitrated.

Therapeutic areas

Fifteen therapeutic areas were selected as conditions with high impact on mobility disability: (i) MS and PD as neurological diseases; (ii) COPD, congestive heart failure (CHF), asthma and pulmonary hypertension (PH) as conditions

primarily affecting exercise capacity; (iii) ankylosing spondylitis, gout, fibromyalgia, obesity, osteoporosis, pathologic fractures, rheumatoid arthritis (RA) as musculoskeletal diseases; (iv) visual disorders and peripheral neuropathies as sensory impairments. Most of these affect older adults.

Data sources

The publicly available list of EPARs (https://www.ema.europa.eu/en/medicines/download-medicine-data), accessed as at 3 September 2020, was used to identify drugs evaluated within the centralised procedure. The EPAR webpages of the medicinal products, which are curated by EMA under their specific trade names, were reviewed to identify the initial marketing authorisation (IMA) documents and all pertinent changes in MA. Documents that did not present new data in clinical efficacy studies were excluded: these included applications for biosimilars, applications for informed consent applications and hybrid applications. Study name/s and -identifiers, primary endpoints and mobility-related secondary endpoints were extracted for all studies designated as 'main studies' within each EPAR. These 'main studies' are considered decisive by EMA for the marketing authorization.

To supplement information provided on these studies, corresponding entries in authoritative study databases were identified. As NCT numbers (unique identification code given to each clinical study upon registration at Clinicaltrials.gov) are the most commonly used study identifiers alongside acronyms, an attempt was made to identify and assign an NCT number to all trials found in the EPAR. Data on primary and secondary mobility endpoints, study start and stop dates were extracted from Clinicaltrials.gov. If no trial registry entry for a study could be found, online study registries of pharmaceutical companies were searched for study synopses produced internally. Information on exclusion/inclusion of age-groups was retrieved if available.

Primary scientific publications were identified using multiple sources: direct links from ClinicalTrials.gov were followed if available, and Pubmed and/or Google Scholar searches using study acronyms, NCT or EudraCT numbers, and suitable search terms related to medicine, investigator and disease names were conducted. Reference lists of relevant articles were manually searched. Review articles on active substances and websites of marketing authorization holders (MAH) were consulted to identify information, especially if EPARs only provided study codes assigned by MAHs. In addition, preliminary designations for active substances used by MAH were used to find pertinent studies. In general, the first publication reporting on the study's efficacy results was selected to extract data on mobility outcomes and information on exclusion/inclusion of certain age-groups.

Definition of mobility

For this analysis, mobility was defined by one's own ability to move, walk or change position via ambulation. This definition does not include the ability to use external means of transportation, which is also considered a dimension of mobility [1]. Endpoints that can indirectly influence mobility, such as fatigue, swollen joints, fractures or the assessment of 'daily activities' were not considered as endpoints assessing mobility in this sense.

Definition and categorisation of mobility assessments and mobility endpoints

Original test descriptions/questionnaires or item content tables were retrieved for all mobility assessments. For a questionnaire to be considered as mobility assessment, a single item asking for ambulation or mobility was sufficient. After verbatim extraction of the endpoints, they were simplified endpoints, they were shortened to the name of the performed test or assessment. Endpoints can be assessed at multiple time points or using different metrics (e.g. percentage responders or time-to-event), with results expressed as percentage change or absolute change in values. These variations were not considered to constitute different endpoints. For example, the American College of Rheumatology-30 (ACR30), ACR50 and ACR70 scores were often listed as distinct secondary endpoints but were counted as one secondary mobility endpoint only: 'ACR'. On the other hand, a separately reported scale (e.g. Short Form-36 Physical Functioning, SF-36-PF) from a larger score was considered to be a distinct endpoint from the score in the physical component summary: SF-36 (PCS).

To evaluate the extent to which independent and distinct reporting on mobility could be inferred from report/score of a mobility assessment, a three-tiered categorisation was used:

Category 1: High level of distinct mobility information: Assessment of specific mobility performance measures that focus on ambulation (e.g. Timed-up & Go test, 6-MWT) or self-reported mobility (e.g. Do you have difficulty when walking 400 metres?). These are reported as separate scores/scales or as single items.

Category 2: Moderate level of distinct mobility information: composite endpoints incorporating specific mobility assessments, but for which summary scores also include non-mobility items (e.g. composite endpoints including category 1 assessments, e.g. a 6-MWT or a T25-FW) OR an assessment with a dimension/scale that specifically informs about the mobility status (e.g. UPDRS II or UPDRS III, MSQOL-54) within a summary score.

Category 3: Low level of distinct mobility information: composite endpoints with a summary score combining both non-mobility items and mobility assessments from category 2 (e.g. ACR), or an assessment or questionnaire including one or more items about mobility-related aspects (e.g. getting up or staying in bed, walking around, working around the house, or being able to move single body parts), but where there is no separate subscale on mobility measures, and only a total score for the assessment is provided (e.g. SGRQ, NYHA functional class).

