

Series: Pragmatic trials and real world evidence: Paper 6. Outcome measures in the real world

Paco M. Welsing^{a,*}, Katrien Oude Rengerink^a, Sue Collier^b, Laurent Eckert^c, Maarten van Smeden^a, Antonio Ciaglia^d, Gaele Nachbaur^e, Sven Trelle^f, Alik J. Taylor^g, Matthias Egger^h, Iris Goetzⁱ, on behalf of Work Package 3 of the GetReal Consortium

^aJulius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

^bGSK, Respiratory R&D, Stockley Park West, Brentford, Middlesex TW89GS, UK

^cHealth Economics and Outcome Research, Sanofi, Avenue Pierre Brossolette, 91385, Chilly-Mazarin, France

^dInternational Alliance of Patients' Organizations, 49-51 East Road, London, N1 6AH, UK

^eMedical Department, GSK, 100 route de Versailles, Marly-le-Roi 78163, France

^fDepartment of Clinical Research, CTU, University of Bern, Finkenhubelweg 11, CH-3012, Bern, Switzerland

^gGlobal Outcomes Research, Takeda Development Centre Europe Ltd, 61 Aldwych, WC2B 4AE, London, UK

^hInstitute of Social and Preventive Medicine, University of Bern, Finkenhubelweg 11, CH-3012, Bern, Switzerland

ⁱGlobal Patient Outcomes & Real World Evidence, Eli Lilly and Company Ltd, Erl Wood Manor, Sunninghill Road, Windlesham, Surrey, GU20 6PH, UK

Accepted 12 December 2016; Published online 11 May 2017

Abstract

Results from pragmatic trials should reflect the comparative treatment effects encountered in patients in real-life clinical practice to guide treatment decisions. Therefore, pragmatic trials should focus on outcomes that are relevant to patients, clinical practice, and treatment choices. This sixth article in the series (see Box) discusses different types of outcomes and their suitability for pragmatic trials, design choices for measuring these outcomes, and their implications and challenges. Measuring outcomes in pragmatic trials should not interfere with real-world clinical practice to ensure generalizability of trial results, and routinely collected outcomes should be prioritized. Typical outcomes include mortality, morbidity, functional status, well-being, and resource use. Surrogate endpoints are typically avoided as primary outcome. It is important to measure outcomes over a relevant time horizon and obtain valid and precise results. As pragmatic trials are often open label, a less subjective outcome can reduce bias. Methods that decrease bias or enhance precision of the results, such as standardization and blinding of outcome assessment, should be considered when a high risk of bias or high variability is expected. The selection of outcomes in pragmatic trials should be relevant for decision making and feasible in terms of executing the trial in the context of interest. Therefore, this should be discussed with all stakeholders as early as feasible to ensure the relevance of study results for decision making in clinical practice and the ability to perform the study. © 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Pragmatic trial; Real-world evidence; Outcome measurement; Generalizability

1. Introduction

Pragmatic trials focus on collecting clinical evidence on the comparative effects of treatments that is relevant to patients, caregivers, and therapeutic decision making rather than on generating evidence of the direct pharmacological or biological effect of the intervention [1–3]. The aim of a pragmatic trial is to ensure generalizability of study results to the target patient group that will be treated in day-to-day clinical practice. The measurement of outcomes should therefore not interfere with routine practice to maintain the real-world character of the trial [4]. For instance,

more frequent study-related follow-up visits could positively influence treatment adherence above that in usual care. Apart from this direct interference, just the awareness of being in a trial can already have an effect on behavior (an example of the Hawthorne effect) [5]; however, following routine clinical practice closely will probably minimize this effect.

It can be challenging to align the real-world aspects of the trial with a methodologically sound trial design, while satisfying all stakeholders regarding their outcome requirements. Defining and operationalizing the appropriate outcomes may therefore be a demanding task. This sixth article in the series (see Box) discusses the challenges and aims to provide guidance for selecting the most

* Corresponding author. Tel.: +31-611104143.

E-mail address: p.m.j.welsing@umcutrecht.nl (P.M. Welsing).

What is new?

Key findings

- Outcomes in pragmatic trials should be relevant for patients, physicians, and clinical decision making. Surrogate endpoints should generally be avoided.
- Typical outcomes include mortality, morbidity, functional status, well-being, and resource use measured over a relevant time horizon irrespective of whether the treatment protocol is followed.
- As pragmatic trials are often open label, the inclusion of objective outcome measures can reduce the risk of bias. Routinely collected outcomes should be prioritized.

What this adds to what was known?

- This article discusses the challenges and provides guidance for selecting the most appropriate outcomes in pragmatic trials and measuring them with minimal impact on routine clinical practice.

What is the implication and what should change now?

- Outcome measurements should interfere with clinical practice as little as possible, but standardization and blinding of outcome assessment should be considered when there is a high risk of bias or high variability is expected.
- Patients who are not treated according to the protocol should remain followed for the collection of outcome data if possible to enable a true intention-to-treat analysis.
- The selection of outcomes in pragmatic trials should be discussed with all stakeholders to ensure the relevance of study results for decision making in clinical practice and the ability to perform the study.

appropriate outcomes in pragmatic trials and measuring them with minimal impact on routine clinical practice.

