Accepted Manuscript

Medical Rehabilitation: Guidelines to Advance the Field with High-Impact Clinical Trials

Marcas M. Bamman, PhD, Gary R. Cutter, PhD, David M. Brienza, PhD, John Chae, MD, Daniel M. Corcos, PhD, Stephanie DeLuca, PhD, Edelle Field-Fote, PT, PhD, Mona N. Fouad, MD, MPH, Catherine E. Lang, PT, PhD, Anne Lindblad, PhD, Robert W. Motl, PhD, Carla G. Perna, CCRA, CCRP, Darcy Reisman, PT, PhD, Kenneth M. Saag, MD, MSc, Sean I. Savitz, MD, Kathryn H. Schmitz, MPH, PhD, Jennifer Stevens-Lapsley, PT, PhD, John Whyte, MD, PhD, Carolee J. Winstein, PT, PhD, Mary E. Michel, PhD

PII: S0003-9993(18)31113-4

DOI: 10.1016/j.apmr.2018.08.173

Reference: YAPMR 57345

To appear in: *ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION*

Received Date: 31 December 2017

Revised Date: 11 July 2018

Accepted Date: 15 August 2018

Please cite this article as: Bamman MM, Cutter GR, Brienza DM, Chae J, Corcos DM, DeLuca S, Field-Fote E, Fouad MN, Lang CE, Lindblad A, Motl RW, Perna CG, Reisman D, Saag KM, Savitz SI, Schmitz KH, Stevens-Lapsley J, Whyte J, Winstein CJ, Michel ME, Medical Rehabilitation: Guidelines to Advance the Field with High-Impact Clinical Trials, *ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION* (2018), doi: 10.1016/j.apmr.2018.08.173.

Running Head: Rehabilitation Clinical Trial Guidelines

Title: Medical Rehabilitation: Guidelines to Advance the Field with High-Impact Clinical Trials

Marcas M Bamman, PhD[1,2,3], Gary R Cutter, PhD[1,2,4], David M Brienza, PhD[5], John Chae, MD[6], Daniel M Corcos, PhD[7], Stephanie DeLuca, PhD[8], Edelle Field-Fote, PT, PhD[9], Mona N Fouad, MD, MPH[1,10], Catherine E Lang, PT, PhD[11], Anne Lindblad, PhD[12], Robert W Motl, PhD[1,2,13], Carla G Perna, CCRA, CCRP[1,2], Darcy Reisman, PT, PhD[14], Kenneth M Saag, MD, MSc[1,15], Sean I Savitz, MD[16], Kathryn H Schmitz, MPH, PhD[17], Jennifer Stevens-Lapsley, PT, PhD[18], John Whyte, MD, PhD[19], Carolee J Winstein, PT, PhD[20], and Mary E Michel, PhD[21]

[1]NIH National Rehabilitation Research Resource to Enhance Clinical Trials (REACT, P2CHD086851); [2]UAB Center for Exercise Medicine, University of Alabama at Birmingham, Birmingham, AL 35205; [3]Department of Cell, Developmental, and Integrative Biology, University of Alabama at Birmingham, Birmingham, AL 35294; [4]Section on Research Methods and Clinical Trials, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294; [5]Department of Rehabilitation Science and Technology and McGowan Institute for Regenerative Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA 15219; [6]Department of Physical Medicine and Rehabilitation, Case Western Reserve University, Cleveland, OH 44106; [7]Department of Physical Therapy and Human Movement Sciences, Northwestern University, Chicago, IL 60611; [8]Virginia Tech Carilion Research Institute, Roanoke, VA 24016; [9]Virginia C. Crawford Research Institute, Shepherd Center, Atlanta, GA 30309 and Division of Physical Therapy, Emory University School of Medicine, Atlanta, GA 30322; [10]Division of Preventive Medicine, Department of Medicine and Minority Health and Health Disparities Research Center, University of Alabama at Birmingham, Birmingham, AL 35294; [11]Physical Therapy, Occupational Therapy, and Neurology Washington University School of Medicine, St. Louis, MO 63108; [12]The Emmes Corporation, Rockville, MD 20850; [13]Department of Physical Therapy, University of Alabama at