Analyses

Data management, calculation of relative and absolute frequencies and descriptive statistics as well as data visualisation
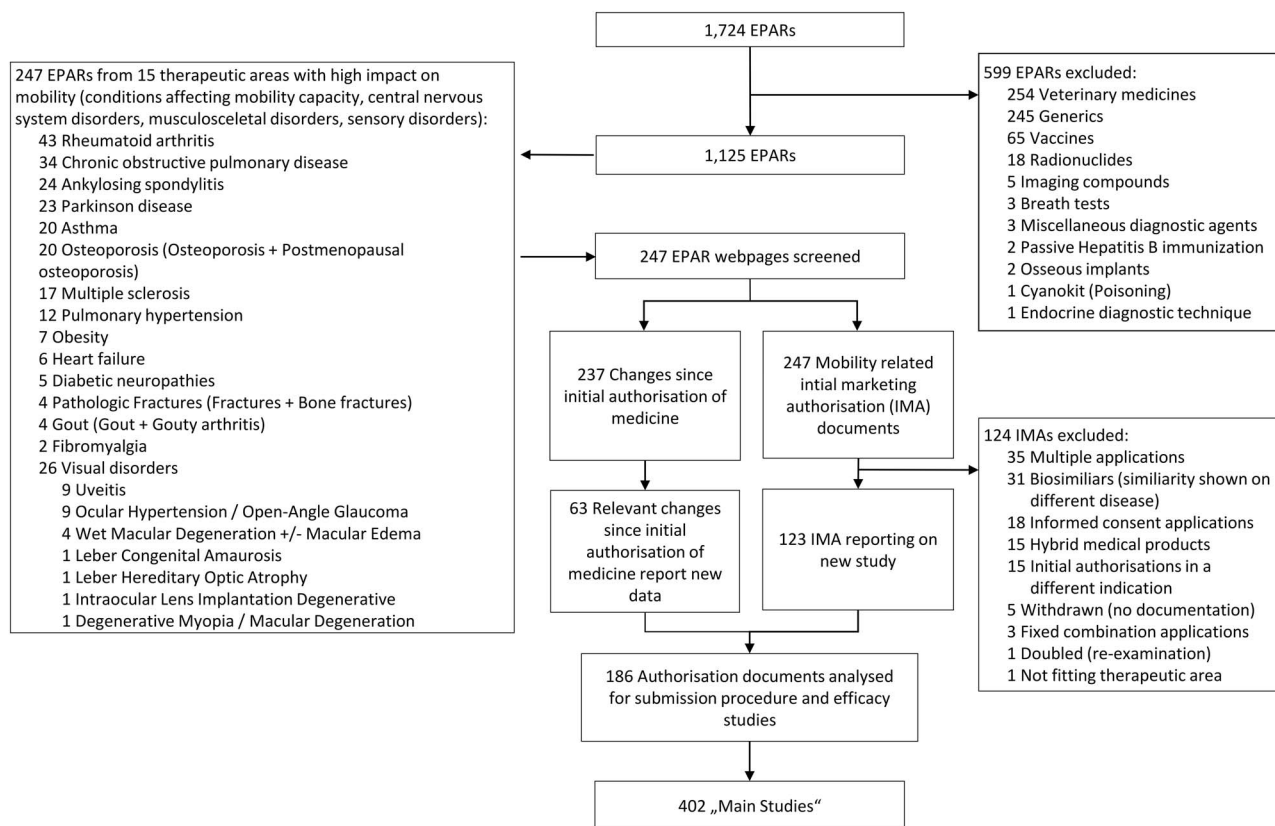
Figure 1. Study selection. Flow chart shows the selection process starting with the publicly available EMA database containing 1,724 entries, downloaded on 3 September 2020 (www.ema.europe.eu.). EPAR, European Public Assessment Report; IMA; initial marketing authorisation.

were done in R, version 4.0.2 [20] using the packages tidyverse, gghalves, version 0.1.1 and ggupset, version 0.3.0 and Inkscape, Version 0.92.

Results

We identified 247 EPAR webpages related to the 15 therapeutic areas of interest (Figure 1). In addition to the 247 IMA documents, 237 changes to MA were screened. One hundred and twenty-three IMAs reported new efficacy results. Together with the 63 changes to MA which reported on new efficacy studies, 186 authorisation documents were screened. These reported a total of 402 'main studies', which were judged to be decisive for a drug's approval by the EMA. Supplementary Table S1 lists all studies with appropriate identifiers (including acronyms, EPAR numbers), their endpoints, and information extracted from trial databases and publications. Overall, the review of EPARs, ClinicalTrials.gov/EudraCT, and publications identified 75 different mobility assessments. Data available for trial eligibility criteria ($n = 377$ studies) showed that 39% set an age limit, which was 70 or below in 17.2%. In MS, 89.4% set an upper age limit, which was 55 or younger in 63.8%. of studies. In contrast, the studies on osteoporosis did recruit the geriatric age group, as only a

minority (8.8%) of studies excluded individuals 80 years or older. No digital mobility endpoints have been reported in EPARs, clinical study databases or primary publications. Ordered by therapeutic area, Supplementary Table S2 lists all mobility tests. Here, we compiled details on the scope and nature of each mobility assessment and its measurement domain. Further its classification as PRO, ClinRO, ObsRO, PerfO, or composite endpoint and the extent to which it provides distinct information on the mobility status is reported.