2. Selection of outcome measures

An outcome is the result of a process or an event. In trials, an outcome (or endpoint) is a measurable change in health or quality of life that can be influenced by treatment [6]. Outcomes in clinical trials generally relate to mortality, pathophysiological manifestations related to the disease process, morbidity, functional status, quality of life, and/or resource use and costs, as shown in Fig. 1 [7,8]. The full impact of a treatment on patients' health is of key

Box Series on pragmatic trials

Pragmatic trials aim to generate real-world evidence on the (relative) effects of treatments, generalizable to routine practice. In this series, we will discuss options and choices for pragmatic trial design, operational consequences, and the interpretation of results.

1. Introduction
2. Setting, sites, and investigator selection
3. Patient selection challenges and consequences
4. Informed consent
5. Usual care and real life comparators
- 6. Outcome measures in the real world**
7. Safety, quality and monitoring
8. Data collection and management

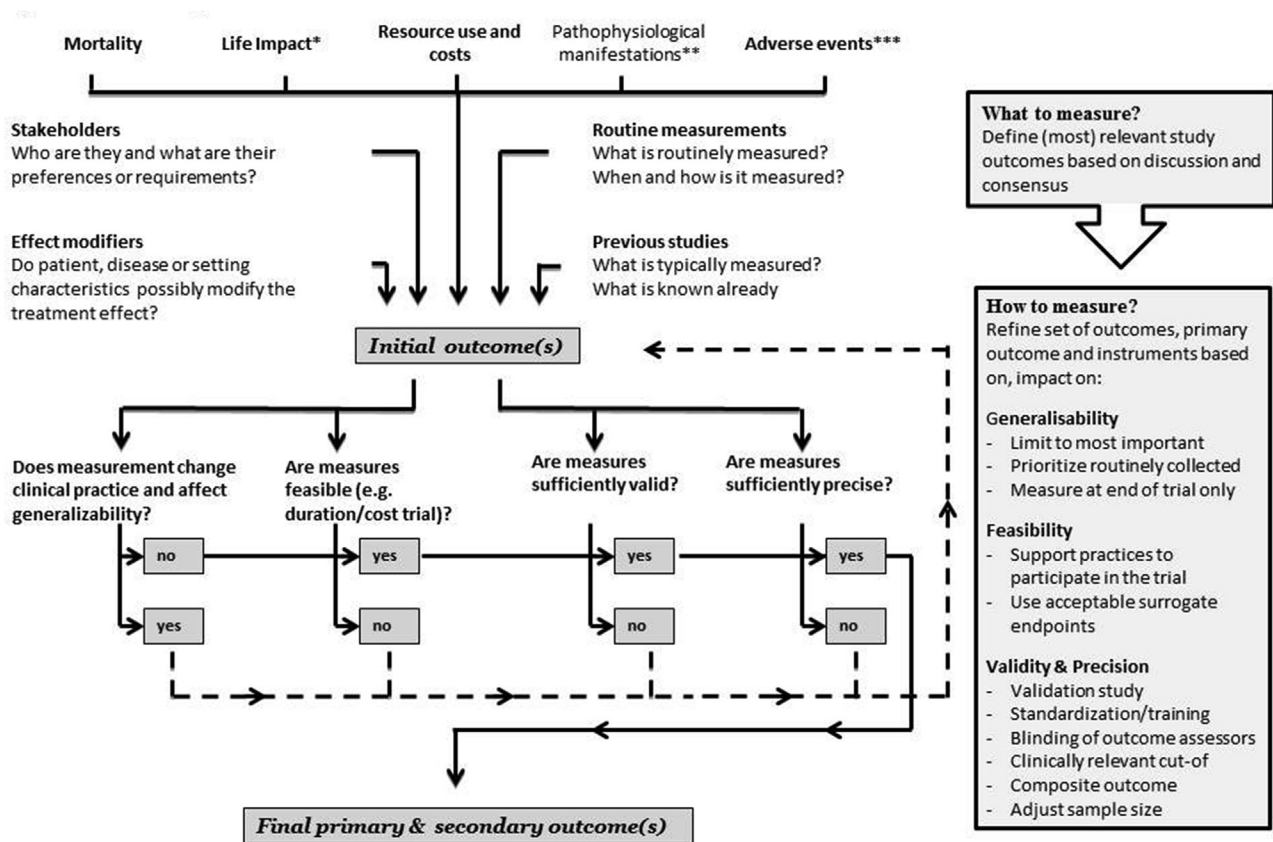
interest but may become apparent only after several years of follow-up, which may challenge the feasibility of the study in terms of its duration and costs.

As depicted in Fig. 1, it is important to consider different types of outcomes and discuss them with all stakeholders including an assessment of what is measured in the usual practice settings and the comparability of outcomes with those used in previous similar studies [7,8]. Feasibility, interference with practice, validity, and precision of measurement instruments should be evaluated to refine the selected set of outcomes and measures. Ideally, pragmatic trials directly provide a valid and sufficiently precise estimate of the treatment effect to support clinical decision making [2,4,9].

The Salford Lung Study (SLS), depicted in Fig. 2 as an example, is phase III pragmatic trial in patients with chronic obstructive pulmonary disease (COPD) that compared usual treatment with a novel dry-powder inhaler. Patients underwent routine clinical care with minimal study visits to assess the annual rate of moderate/severe COPD exacerbations as the primary outcome. Primary endpoints were obtained directly through the electronic medical record, whereas secondary endpoints such as quality of life and number of self-reported instances of exacerbations were recorded at the final 12-month follow-up study visit (as to not interfere with clinical practice during the study). The study and outcome measurement were discussed with regulators and a health technology assessment (HTA) agency as the main stakeholders for this prelaunch pragmatic trial [10].