Birmingham, Birmingham, AL 35294; [14]Department of Physical Therapy, University of Delaware, Newark, DE 19716; [15]Division of Clinical Immunology and Rheumatology, Department of Medicine and Center for Outcomes and Effectiveness Research and Education, University of Alabama at Birmingham, Birmingham, AL 35294; [16]Department of Neurology, University of Texas Health Science Center, Houston, TX 77030; [17]Department of Public Health Sciences, Pennsylvania State University, State College, PA 16801; [18]Physical Therapy Program, University of Colorado, Denver, CO 80045; [19]Moss Rehabilitation Research Institute, Einstein Medical Center, Philadelphia, PA 19141; [20]Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA 90089; [21]National Center for Medical Rehabilitation Research, National Institutes of Health, Bethesda, MD 20847 (retired)

Acknowledgments: We sincerely thank Dr. Alison Cernich, Director, and Dr. Ralph Nitkin, Program Official, for their NCMRR leadership, and for their support of the clinical trials workshop summarized in this manuscript. Supported in part by NIH Grant P2CHD086851.

Conflicts of Interest: The authors disclose the following relationships:

J Chae discloses a relationship with: SPR Therapeutics.
G Cutter discloses relationships with: Pythagoras (Board membership), Amo Pharmaceuticals, Teva Neuroscience, EMD Serono, Novartis, Pfizer, Sanofi, AMO, Genzyme, Medimmune, Receptos, Gilead, Sanofi-Aventis, Apotex/Modigenetec, Opko, NHLBI, NICHD, Ono/Merck, Genentech, GSK, Transparency Life Sciences, Roche, Ophazyme, Somahlution, Horizon Pharma, Reata Pharma, PTC Therapeutics, Klein-Buendel, Merck/Pfizer, UT Houston, Medday, Roche, Sciflour, Axon, and GW Pharmaceuticals.
S Savitz discloses relationships via his institution (UTHealth) with: Athersys, Genentech, Pfizer, Dart Neuroscience, SanBio, Let's Cure CP, the Texas Institute for Rehabilitation and Research Foundation, Cord Blood Registry Systems, Neuralstem, Mesoblast, BlueRock, ReNeuron, Lumosa, Celgene, Dart Neuroscience, and Aldagen.

Word Counts: Main Text = 6159; Online Appendix = 2242

Address for Correspondence: Marcas M Bamman, PhD, UAB Center for Exercise Medicine, 1313 13th Street South, Room 300, University of Alabama at Birmingham, Birmingham, AL 35205-5327

Phone 205-975-9042, mbamman@uab.edu

Title: Medical Rehabilitation: Guidelines to Advance the Field with High-Impact Clinical Trials

ABSTRACT

The purpose of this Special Communication is to summarize guidelines and recommendations stemming from an expert panel convened by the National Institutes of Health (NIH), National Center for Medical Rehabilitation Research (NCMRR) for a workshop entitled, The Future of Medical Rehabilitation Clinical Trials, held 29-30 September 2016 at the NCMRR offices in Bethesda, Maryland. The ultimate goal of both the workshop and this summary is to offer guidance on clinical trials design and operations to the medical rehabilitation research community, with the intent of maximizing the impact of future trials.

KEY WORDS

Medical rehabilitation; Clinical trials; NIH workshop

PURPOSE AND OVERVIEW

The conduct guidelines, review processes, monitoring, and ulitimate outcome expectations of clinical trials sponsored by the National Institutes of Health (NIH) have evolved substantially over the past decade. This has occurred alongside a rapidly evolving landscape of clinical and translational research in medical rehabilitation – a field that is burgeoning to meet the growing demands of both chronic and traumatic disease management. The medical rehabilitation field finds itself in a unique and challenging position of advancing interventions that extend healthspan, but must do so with high quality clinical trials. To that end, the NIH National Center for Medical Rehabilitation Research (NCMRR) – in conjunction with one of its P2C national resource centers, the Rehabilitation Research Resource to Enhance Clinical Trials (REACT, P2CHD086851) – convened an expert panel for a workshop, The Future of Medical Rehabilitation Clinical Trials 29-30 September 2016 in Bethesda, Maryland. (REACT is one of six NCMRR-supported national P2C research resource centers which form the Medical Rehabilitation Research Resource (MR3) Network). The interdisciplinary panel approached the workshop with a broad view of medical rehabilitation – embracing the full spectrum of interventional strategies (behavioral, device-driven, pharmaceutical, multimodal, etc) intended to treat acute or chronic conditions with the aim of improving or restoring functional status (physical, cognitive, emotional), self-sufficiency, and ultimately healthspan.