According to the EPAR documents, a total of 524 primary or co-primary endpoints were employed in the 402 'main studies' that were referenced in the EPAR documents (Table 1). One hundred and fifty-three of 524 primary endpoints (29.9%) included components assessing mobility. Between therapeutic areas, large differences in the reporting of mobility outcomes were noted. In nearly half of the therapeutic areas (asthma, diabetic neuropathies, pathologic fractures, heart failure, gout, obesity and osteoporosis) none of the primary endpoints contained an assessment of mobility. In contrast, the majority of studies in rheumatic diseases reported use of mobility as a primary endpoint (ankylosing spondylitis 82.4%, rheumatoid arthritis 60.7%). Studies in neurological diseases also employed primary mobility outcomes (PD 55.8%, MS 28.3%).

Table I. Type of primary endpoints

Therapeutic areas	Area of mobility impairment	Included studies	All endpoints	PRO	ClinRO	PerfO	Biomarker	Composite endpoint	Mobility endpoints
Multiple sclerosis	Central nervous system	47	51	1 (2.0%)	43 (84.3%)	2 (4.0%)	5 (9.8%)	–	13 (25.5%)
Parkinson disease	Central nervous system	35	52	7 (13.5%)	29 (55.8%)	–	–	16 (30.1%)	29 (55.8%)
Asthma	Condition affecting mobility capacity	28	30	–	18 (60.0%)	12 (40.0%)	–	–	–
COPD	Condition affecting mobility capacity	35	45	3 (6.7%)	11 (24.4%)	31 (68.9%)	–	–	6 (13.3%)
Heart failure	Condition affecting mobility capacity	3	4	2 (50.0%)	2 (50.0%)	–	–	–	–
Pulmonary hypertension	Condition affecting mobility capacity	21	23	0	1 (4.3%)	13 (56.5%)	4 (17.4%)	5 (21.7%)	17 (73.9%)
Ankylosing spondylitis	Musculo-sceletal	9	17	15 (88.2%)	–	–	2 (11.8%)	–	14 (82.4%)
Fibromyalgia	Musculo-sceletal	4	10	10 (100%)	–	–	–	–	2 (20.0%)
Pathologic fractures	Musculo-sceletal	12	12	–	12 (100%)	–	–	–	–
Gout	Musculo-sceletal	8	10	2 (20.0%)	2 (20.0%)	–	6 (60%)	–	–
Obesity	Musculo-sceletal	24	36	–	–	–	36 (100%)	–	–
Osteoporosis	Musculo-sceletal	37	43	–	2 (4.7%)	–	41 (95.3%)	–	–
Rheumatoid arthritis	Musculo-sceletal	73	123*	22 (17.9%)	18 (14.6%)	–	14 (11.4%)	69 (56.1%)	71 (57.7%)
Diabetic neuropathies	Sensory	2	2	2 (100%)	–	–	–	–	–
Visual disorders	Sensory	64	66	2 (3.0%)	3 (4.5%)	2 (3.0%)	59 (89.4%)	–	1 (1.5%)
All	**Total**	**402**	**524**	**66**	**141**	**60**	**167**	**90**	**153 (29.9%)**

*Information in EPAR for four studies on Arava was lacking unequivocal information whether endpoints were primary or secondary endpoints.

For therapeutic areas with at least one primary mobility endpoint, Figure 2 shows the proportion of the total number of primary endpoints which contain mobility components by therapeutic area, as well as the use of specific types of mobility assessment within each therapeutic area. To measure mobility, PROs only were used in ankylosing spondylitis (Assessment in Ankylosing Spondylitis) and fibromyalgia [Fibromyalgia Impact Questionnaire, SF-36(PCS)]. In MS, a ClinRO (the EDSS, Extended Disability Status Score) was the most common primary mobility endpoint. In PH a PerfO measuring mobility was very frequently used [(6-MWT in 18 out of 21 studies, 85.7%)]. Composite mobility endpoints were most common in PD and RA, with a mobility component included in traditional disease-specific scores such as the Movement Disorders Society-revised version of the Unified Parkinson Disease Rating Scale (UPDRS) and the ACR score. When mobility was assessed within a composite score, this was most frequently done within a ClinRO (11/17, 64.7%).

Assessment of mobility was reported more frequently as a secondary endpoint in the EPARs. Figure 3A shows the number of secondary mobility endpoints reported by therapeutic area. The number of secondary endpoints assessing mobility in asthma (22 secondary mobility endpoints/28 studies), obesity (12/24) and visual disorders (33/64) was relatively small. In osteoporosis, only three secondary mobility endpoints (EQ-5D, Qualiost® and 'Disability due to a fracture or to back pain') were reported in 37 studies. In contrast, in PD (108 secondary mobility endpoints/35 studies), PH (48/21) and RA (150/73) secondary mobility endpoints were used more frequently.

In terms of types of endpoints, 332 of 581 (57.1%) secondary mobility endpoints were PROs, 136 (23.4%), were ClinROs, 78 (13.4%) were composite endpoints and 35 (6.0%) were PerfOs. In asthma, diabetic neuropathies, fibromyalgia, gout, and obesity, mobility was exclusively assessed by PROs. The 6-MWT (in PH) and the MSFC (in MS) were the most frequently used PerfOs. Figure 3A illustrates that some generic quality of life questionnaires—such as EQ-5D, and SF-36(PCS)—are used across multiple therapeutic areas. The number of unique mobility assessments varied by therapeutic area (e.g. 14 disease-specific mobility endpoints in Parkinson's compared to zero in diabetic neuropathies). Figure 3B provides more detail on therapeutic areas with more than three different unique mobility assessments (ankylosing spondylitis, COPD, MS, PD, PH). For instance, in addition to EDSS and MSFC, 12 other indication-specific mobility assessments were reported in MS trials.