2.1. Surrogate outcomes

A surrogate outcome is a substitute for a direct measurement of a clinical endpoint. It can be a laboratory measurement, imaging measurement, or a physical sign as a substitute for an endpoint that relates to how a patient feels, functions, or survives [11]. Bone density, for instance, can be measured as a surrogate for osteoporotic fractures.



* Life impact outcomes are outcomes that directly measure how a patient feels or functions in daily life ** General considered less important in pragmatic trials, including surrogate measures *** Special type of outcome with many regulations see paper on monitoring of safety in this series Adapted from Boers et al. 2014 (7)

Fig. 1. Choosing what outcomes to measure and how to measure them.

Changes induced by a therapy on a surrogate outcome are assumed to reflect changes in the clinical endpoint of primary interest. However, the relation between the surrogate endpoint and the clinical endpoint is usually uncertain, and obtaining a precise estimate of the treatment effect relevant for clinical decision making is generally not possible using surrogate endpoints [1,4,12,13]. Surrogate outcomes in a pragmatic trial, therefore, may only be acceptable as primary endpoint when measuring a clinical endpoint is not possible and high-quality evidence exists on their relation with endpoints directly relevant for decision making [8,12]. An example of acceptable surrogate endpoint is the viral load for trials evaluating human immunodeficiency virus treatment [14,15].

2.2. Patient-reported outcomes

Patient-reported outcomes (PROs) are outcomes provided directly by the patient without any interpretation by others. These outcomes often assess symptoms, such as pain or fatigue, and their impact, interference with activities in a patients' daily lives, or treatment satisfaction [16,17]. Health-related quality of life (HRQoL) is a specific type of PRO that measures the degree to which the medical

condition or its treatment impacts an individual's well-being, including physical and mental domains, and can be regarded the ultimate effect of health care [18]. Measures may provide preference values, expressed in utility scores, which allow the calculation of quality-adjusted life years that are often used in economic evaluations of a treatment. In SLS, a COPD Assessment Test and a utility HRQoL instrument, the EuroQol-5 dimension, were used as outcomes (Fig. 2).

PROs often are not routinely assessed in clinical practice, but rather as part of the informal assessment of the patient's well-being. If they can be assessed without major interference of routine care, PROs may be optimally used as outcomes in a pragmatic trial [16,17,19,20]. For example, assessing PROs that are relevant for disease management might be beneficial for the quality of health care and simultaneously offer the opportunity to use these data for outcome assessment in clinical trials. In this regard, the core PROs of the National Institutes of Health Collaboratory is a worthy initiative (Table 1) [25].

2.3. Primary and secondary outcomes

Clinical trials generally assess multiple outcomes or endpoints to evaluate differences between treatments.