The **overarching aim** of the workshop was to examine trial design and conduct considerations viewed by the panel as essential to the success and ultimate impact of clinical trials in medical rehabilitation, and to summarize recommendations to the researcher seeking to lead such trials. Thus, the **specific aims** were to: (i) aid investigators in targeting an appropriate study design to meet the study objectives at any stage of development; (ii) offer strategies for defining key outcome measures at a given phase of translation; (iii) emphasize the value of learning from the inevitable inter-individual response heterogeneity to any intervention, to both streamline improvements in intervention design,

and facilitate development of precision rehabilitation strategies; (iv) provide proven strategies for optimizing participant recruitment and retention; (v) describe challenges and opportunities for maximizing sustainability; (v) discuss both the value and potential risks of leveraging existing and emerging technologies; and (vi) overview a number of key, practical approaches to increase the rigor and reproducibility of medical rehabilition trials (see appendix). This workshop summary and appendix encapsulate the panel's discussion and recommendations in each of these areas, with the intent of maximizing the impact of future trials. A videocast of the workshop is archived at https://www.nichd.nih.gov/about/org/ncmrr/Pages/highlights.aspx.

DESIGN CONSIDERATIONS / STAGES OF DEVELOPMENT

Types of Designs

An essential first step in developing plans for a clinical trial is to recognize the stage of development and select a concordant study design that will meet the study's aims. The typical randomized controlled trial (RCT) is a design in which participants are randomly assigned to a treatment or untreated control, and studied in parallel. Random assignment is used to mitigate bias, and the control condition is used to account for potential influential factors (on outcomes) independent of the treatment (e.g., seasonal variation, learning that improves performance on a test due to repeat exposure to the test, etc). Such standard parallel group designs are well-known and well-used, both appropriately and inappropriately. However, many other designs are suitable for rehabilitation trials. Here we consider newer or less common designs that may enable the medical rehabilitation researcher to most effectively address a primary question.

Targeted or Enrichment. Similar to some cancer trials where responses can purportedly be predicted by genetics or tumor responsiveness *ex vivo*, particular rehabilitation interventions may be suited to participants based on genetics or type of injury. These "targeted designs"[1] or enrichment designs are more efficient than classical parallel designs by selecting participants with high likelihood for

response to therapy. However, there are tradeoffs between the costs of screening and recruitment vs. a design to produce improved results with a purportedly more responsive population. Some considerations should be: (i) accuracy of identifying the responsive subgroup; (ii) differential effect of the proposed treatment in the responsive subgroup, and (iii) costs of screening and resultant sample sizes. When the projected differential between the response to therapy in the target group vs. non-target group is great and the cost of the therapy is consequential, then such a design can lead to smaller sample sizes and improve efficiency. When screening cost is high, the benefit of the target design requires a large difference in efficacy between the target and non-target group, and when the proportion with the target exceeds 50%, the benefits of a target design diminish rapidly, particularly if there is some responsiveness by the non-target group.

Adaptive. Adaptive designs are popular, but have numerous definitional interpretations. Adaptive designs typically imply modifying sample size and/or dropping treatment arms based on information acquired. Such designs examine futility or dropping a treatment;[2] declare effectiveness or efficacy at the interim time point;[3] and adjust the sample size to achieve the expected result.[4] With the latter case, there are generally two approaches: (i) adjust the sample size based on design assumptions and do not examine treatment differences; or (ii) examine the actual differences and increase sample size if necessary to achieve the power to reject the original null hypothesis. There is no statistical penalty in the first approach since adjustments are based on the assumptions of the trial; however, the logistics and analysis strategy must be carefully planned in advance. With the second approach, the planning for interim sample size reassessment – which uses actual differences between groups and may require adjustments to the Type I errors – is extensive and requires careful decision-making about unmasking data, i.e. who can see which results.