Figure 4 shows the year of publication of the individual studies in a scientific journal, ordered by therapeutic area. There was a large number of new studies in asthma, MS, RA, PH and osteoporosis. Notably, the advent of new biologicals for eosinophilic asthma led to a recent peak in new studies that were evaluated in the EPARs. For 105 of the 402 studies (26.1%), EPARs did not report any assessment of mobility (grey dots). Only a minority of reported assessment tools
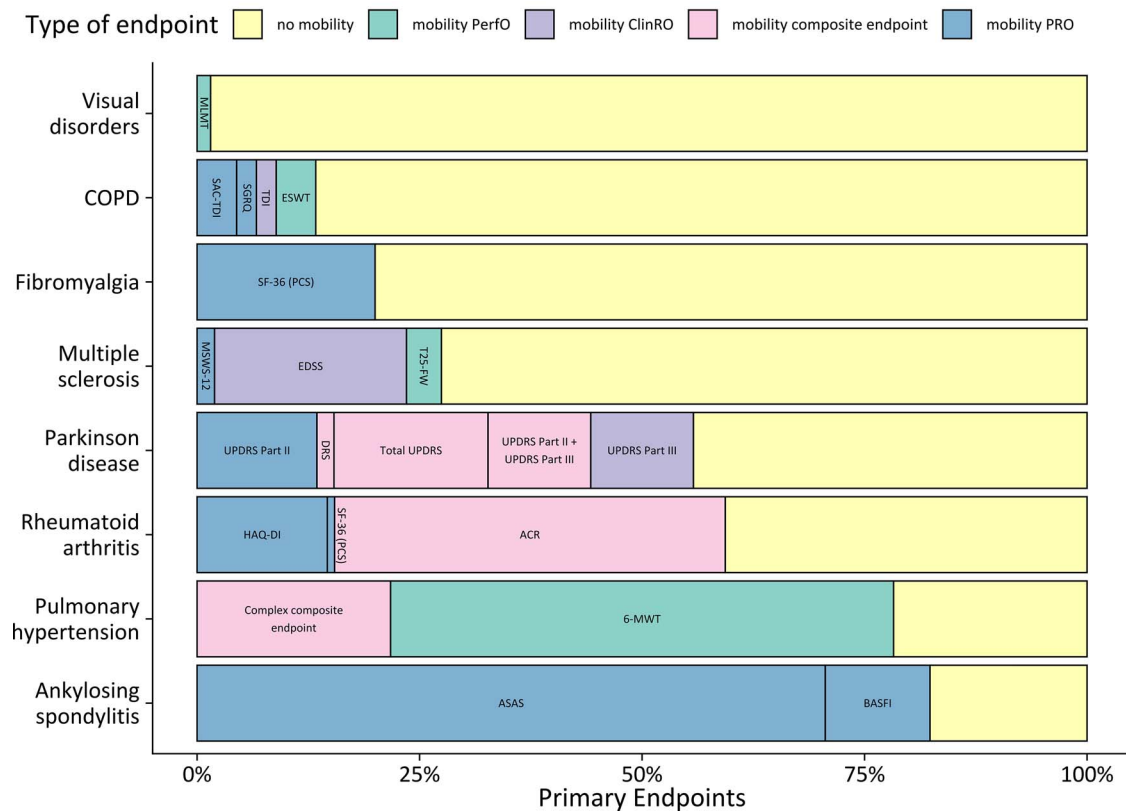
Figure 2. Mobility endpoint as a primary endpoint. Percentage of primary endpoints which assessed a mobility aspect in relation to total primary endpoints. Only therapeutic areas with at least one primary mobility endpoint are shown. The stacked bars depict subtypes of mobility endpoints: mobility PRO (patient-reported outcome), mobility PerfO (performance outcome), mobility ClinRO (clinician-reported outcome) and composite endpoints containing a mobility assessment component.

(11 out of 70 (15.7%) provide highly distinct information about the patients' mobility status (category 1). Twenty-one (30%) of the assessment tools included tests or subscores assessing mobility, but with less specific information on patient's mobility status (category 2). More than half of assessments ($n = 38$, 54.3%) fell in category 3, conferring the least distinct information on the patient's mobility status.

The reported frequencies and distributions of primary and secondary mobility endpoints are based on information about the studies given in the EPAR documents. To address a possible publication bias, endpoints in study registries and scientific publications were additionally extracted and matched with the data from EPARs. The majority of mobility endpoints were mentioned across all sources. However, there was a small number of mobility assessments ($n = 5$) that were only reported in study registries or scientific publications: the HUI-3 questionnaire in ankylosing spondylitis, the QUAL-EFFO questionnaire in osteoporosis, the Lewin-Technology Assessment Group in obesity, the BODE-index in COPD, and the multi-dimensional health assessment questionnaire in fibromyalgia.

Discussion

This study investigated the role of mobility outcomes in regulatory approvals by the EMA, using publicly available EPARs, supplemented by data from registries and publications. Four hundred and two studies of 116 different drugs in 15 therapeutic areas were included.