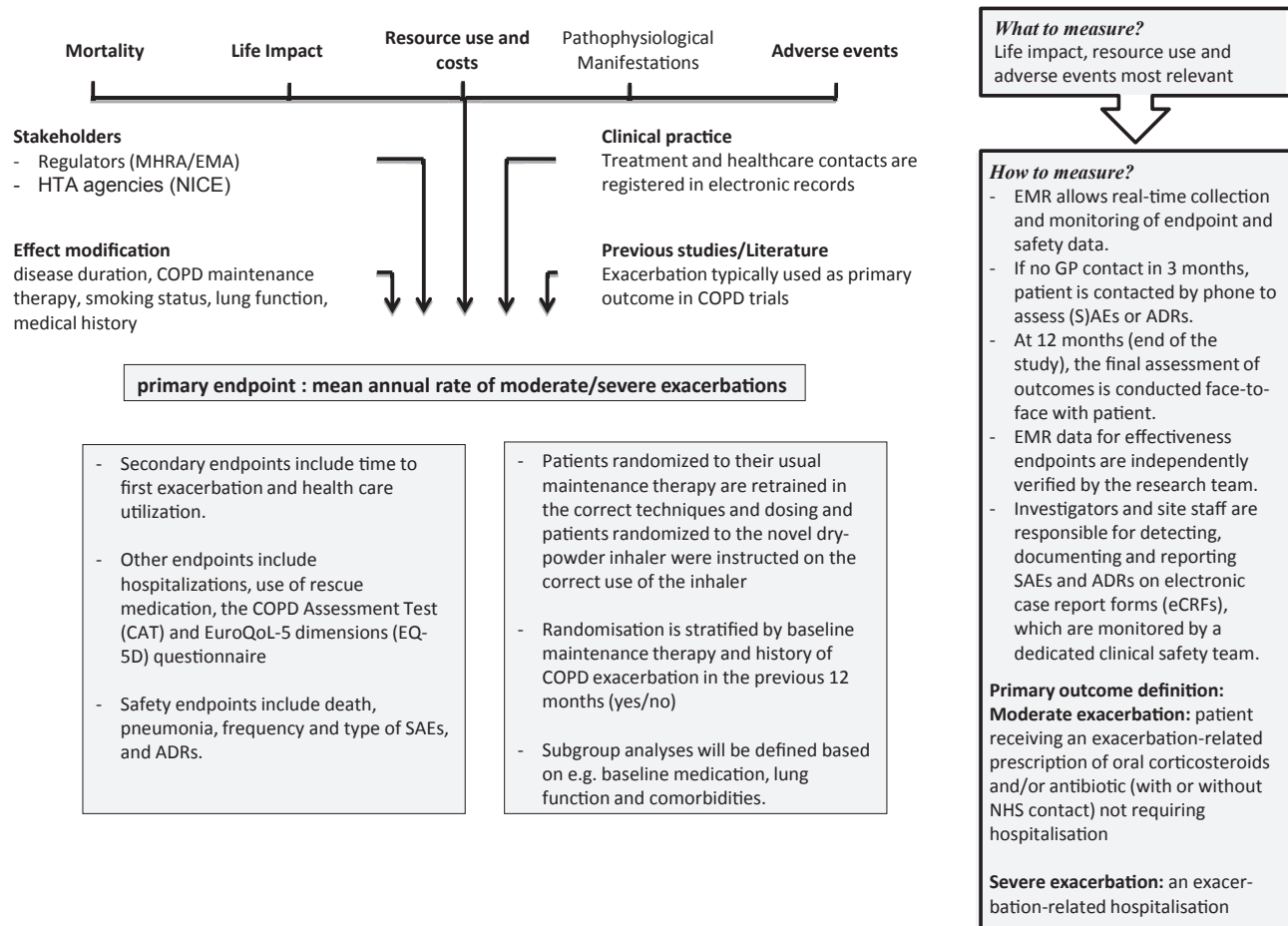


Fig. 2. Outcomes in the Salford Lung Study. MHRA, Medicines and Healthcare products Regulatory Agency; EMA, European Medicines Agency; HTA, health technology assessment; NICE, National Institute for Health and Clinical Excellence; COPD, chronic obstructive pulmonary disease; SAEs, serious adverse events; ADRs, adverse drug reactions; EMR, electronic medical record; GP, general practitioner; NHS, National Health Service.

The sample size of the study is determined to obtain sufficient statistical power to detect or refute a relevant difference between treatments on the primary outcome. Secondary outcome measures are evaluated more exploratively, and a larger sample size may be needed to provide sufficient statistical power. Therefore, using outcomes of interest as secondary outcomes only may result in imprecise effect estimates complicating decision making. The primary outcome in a pragmatic trial should therefore be the outcome most influential for the clinical decision, the total set of outcomes measured in a trial should reflect all important effects of the intervention, and ensuring sufficient power for important secondary outcomes may be needed.

Schwartz and Lellouch [1] suggested that the outcome of a pragmatic trial should be a single criterion, or a combination of multiple weighed criteria, relevant for making a treatment decision in routine clinical practice. Different outcomes can be relevant for decision making, including outcomes such as adverse events, treatment adherence, and comedication. Assigning weights to their

individual importance may be difficult however; and a different mix of outcomes can lead to the same result on such composite outcome, which might complicate the interpretation [8]. An advantage of a composite endpoint is that it may yield higher power and feasibility of executing the trial compared with separate outcomes with a low incidence. Although different outcomes with congruent results will strengthen the conclusions of the study, contradicting results can complicate it. For example, in a study, it may be found that a new antiallergy treatment decreases the symptom score but simultaneously increases the use of concomitant immunotherapy [30]. In addition, the evaluation of multiple outcomes will increase the overall chance of type I errors. To control this type I error, several options are available, but they may decrease precision of the results or necessitate a larger sample size to have sufficient power to detect a difference on all outcomes [31,32]. More conservative *P*-values, decision rules specified a priori, or gatekeeping procedures can be used. An example of a decision rule can be that success may only be claimed if there is a

Table 1. International initiatives for harmonization of outcome measurements

| | |
|--|---|
| The Core Outcome Measures in Effectiveness Trials www.comet-initiative.org | Brings together standardized sets of outcomes in a database that can be searched on disease, type of intervention, and stakeholders involved. These sets represent the minimum that should be reported in all clinical trials of a specific condition, making it easier for the results of trials to be compared, contrasted, and/or combined. This can be a useful resource for selecting outcomes, and the website contains useful references for core outcome set developers as well [21]. |
| Outcome Measures in Rheumatology (OMERACT) http://www.omeract.org/omeract_publications.php | Is one of the oldest initiatives standardizing outcome measures in clinical trials that is still active. Although this initiative focuses on rheumatology, several resources relevant for a broader disease range are available on their website like the OMERACT Handbook on developing core outcome measurement sets [7]. |
| International Consortium for Health Outcomes Measurement (ICHOM) https://www.linkedin.com/company-beta/2760048/ | Aims to transform health care systems worldwide by measuring and reporting patient outcomes in a standardized way in daily practice. Focus is on outcomes related to questions patients have regarding the impact of treatment on their life. ICHOM advocates to define standard sets of outcomes per medical condition and then drives adoption to enable health care providers globally to compare, learn, and improve [22,23]. |
| Healthmeasures.net http://www.healthmeasures.net/ | A useful resource bringing together four flexible and comprehensive measurement systems. These assess physical, mental, and social health, symptoms, well-being, and life satisfaction; along with sensory, motor, and cognitive functions. It also provides guidance for their use [24]. |
| Core Patient Reported Outcomes of the NIH Collaboratory https://www.nihcollaboratory.org/cores/Pages/default.aspx | Aims to create guidelines and define best practices with respect to selecting the most appropriate PRO measures and stimulates the development of new instruments when needed, creating efficient, high-quality PRO data collection systems compatible with electronic health records and registries and conducting statistical analyses of PRO endpoints [25]. |
| The CONsensus-based Standards for the selection of health Measurement Instruments http://www.cosmin.nl | Aims to improve the selection of health measurement instruments among others by providing a tool on standards for evaluating the quality of studies on measurement properties of measurement instruments [26]. |
| Global HealthCare system roadmaps and Pharmacoeconomic guidelines around the world Eldessouki R, Smith MD. Health care system information sharing: a step toward better health globally. <i>Value Health Reg Issues</i> 2012;1:118–129. http://www.ispor.org/HTARoadmaps/Default.asp http://www.ispor.org/peguidelines/index.asp | These tools from the International Society of Pharmacoeconomics and Outcomes Research may be a useful resource for evaluating and comparing country- and agency-specific outcome preferences. On these website, HTA requirements and decision processes with country-/region-specific information on preferred outcomes and preference for efficacy or effectiveness and generalizability of results among many more characteristics of the evidence requirements can be searched and compared in tables [27–29]. |