Traditional phase I, II and III designs are adaptive designs in that between each phase, adjustments can be made (and usually are), but they are not seamless. Today's terminology implies that

changes are made as data become available. Alternatively, designs can be adaptive in their randomization, dynamically balancing assignments based on what has happened to date. While useful in achieving balance on known factors, these can also lead to critical imbalances on a variety of factors and should be considered carefully before implementation. Simulation based on existing datasets may help avoid situations where implementation of these procedures is actually harmful.

While there are numerous approaches, one central principle applicable across all adaptive designs is the importance of extensive planning. Key considerations include: (i) Who will examine the data? (ii) How will the decision to increase sample size be made? (iii) How will the decision to drop or add a treatment arm be made, and who will make it? (iv) What are the statistical implications of examining the data in terms of Type I error and power to make the correct decision? and (v) Once a decision is made, what implications exist for: participants, investigators, the sponsor, Data and Safety Monitoring Board (DSMB) (if applicable), and Institutional Review Boards (IRBs).

One prerequisite of adaptive designs is having all data entered and adjudicated; this necessitates orchestration, and there are several caveats: (i) additional time and extra pressure on data management for complete and accurate data; (ii) timing of analyses and the size of the interim sample used to adjust the overall sample size; (iii) analyzing data too soon with too small a sample size can lead to false positives, or to increasing the sample size when it wouldn't otherwise be necessary; and (iv) there is usually a cap on expansion of sample size for practical and financial reasons and therefore this is not a panacea for lack of effectiveness.

Sequential, Multiple Assignment, Randomized Trial (SMART). Another design, the sequential, multiple assignment, randomized trial (SMART), seeks to improve treatment paradigms for providers and participants. SMART designs are special cases of adaptive designs appropriate for chronic conditions where treatments work, but may require variation over time. SMART designs are often implemented in mental health trials where treatments are switched over time. These designs generally re-randomize

non-responders to alternative treatments and are appropriate when there is high heterogeneity in treatment responses both within and among participants. These designs should focus on the most important primary hypotheses, as powering a study for every potential pattern of treatment is impractical. Outcomes are usually binary, indicating success or failure with the intervention; for example the proportion successful after the first-line treatment. Subsequent patterns of treatment failures may emerge and, while understanding them could be important, powering the trial for these subsequent successes and failures would inevitably lead to excessively large sample sizes. Secondary questions further develop the adaptive intervention and take advantage of sequential randomization to eliminate confounding.

Multimodal Interventions

Often, two interventions are likely to be associated with positive benefits, and there is sometimes value in combining them in clinical trials. For example, clinical trials of biologic and pharmacologic interventions in spinal cord injury (SCI) often receive a great deal of media attention. However, a systematic review of these interventions indicates the strongest evidence for efficacy in multimodal interventions that include a physical rehabilitation component.[5] Conversely, combining interventions sometimes results in outcomes that are less beneficial than interventions applied independently. An example of deleterious interaction, from the SCI literature, is the interaction between monosialic ganglioside and methylprednisolone. In the U.S., methylprednisolone was once the widely accepted standard of care based on evidence that it reduced lesion volume. Pre-clinical studies suggested that monosialic ganglioside could improve neurological recovery. However, one study identified a negative interaction wherein monosialic ganglioside blocked the effect of methylprednisolone.[6] Consequently, in human trials it was deemed necessary to delay administration of monosialic ganglioside, possibly decreasing its value.[7] Generally, studies assessing multimodal interventions are most useful when: (i) the effects of each intervention have been well-characterized in

isolation; (ii) effects have been characterized in the study population of interest; (iii) there is theoretical or mechanistic reason to believe there will be synergism and effects will be cumulative; and (iv) there is no evidence for a negative interaction between interventions.

Control or Comparison Groups

The experimental rigor of testing a treatment in a clinical trial is strengthened by the inclusion of a comparator. Depending on the design, this may either be a comparison group or a control group. A **comparison group** to the experimental group, on the outcome of interest, is not selected randomly and does not receive the intervention that is being investigated. In contrast, a **control group** is comprised of individuals who could have been part of the experimental group, but through random assignment were allocated to control. Whether a comparison group or truly randomized control group is applied must be carefully considered in trial design; weighing the pros and cons of each. It is sometimes desirable and appropriate to utilize a comparison group for practical or other reasons, but one must be aware of potential biases that can be introduced if the comparison group is a "convenience sample" (e.g., participants who could not be randomized to intervention for practical reasons, such as driving distance to the intervention facility).