Our analysis shows that even in diseases that result in significant mobility restrictions, mobility endpoints still play a minor role in MA of new drugs. In 26.1% of the studies, EPARs did not report any mobility assessment as a primary or secondary endpoint. Although mobility assessments were employed more frequently as secondary endpoints (a total of 581 secondary mobility endpoints were found in our search), these are of less importance for regulatory approval. A significant proportion of the reported secondary mobility endpoints are generic quality of life assessments containing mobility aspects as only one component, e.g. EQ-5D, SF-36(PCS). These are not primarily employed to provide information about mobility, but may form the basis for subsequent evaluations for use in health technology assessments.

Importantly, the reported mobility endpoints differ considerably in providing distinct information on mobility impairment or the subjects' mobility capability. In the studies included, only a minority of mobility assessments (15.7%) provided distinct information on mobility status, impairment and disability, e.g walking distance covered during the 6-MWT or the 12-item MS walking scale. In contrast, considerably less distinct information is available from assessments classified as moderate (category 2; 30%) or
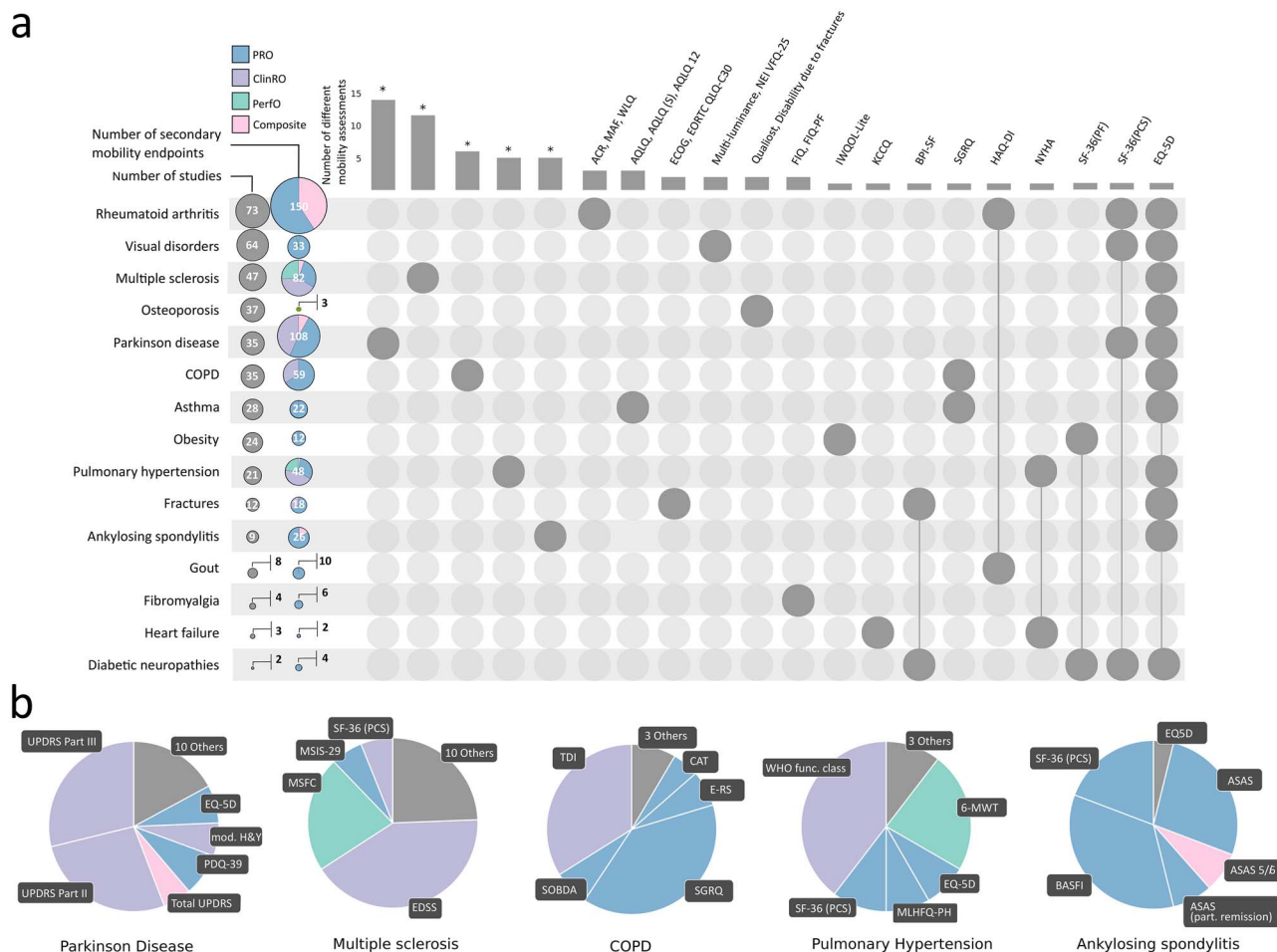
Figure 3. Secondary mobility endpoints. Panel A: Left column: therapeutic areas ordered by number of studies. Adjacent column: total numbers of mobility endpoints and their subtypes. The sizes of the circles are proportional to the total numbers. Top: column chart: number of different mobility assessments performed by therapeutic area; names of mobility tests are printed above the column if their number does not exceed three; therapeutic areas with more than three different mobility assesements are marked with an asterisk (∗). Tests used in multiple therapeutic areas are connected by lines. The therapeutic area corresponding to the column is identified by a grey dot in the network plot below. Panel B: therapeutic areas with more than three different mobility tests. Pie charts show the total number of secondary mobility endpoints employed in the studies of the five therapeutic areas. 'Other' denotes secondary mobility endpoints ocurring in less than 5% of all secondary mobility endpoints within each therapeutic area. ACR, American College of Rheumatology; ASAS, Assessment in Ankylosing Spondylitis; BASFI, Bath Ankylosing Spondilitis Functional Index; CAT, COPD Assessment Test; E-RS, Evaluating Respiratory Symptoms in Chronic Obstructive Pulmonary Disease; SOBDA, Shortness of Breath with Daily Activities questionnaire; MSIS, Multiple Sclerosis Impact Scale; MSFC, Multiple Sclerosis Functional Composite; UPDRS, Unified Parkinson's Disease Rating Scale; mod.H&Y, modified Hoehn and Yahr; PDQ, Parkinson's Disease Questionnaire; MLHFQ-PH, Minnesota Living With Heart Failure Questionnaire; 6-MWT, Six Minute Walk Test; MAF, Multi-dimensional Assessment of Fatigue; WLQ, Work Limitations Questionnaire; AQLQ, Asthma quality of Life Questionnaire; ECOG, Eastern Co-operative of Oncology Group Performance Score; EORT, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core; NEI-VFQ, National Eye Institute 25-item Visual Function Questionnaire; FIQ, Fibromyalgia Impact Questionnaire Physical Functioning; WLQ, Work Limitations Questionnaire; KCCQ, Kansas City Cardiomyopathy Questionnaire—Clinical Summary Score; BPI, Brief Pain Inventory; SGRQ, St. George's Respiratory Questionnaire; HAQ-DI, Health Assessment Questionnaire—Disability Index; NYHA, New York Heart Association Functional Class; SF-36(PF), Short Form 36 Physical Functioning Subscale; SF-36(PCS), Short Form 36 Physical Component Summary; EQ-5D, EuroQol 5 Dimensions.