Abbreviations: NIH, National Institutes of Health; PRO, patient-reported outcome; HTA, health technology assessment.

significant improvement in all (primary) outcomes, gate-keeping procedures entail organizing primary and secondary hypotheses of interest a priori, and only proceed with testing when the significance criteria for the previous hypothesis are satisfied. The final selection of outcomes in pragmatic trials should be relevant for decision making and feasible in terms of trial conduct.

2.4. Satisfying all stakeholders

Stakeholders in the prelaunch and postlaunch phases may have different priorities or requirements for study outcomes. The pharmaceutical industry and regulatory agencies are the most important stakeholders for prelaunch drug trials, whereas after market release, these are mainly HTA agencies, payers, health care providers, and patients. Study outcomes for a pragmatic trial performed in the prelaunch phase should probably aim to satisfy outcome

requirements for prelaunch and postlaunch stakeholders requiring careful discussions with these stakeholders as done in the SLS (Fig. 2). Regulatory agencies typically assess efficacy and safety using (prelaunch) explanatory randomized controlled trials, whereas HTA agencies also evaluate more pragmatic or observational studies and may use meta-analytic approaches. The European Medicine Agency states that HRQoL measures are generally not acceptable as a primary outcome, and they also state that surrogate endpoints may be used as primary endpoints [33,34]. The EUnetHTA guidelines recommend to include a disease- or population-specific and a generic HRQoL measure that adequately captures the impact of a disease on daily life. The relative effectiveness assessment of pharmaceuticals should be based where possible on patient-relevant clinical endpoints such as morbidity and overall mortality [8]. Furthermore, outcomes that are considered important by research physicians are not

necessarily important for patients [35–38]. For example, it was found that heart failure patients deem the absence of dyspnea important, although this was generally not an endpoint in heart failure trials [39]. Patients with diabetes find mortality, the development of kidney failure, and dialysis dependence more important than HbA1c levels, a surrogate outcome typically used as a primary outcome, although these preferred outcomes may influence the feasibility of conducting the trial [13]. Collaboration between patients with rheumatoid arthritis and the rheumatology research community has identified that fatigue is often an outcome of top priority, often ranking as high or higher than pain as outcome. These findings have led to international consensus that fatigue should be measured in all clinical trials in addition to the existing core set of outcome measures for clinical trials [40].

The different opinions and requirements related to outcomes should be discussed with all stakeholders (Fig. 1). Several international initiatives and resources exist bringing together the so-called core outcome sets, stimulating the consistent measurement of relevant outcomes in clinical practice and trials. These core outcome sets also enable easier comparison of study results and are generally developed with involvement of different stakeholders. They are directly relevant for pragmatic trials when all relevant stakeholder views are taken into account, and measurements do not interfere with usual practice to an extent that results are no longer generalizable. Otherwise existing sets and resources can be helpful in selecting and agreeing on a set of outcomes and how to measure them for a specific study and indication, group of stakeholders, and/or region (Table 1).

3. Measuring outcomes

Instruments to assess the outcome should be sufficiently accurate (ie, valid, reliable, and responsive) and feasible in terms of costs and complexity [41–43]. Deviations from routine clinical practice should be minimized, also to promote the acceptability of measurements to patients and care providers who participate in a study [7,44–46]. It may be beneficial for decision making to apply a predefined cutoff value for continuous outcome measures in a pragmatic trial when resulting categorical outcomes have a clearer meaning for interpretation of trial results at patient level (ie, the probability of recovery). For example, in an explanatory trial studying the biologic effect of treatment for high blood pressure, the decrease in millimeters of mercury in systolic blood pressure could be the main outcome, whereas in a pragmatic trial, a drop in systolic blood pressure below 130 mm Hg may be more appropriate as primary outcome. For some outcomes, the concept of minimally clinically important difference may be useful to define such cutoff [47,48]. Using a predefined cutoff also poses disadvantages such as loss of precision [49].