Numerous factors may influence the outcome of interest, and accordingly it is important to isolate the "active ingredient"— the component(s) of the intervention that is(are) thought to be directly responsible for the effects on the outcome(s) of interest — so the true value of the intervention can be discerned. This is why a control or comparison group that is not engaged in the study, other than for testing sessions, is not acceptable, as this does not control for possible confounding effects of interactions that may influence behavior, attitudes, perceptions, and outcomes. Even participant expectations are known to influence outcomes including, (i) placebo effect, wherein outcomes arise from subject beliefs about the treatment rather than the treatment itself; (ii) Hawthorne effect, wherein

subjects alter their behavior as a consequence of being observed; and (iii) Pygmalion effect, wherein subjects perform at the level that they believe others expect of them.

Study outcomes are most robust when the control or comparison group is actively engaged (i.e. with a placebo intervention) to the same extent as the experimental group, with the only difference between groups being the active ingredient under study. However, even when the active ingredient is well-isolated, there may be factors that confound outcomes, for example: (i) Was the dose sufficient? (ii) Did subjects attend all sessions? (iii) Were subjects immersed and engaged? (iv) Did subjects develop skill? (v) Did subjects use or practice the new skill outside of the training sessions? Accordingly, defining the control condition is among the most critical aspects of study design.

Alternatives to Typical Randomized Control Group. Alternatives to the typical RCT design include delayed-intervention, crossover, run-in (or wash-in), and N of 1 randomized designs.[8] The major distinction of these designs, from the randomized control, is that all subjects eventually receive the experimental intervention. These alternatives can be particularly attractive in the advanced stages of medical rehabilitation trials, when prohibiting a promising experimental treatment may be viewed by some as unethical and/or a major road-block to recruitment. ***Delayed-intervention***. Subjects are randomly assigned to an immediate-intervention group or a delayed-intervention group. The delayed group is tested at two or more timepoints prior to receiving the intervention. These test-retest measures provide control data to which the outcomes of the immediate-intervention group can be compared. As with the classic RCT, this design is strongest when the delayed-intervention group is engaged in a placebo intervention. Some considerations are that enrollment may suffer if subjects are unwilling to wait, and the delayed-intervention group may be at higher risk for dropout. ***Crossover***. This typically involves two periods with two interventions (active; placebo) although some designs have three periods. The order in which the subjects participate in each period is randomized. This design can be highly efficient as each subject serves as their own control, thereby accounting for inter-subject

variability. The design is suited for studies of symptom control (e.g., pain), however it is not appropriate for interventions that resolve the health condition. When there is a possibility of persistence of effects (i.e. carryover), a washout period is required between intervention periods, and it can be difficult to estimate the duration of the washout period needed to eliminate carryover effects. Whereas it is possible to test for carryover effects, these tests are not powerful with small sample sizes. ***Run-in (wash-in)***. Here all subjects participate in an initial period wherein they are engaged in a placebo intervention. This design is particularly valuable for study populations that have been inactive, in whom any intervention is likely to result in change. The run-in is useful for assessing stability of baseline measures, or trends in change associated with a placebo intervention. The design is most effective when multiple baseline measures are obtained in both the wash-in and experimental intervention periods.[9,10] ***N of 1 randomized***. This is a type of single-subject design wherein there are repeated observations across time in a single subject, with intervention effects being reversible upon withdrawal of the intervention. Typically, there are different levels of one intervention (AB or ABA; the latter referred to as a "reversal design") with one outcome measure of interest; although there are variations on this approach. N of 1 studies typically have a baseline, intervention, and post-intervention period, each with at least two observation/measurement timepoints.[11] This form of the N of 1 study design is most robust when the condition being addressed is chronic/stable, and when the effect of the intervention is rapid so that optimal treatment duration is achieved within the study timeframe.[12] Another type of N of 1 design is the multiple-baseline design,[13] wherein measurements are observed at the individual subject level but comparisons are made across multiple subjects. This approach represents a variation on the delayed-intervention design, and the delay period is different across subjects. Treatment effects are indicated by similar responses across subjects in the baseline (control) period, and in the intervention period. The multiple baseline design is valuable when the intervention effects are not reversible, or in situations wherein the intervention should not be withdrawn.