low (category 3; 54.3%). Even if single items (e.g. a mobility-specific questions) or mobility performance assessments in composite endpoints are included, the summary score reported remains difficult to interpret in relation to the patient's mobility status. A low total score may still be caused by impairments unrelated to mobility.

Our analysis shows that mobility is mostly considered within a broader context of general quality of life
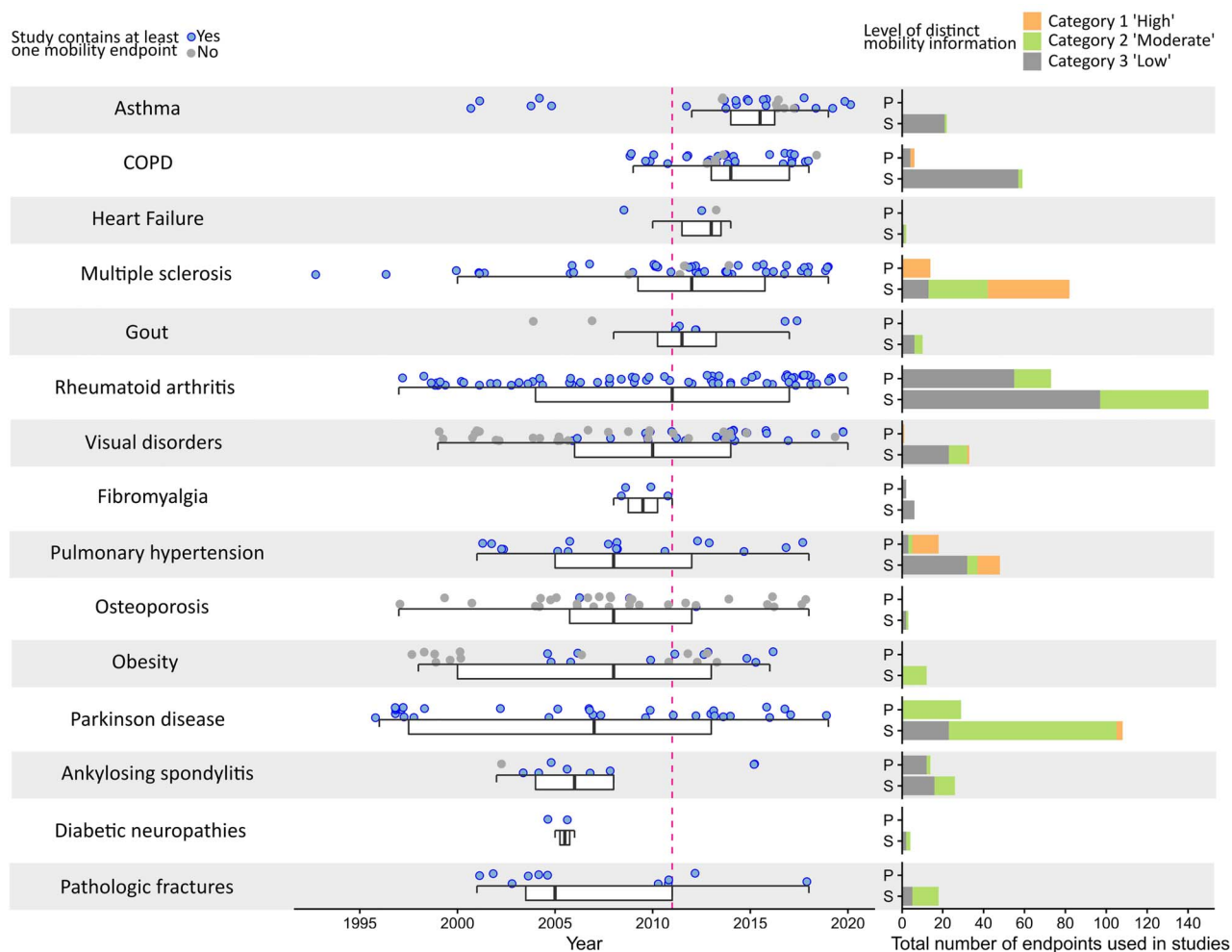
Figure 4. Key time points for included studies, EPARs and scientific publications. Left column: Publication date of first efficacy results of clinical studies. Studies that report on mobility are coloured blue. Boxplots show the median as thick bar, the lower and upper hinges of the box correspond to the 25th and 75th percentiles, whiskers extend to 1.5×inter-quartile range. For 20 studies (Osteoporosis *n* = 4, Visual disorders *n* = 10, Parkinson disease *n* = 2, Rheumatoid arthritis *n* = 2, Multiple sclerosis *n* = 1 and Pulmonary hypertension *n* = 1) dates on study start/stop were not available. Right column: Specific information on mobility disability deducible from primary (p) and secondary (s) mobility endpoints. All mobility endpoints were categorised into three tiers. Category 1: High level of distinct mobility information. Category 2: Moderate level of distinct mobility information. Category 3: Low level of distinct mobility information. For the complete definition of categories, please see Methods section.