3.1. Routinely collected data to define outcomes

Data that are recorded in (electronic) medical files during routine clinical practice should be preferred as outcomes in pragmatic trials. Many algorithms have been created to define outcomes based on data that have been routinely recorded in electronic health files [50–53]. Ricketts et al. [50] developed an algorithm to define progression-free survival in patients with head and neck cancer, using data on patient admission, chemotherapy, and radiotherapy. This automated technique resulted in progression-free survival estimates in line with what was found by manual chart review in 82% of the patients [50]. Based on the instances of diarrhea and fatigue recorded in the electronic health record, nonresponse, partial, or complete response have been calculated after 1 year of treating patients with Crohn disease or ulcerative colitis with antibodies to tumor necrosis factor alpha, which correlated well with the need for surgery and hospitalization [51]. Rubbo et al. [52] reviewed 31 studies on the validity of an automated diagnosis of acute myocardial infarction and found positive predictive values of more than 70%. Although these outcome definitions can be useful in pragmatic trials, such algorithms are generally only validated for the specific data set and may need to be validated or newly developed and validated for a specific trial [53], which can increase complexity and may delay the trial.

3.2. Accurate and complete outcome assessment

Accurate and complete measurement of relevant outcomes is key for the validity and precision of trial results and applicability of the results in clinical practice. In contrast to explanatory trials, outcomes in a pragmatic trial should be measured with the inclusion of extraneous effects of comedication, nonadherence, and placebo effects. In pragmatic trials, therefore, patients and physician are often aware of the treatment group (ie, not blinded) as knowledge of the treatment (including related placebo effects and effects on care) is part of real-world practice. Patients who are not treated according to the protocol should also remain followed for the collection of outcome data if possible to enable a true intention-to-treat analysis [3,54].

Self-reported and subjective outcomes can be valuable indicators of disease impact; however, they may be sensitive to observer bias. For example, when opinions about the merits and disadvantages differ between treatments compared, physicians and patients may (unintentionally) report the severity of the disease differently or tend to report safety concerns quicker on new treatment [55,56]. Although patients and physicians should not be blinded, blinded outcome assessment, if possible, is recommended to minimize observer bias in pragmatic trials [57]. Measurements in practice may also be performed in a less uniform way, leading to higher variability in outcomes necessitating a higher sample size for the study or even resulting in biased estimates of effectiveness [58,59]. When

outcome measurement is systematically underestimating/overestimating the actual outcome, this results in biased estimates of comparative treatment effect. Standardization of measurement and/or training of assessors may increase the precision of measurements and a measurements' validity and may make it less subjective. Standardization and training also increase complexity and may compromise generalizability but should be considered when not doing so would compromise the precision and/or validity of results. An external adjudication committee can be installed for a blinded and uniform assessment of the outcome, which may take away the need to train physicians on standardized outcome assessment. However, adjudicating intermediate outcomes (eg, osteoporosis) does not reflect clinical practice, and the adjudicated outcome does not represent future actions based on the presence of osteoporosis and the development of final outcomes (eg, fractures). The value of adjudication for validity and precision of trial results is not certain, and it also increases the complexity and decreases the feasibility of the trial, and therefore, its use should be determined on a per-trial basis [60].

When clinical examinations are only prompted by physical complaints, this may lead to missing outcome data that can jeopardize the validity, generalizability, and/or precision of the study results. The impact of missing data on the results of the study depends on the reason why data are missing and can be addressed by several approaches as described elsewhere [61]. Data that are missing completely at random will not systematically affect or bias the outcome measurement (but do decrease precision). When data are missing at random, biases can be overcome using methods such as multiple imputation, where missing values are replaced by appropriate estimations based on nonmissing data, that allow individuals with incomplete data to be included in analyses. Biases caused by data that are missing not at random can only be addressed by sensitivity analyses that examine the effects of different missing data assumptions. As pragmatic trials follow routine practice as closely as possible, measurements may be more variable, and missing values may be more pertinent than in explanatory trials when not routinely performed. A feasibility study may identify the key challenges and can be helpful to adapt the design where needed to obtain valid and precise outcome measurements (Fig. 1).

4. Additional measurements

Characteristics of the patient, disease, and setting of the trial can modify the treatment effect. Measuring potential effect modifiers, such as age, socioeconomic status, ethnicity, type of medical practice, comorbidity, concomitant medication, and treatment adherence, can help to clarify whether the study population in the pragmatic trial is comparable to the patients who are encountered in clinical practice (Fig. 1). Furthermore, exploring the relation between

treatment effect and effect modifiers could improve the understanding of the treatment effect in subgroups and may help to explain differences between pragmatic and explanatory clinical trials investigating the same intervention [62]. Formally studying effect modification in the study will affect sample size or require oversampling of subgroups [62].

5. Conclusion

Estimates of treatment effect in pragmatic trials should be directly generalizable to clinical practice and relevant to patients and decision making in real-world clinical practice. The primary outcome should typically be the outcome most relevant for the research question and decision problem at hand and analysis based on the true intention-to-treat population. Outcome measurement should not interfere with usual practice, and routinely collected outcome measures should therefore be preferred. Relevance, feasibility, interference with practice, validity, and precision of outcome measurement should be evaluated with all stakeholders, and the final set of outcomes and measures may be more tailor-made, and choices should therefore be reported clearly.