assessments, but rarely as a stand-alone specific clinical outcome measure. Considering the relevance of mobility for health, independence, and high levels of quality of life, this is regrettable. Regulators, like EMA, could have enough leverage to promote their use, but guidance documents regarding the use of endpoints, e.g. 'EMA guidelines on the clinical investigation of medicinal products' (for the list of documents, see Supplementary Table S3) do rarely recommend distinct mobility assessments. Companies can also seek scientific advice from EMA during the drug development. However, as demonstrated by one of the most recent MA of romosuzumab in osteoporosis, even after scientific advice for clinical aspects, distinct mobility assessments were not performed [21, 22]. In addition,

studies conducted for MA of biosimilars, for example in RA (Amgevita/Solymbic, Amsparity, Benepali, Flixcabi, Hulio, Imraldi, Inflectra/Remsima and Nepexto) have not taken the opportunity to incorporate distinct mobility assessments neither. By contrast, the recently published VITALITY-HFpEF study of Vericiguat in heart failure gave high priority to mobility. In this study, the physical limitation score of the Kansas City Cardiomyopathy Questionnaire, a mobility PRO, was evaluated as a primary endpoint and a PerfO (the 6-MWT) was used as a key secondary endpoint [23].

The importance of mobility loss especially in older age is widely acknowledged, as is the fact that representation of the geriatric population in pivotal studies must be improved (e.g. [24]). Across our set of studies and therapeutic areas, 17.2%

excluded patients older than 70 from enrolment. Thus, recruitment of geriatric patients might be more restricted by co-morbidities and other factors such as sensory deficits or mobility disability.

Precise assessment of mobility in the older people might benefit especially from new technologies. In an era when smartphones and wristband wearables are being purchased in their hundreds of millions, a comprehensive assessment of mobility could include not only perception (by PROs) or capacity (by PerfO), but also objectively reflect the real-world mobility. Yet, in our analysis, we were not able to identify a single study using DMOs such as step counting, real-world walking speed or cadence [25], even when reviewing tertiary and exploratory endpoints as far as they have been reported. The inclusion of DMOs adds an objective, real-time dimension to the measurement of mobility that cannot be replaced by other clinical outcome assessments. Mobility perceptions recorded by PROs are quite susceptible to external influences, and performance outcomes, usually measured in clinical settings, may not reflect mobility capacity in patients' real-life environment. Assessing mobility using PROs where data are obtained retrospectively is also limited by recall bias, which is especially important in older people due to cognitive deterioration. The capability of DMOs to establish dose–response relationships and to detect non-linear relationships with relevant outcomes such as mortality or hospital admission is also well-established [4]. The sensitivity and statistical characteristics of obtained parameters from digital monitoring promise high responsiveness, making DMOs very promising endpoints for use in clinical trials. Taken together, considering the high validity, broad accessibility, ease of use, and relatively low costs, DMOs should be considered as important assessment measures in future studies. An early example is the phase 2 study conducted in 2008–2009 of ataluren (Translarna®) in Duchenne's muscular dystrophy, where a step activity monitor was already used to measure physical mobility and a qualification opinion on stride velocity as a secondary endpoint was published by the EMA (EMEA/H/C/002720).

Since EPARs reflect the MA process, it can be assumed that endpoints not mentioned in an EPAR were not considered relevant for regulatory approval. However, given that versions of the EPAR documents were very condensed in the early 2000s, detailed reporting of secondary endpoints could have been judged as dispensable at that time. It was hypothesised that information about mobility endpoints not reported in the EPAR may have been reported in study registries and/or scientific publications. However, only five mobility assessments unreported in the EPARs could be extracted from these sources.