Acknowledgments

The work leading to these results has received support from the Innovative Medicines Initiative Joint undertaking under grant agreement no. 115546, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007/2013) and European Federation of Pharmaceutical Industries and Associations (EFPIA) companies in kind contribution. The research leading to these results was conducted as part of the GetReal consortium and included literature review and interviews with stakeholders from academia, research institutions, contract research organizations, pharmaceutical industry, regulatory authorities, health care insurers, HTA agencies, general practitioners, and patient organizations. For further information, please refer to <http://www.imi-getreal.eu/>.

References

- [1] Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Clin Epidemiol* 2009;62:499–505.
- [2] Sackett DL. Explanatory and pragmatic clinical trials: a primer and application to a recent asthma trial. *Pol Arch Med Wewn* 2011;121: 259–63.
- [3] Zuidgeest MGP, Goetz I, Groenwold RHH, Irving E, van Thiel GJM, Grobbee DE. Series: Pragmatic trials and real world evidence: Paper 1. Introduction. *J Clin Epidemiol* 2017;88:7–13.
- [4] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;350:h2147.

- [5] McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol* 2014;67:267–77.
- [6] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390.
- [7] Boers M, Kirwan JR, Wells G, Beaton D, Gossece L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT Filter 2.0. *J Clin Epidemiol* 2014;67:745–53.
- [8] EUnetHTA: Endpoints used for relative effectiveness assessment of pharmaceuticals. Available at <http://www.eunetha.eu/eunetha-guidelines>. Accessed May 29, 2017.
- [9] Institute of Medicine. Initial national priorities for comparative effectiveness research. Washington, DC: The National Academies Press; 2009.
- [10] Bakerly ND, Woodcock A, New JP, Gibson JM, Wu W, Leather D, et al. The Salford Lung Study protocol: a pragmatic, randomised phase III real-world effectiveness trial in chronic obstructive pulmonary disease. *Respir Res* 2015;16:101.
- [11] De Gruttola VG, Clax P, DeMets DL, Downing GJ, Ellenberg SS, Friedman L, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. Summary of a National Institutes of Health Workshop. *Contemp Clin Trials* 2001;22:485–502.
- [12] Chokshi DA, Avorn J, Kesselheim AS. Designing comparative effectiveness research on prescription drugs: lessons from the clinical trial literature. *Health Aff* 2010;29:1842–8.
- [13] Murad MH, Shah ND, Van Houten HK, Ziegenfuss JY, Deming JR, Beebe TJ, et al. Individuals with diabetes preferred that future trials use patient-important outcomes and provide pragmatic inferences. *J Clin Epidemiol* 2011;64:743–8.
- [14] Human immunodeficiency virus type 1 RNA level and CD4 count as prognostic markers and surrogate end points: a meta-analysis. HIV Surrogate Marker Collaborative Group. *AIDS Res Hum Retroviruses* 2004;16:1123–33.
- [15] Antiretroviral Therapy Cohort Collaboration (ART-CC), Vandenhende MA, Ingle S, May M, Chene G, Zangerle R, Van Sighem A, et al. Impact of low-level viremia on clinical and virological outcomes in treated HIV-1-infected patients. *AIDS* 2015;29:373–83.
- [16] Basch E. New frontiers in patient-reported outcomes: adverse event reporting, comparative effectiveness, and quality assessment. *Annu Rev Med* 2014;65:307–17.
- [17] Lipscomb J, Reeve BB, Clauser SB, Abrams JS, Bruner DW, Burke LB, et al. Patient-reported outcomes assessment in cancer trials: taking stock, moving forward. *J Clin Oncol* 2007;25:5133–40.
- [18] Osoba D, King M. Meaningful differences. In: Fayers P, Hays RD, editors. *Assessing quality of life in clinical trials*. New York, NY: Oxford Press; 2005:244–57.
- [19] Kroenke K, Monahan PO, Kean J. Pragmatic characteristics of patient-reported outcome measures are important for use in clinical practice. *J Clin Epidemiol* 2015;68:1085–92.
- [20] Wu AW, Kharrazi H, Boulware LE, Snyder CF. Measure once, cut twice adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. *J Clin Epidemiol* 2013;66(8 Suppl):S12–20.
- [21] Available at www.comet-initiative.org. Accessed May 29, 2017.
- [22] Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med* 2016;374:504–6.
- [23] Available at <https://www.linkedin.com/company-beta/2760048/>. Accessed May 29, 2017.
- [24] Available at <http://www.healthmeasures.net/>. Accessed May 29, 2017.
- [25] Available at <https://www.nihcollaboratory.org/cores/Pages/default.aspx>. Accessed May 29, 2017.
- [26] Available at <http://www.cosmin.nl>. Accessed May 29, 2017.
- [27] Eldessouki R, Smith MD. Health care system information sharing: a step toward better health globally. *Value Health Reg Issues* 2012;1:118–29.
- [28] Available at <http://www.ispor.org/HTARoadmaps/Default.asp>. Accessed May 29, 2017.
- [29] Available at <http://www.ispor.org/peguidelines/index.asp>. Accessed May 29, 2017.
- [30] Pfaar O, Anders C, Klimek L. Clinical outcome measures of specific immunotherapy. *Curr Opin Allergy Clin Immunol* 2009;9:208–13.
- [31] Bender R, Lange S. Adjusting for multiple testing—when and how. *J Clin Epidemiol* 2001;54:343–9.
- [32] Mascha EJ, Turan A. Joint hypothesis testing and gatekeeping procedures for studies with multiple endpoints. *Anesth Analg* 2012;114:1304–17.
- [33] General Considerations for Clinical Trials (ICH Topic E 8). Available at <http://www.ema.europa.eu/ema/>. Accessed May 29, 2017.
- [34] Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products. EMA 2005, Guidance document ref nr: CPMP/EWP/139391/04, Available at http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_001238.jsp&mid=WC0b01ac0580032ec4.
- [35] Janse AJ, Gemke RJ, Uiterwaal CS, van der Tweel I, Kimpen JL, Sinnema G. Quality of life: patients and doctors don't always agree: a meta-analysis. *J Clin Epidemiol* 2004;57:653–61.
- [36] Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication. *JAMA* 2008;288:3027–34.
- [37] Arora NK, Jensen RE, Sulayman N, Hamilton AS, Potosky AL. Patient-physician communication about health-related quality-of-life problems: are non-Hodgkin lymphoma survivors willing to talk? *J Clin Oncol* 2013;31:3964–70.
- [38] Petersen MA, Larsen H, Pedersen L, Sonne N, Groenvold M. Assessing health-related quality of life in palliative care: comparing patient and physician assessments. *Eur J Cancer* 2006;42:1159–66.
- [39] Pisa G, Eichmann F, Hupfer S. Assessing patient preferences in heart failure using conjoint methodology. *Patient Prefer Adherence* 2015;9:1233–41.
- [40] Kirwan J, Minnock P, Adebajo A, Bresnihan B, Choy E, de Wit M, et al. Patient perspective: fatigue as a recommended patient-centred outcome measure in rheumatoid arthritis. *J Rheumatol* 2007;34:1174–7.
- [41] Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol* 2001;54:1204–17.
- [42] Walton MK, Powers JH, Hobart J, Patrick D, Marquis P, Vamvakas S, et al. ISPOR TASK FORCE report clinical outcome assessments: conceptual foundation—report of the ISPOR clinical outcomes assessment—emerging good practices for Outcomes Research Task Force. *Value Health* 2015;18:741–52.
- [43] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *J Clin Epidemiol* 2010;63:737–45.
- [44] Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;22:1889–905.
- [45] Oude Rengerink K, Kalkman S, Collier S, Ciaglia A, Worsley SD, Lightbourne A, et al. Pragmatic trials and real world evidence: Paper 3. Patient selection challenges and consequences. *J Clin Epidemiol* 2017;89:173–80.
- [46] Worsley SD, Oude Rengerink K, Irving E, Lejeune S, Mol K, Collier S, et al. Series: Pragmatic trials and real world evidence: Paper 2. Setting, sites, and investigator selection. *J Clin Epidemiol* 2017;88:14–20.
- [47] Curtis JR, Yang S, Chen L, Pope JE, Keystone EC, Haraoui B, et al. Determining the minimally important difference in the clinical disease

- activity index for improvement and worsening in early rheumatoid arthritis patients. *Arthritis Care Res (Hoboken)* 2015;67:1345–53.
- [48] Conijn AP, Jonkers W, Rouwet EV, Vahl AC, Reekers JA, Koelemay MJ. Introducing the concept of the minimally important difference to determine a clinically relevant change on patient-reported outcome measures in patients with intermittent claudication. *Cardiovasc Intervent Radiol* 2015;38:1112–8.
- [49] Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006;332:1080.
- [50] Ricketts K, Williams M, Liu ZW, Gibson A. Automated estimation of disease recurrence in head and neck cancer using routine health-care data. *Comput Methods Programs Biomed* 2014;117:412–24.
- [51] Ananthakrishnan AN, Cagan A, Cai T, Gainer VS, Shaw SY, Savova G, et al. Identification of nonresponse to treatment using narrative data in an electronic health record inflammatory bowel disease cohort. *Inflamm Bowel Dis* 2016;22:151–8.
- [52] Rubbo B, Fitzpatrick NK, Denaxas S, Daskalopoulou M, Yu N, Patel RS, et al. Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: a systematic review and recommendations. *Int J Cardiol* 2015;187:705–11.
- [53] Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol* 2012;65:343–349.e2.
- [54] Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials* 2007;4:350–6.
- [55] Ware JH, Hamel MB. Pragmatic trials—guides to better patient care? *N Engl J Med* 2011;364:1685–7.
- [56] van Staa TP, Dyson L, McCann G, Padmanabhan S, Belatri R, Goldacre B, et al. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess* 2014;18:1–146.
- [57] Minns Lowe CJ, Wilson MS, Sackley CM, Barker KL. Blind outcome assessment: the development and use of procedures to maintain and describe blinding in a pragmatic physiotherapy rehabilitation trial. *Clin Rehabil* 2011;25:264–74.
- [58] Buzkova P. Measurement error and outcomes defined by exceeding a threshold: biased findings in comparative effectiveness trials. *Pharm Stat* 2012;11:429–41.
- [59] Lachin JM. The role of measurement reliability in clinical trials. *Clin Trials* 2004;1:553–66.
- [60] Pogue J, Walter SD, Yusuf S. Evaluating the benefit of event adjudication of cardiovascular outcomes in large simple RCTs. *Clin Trials* 2009;6:239–51.
- [61] Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Br Med J* 2009;338:b2393.
- [62] Rothwell PM. Treating individuals 2. Subgroup analysis in randomized controlled trials: importance, indications, and interpretation. *Lancet* 2005;365:176–86.