Some limitations of our analysis should be acknowledged. This review includes only reports on drugs for which companies have applied via the centralised procedure and EPARs have been published. Information is not included from reports from approval procedures by national bodies, whose policies and attitudes on mobility might differ. In a large majority of cases, study protocols were not available. Study protocols may contain information on exploratory or additional secondary endpoints which have not been mentioned in registries or EPARs because these data were used for internal investigation only or to support discussions with payers. This study has limited the review journal publications to the first main publication, and subsequent publications may have highlighted specific mobility aspects as well as quality of life assessments incorporating mobility items.

Our study on the use of mobility endpoints in the MA of new medicines within EMA's centralised procedure indicates the discrepancy between the rising recognition of mobility and physical activity in the health sector and broader society and the limited priority given to their assessment by clinical trialists, regulators and pharmaceutical companies.

References

1. World Health Organisation. International Classification of Functioning, Disability and Health: ICF Online, http://id.who.int/icd/entity/2048203604 (23 November 2021, date last accessed).

2. Piercy KL, Troiano RP, Ballard RM *et al.* The physical activity guidelines for Americans. JAMA 2018; 320: 2020–8.

3. Klenk J, Kerse N. Every step you take. BMJ 2019; l5051. https://doi.org/10.1136/bmj.l5051.

4. Ekelund U, Tarp J, Steene-Johannessen J *et al.* Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ 2019; 366: l4570.

5. Kohl HW, Craig CL, Lambert EV *et al.* The pandemic of physical inactivity: global action for public health. Lancet 2012; 380: 294–305.

6. Booth FW, Roberts CK, Laye MJ. Lack of exercise is a major cause of chronic diseases. Compr PhysiolWiley 2012; 2: 1143–211.

7. Studenski S, Perera S, Patel K *et al.* Gait speed and survival in older adults. JAMA - J Am Med Assoc 2011; 305: 50–8.

8. Walton MK, Powers JH, Hobart J *et al.* Clinical outcome assessments: conceptual foundation—report of the ISPOR clinical outcomes assessment – emerging good practices for outcomes research task force. Value Health 2015; 18. https://doi.org/10.1016/j.jval.2015.08.006.

9. Stone AA, Shiffman S, Schwartz JE *et al.* Patient non-compliance with paper diaries. BMJ 2002; 324: 1193–4.

10. Bottomley A, Jones D, Claassens L. Patient-reported outcomes: assessment and current perspectives of the guidelines of the Food and Drug Administration and the reflection paper of the European Medicines Agency. Eur J Cancer 2009; 45: 347–53.

11. Cerreta F, Ritzhaupt A, Metcalfe T *et al.* Digital technologies for medicines: shaping a framework for success. Nat Rev Drug Discov 2020; 19: 573–4.

12. Hillel I, Gazit E, Nieuwboer A *et al.* Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring. Eur Rev Aging Phys Act 2019; 16: 6. https://doi.org/10.1186/s11556-019-0214-5.

13. Warmerdam E, Hausdorff JM, Atrsaei A *et al.* Long-term unsupervised mobility assessment in movement disorders. Lancet Neurol 2020; 19: 462–70.

14. Viceconti M, Hernandez Penna S, Dartee W *et al.* Toward a regulatory qualification of real-world mobility performance biomarkers in Parkinson's patients using digital mobility outcomes. Sensors 2020; 20: 5920.

15. Lord S, Galna B, Rochester L. Moving forward on gait measurement: toward a more refined approach. Mov Disord 2013; 28: 1534–43.

16. Rochester L, Mazzà C, Mueller A *et al.* A roadmap to inform development, validation and approval of digital mobility outcomes: the mobilise-D approach. *Digit*. Biomarkers 2020; 4: 13–27.

17. Eyding D, Lelgemann M, Grouven U *et al.* Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. BMJ 2010; 341: c4737. https://doi.org/10.1136/bmj.c4737.

18. Moreno SG, Sutton AJ, Turner EH *et al.* Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. BMJ 2009; 339: b2981. https://doi.org/10.1136/bmj.b2981.

19. Moher D, Liberati A, Tetzlaff J *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009; 339: 332–6.

20. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2021. https://www.R-project.org/.

21. Amgen Inc. Statistical Analysis Plan 201101142. 2017. https://clinicaltrials.gov/ProvidedDocs/14/NCT01631214/SAP_002.pdf.

22. European Medicines Agency. EMEA/H/C/004465/0000 Evenity EPAR. 2019. https://www.ema.europa.eu/en/documents/assessment-report/evenity-epar-public-assessment-report_en.pdf.

23. Armstrong PW, Lam CSP, Anstrom KJ *et al.* Effect of Vericiguat vs placebo on quality of life in patients with heart failure and preserved ejection fraction: the VITALITY-HFpEF randomized clinical trial. JAMA - J Am Med Assoc 2020; 324: 1512–21.

24. van Marum RJ. Underrepresentation of the elderly in clinical trials, time for action. Br J Clin Pharmacol 2020; 86: 2014–6.

25. Polhemus AM, Delgado-Ortiz L, Brittain G *et al.* Walking on common ground: a cross-disciplinary scoping review on the clinical utility of digital mobility outcomes. npj Digit Med 2021; 4: 149.