Limitations of Typical RCTs

Typical RCTs are deployed widely and justifiably to assess intervention efficacy, but are not without limitations. **Cost / Efficiency**. RCTs can be expensive and inefficient.[14] For example, the average cost of taking a drug from bench to bedside was $500-800 million in 2007, with RCTs accounting for 60% of the total cost. By 2013 average cost ballooned to $1.39 billion, with steady increases on the horizon.[15,16] High RCT costs make the US less competitive worldwide, and much of the cost burden finds its way to consumers in the form of higher cost of therapy. **Recruitment / Retention**. Challenges of recruitment/retention include: (i) stringent inclusion/exclusion criteria limiting populations; (ii) need for participants to attend specialized research settings that ensure the collection of standardized data – differing widely from real world environments[17.]; (iii) difficulties in recruiting adequate numbers of participants in a timely manner; and (iv) lack of broad clinician participation.[14] **Generalizability**. Typical RCTs examine whether an intervention works under ideal circumstances, using strict protocols often with placebo run-ins, in selected populations, with tightly controlled follow-up assessment, placebo, and methods to encourage high adherence.[18] *Typical RCTs therefore have very high internal validity, but can suffer from low generalizability*.[19,20] **Safety**. While a premium is placed on the monitoring of safety (e.g., adverse events (AEs)) in RCTs, adequate assessment of AEs can be influenced by small event numbers and short follow-up duration, healthy-person bias, absence of important subgroups, use of surrogate endpoints rather than clinical outcomes, and use of placebo control.

Pragmatic Clinical Trials

Some shortcomings of typical RCTs might be overcome by "pragmatic" clinical trials (PCTs), sometimes called large "simple" trials.[21] PCTs are randomized effectiveness trials that enroll large numbers of participants, have simplified protocols, and measure participant-centered outcomes. These trials investigate the effectiveness of approved interventions, and the FDA requirements for collection of safety data are less strict resulting in less investigator burden and lowers trial costs.[22] PCTs can use

existing databases and platforms – e.g., administrative claims databases, electronic health record (EHR) data, and PCORnet – that facilitate recruitment and outcome ascertainment, thereby further reducing costs.

In contrast to RCTs which often examine intervention ***efficacy*** under ideal circumstances, PCTs examine the value of an intervention compared with other existing interventions under usual clinical circumstances (i.e. ***effectiveness***). The distinction between PCTs and traditional RCTs is not a true dichotomy. Trial design lies along a continuum across a number of different dimensions; thus it may be useful during the design phase to leverage the PRagmatic-Explanatory Continuum Indicator Summary (PRECIS).[23] The PRECIS tool yields a wheel or spider diagram to illustrate where a given trial design lies on the pragmatic/explanatory continuum, based on weighing several key domains. A recent revision of the original PRECIS, called PRECIS-2, weighs nine domains: eligibility, recruitment, setting, organization, flexibility of intervention delivery, flexibility of adherence, follow-up, primary outcome, and primary analysis.[24]

There are certainly challenges and barriers to all trials along this continuum. In an effort to identify and alleviate barriers to conducting PCTs, the perspectives of various stakeholders (potential participants, physicians, researchers/study administrators, and policymakers) were collected on issues such as site and participant recruitment, consent and randomizations, study follow-up, and outcomes assessment.[25] Practice-based research networks emerged as a way to encourage more clinical practices in the community to become involved as clinical trial sites. Informatics was also identified as critical for improving efficiency, for example, deploying electronic informed consent or linking a participants' trial outcomes to their own health data to validate PCT findings.

Ultimately, the panel strongly encourages each investigative team to recognize their stage of development and select a concordant study design that will most effectively address the primary question; providing the field with key information that will enable advancement to the next phase.

OUTCOME MEASURES FOR DIFFERENT PHASES OF TRANSLATION

The translational process in medical rehabilitation research rarely follows the phase I, phase II, and phase III trial sequence typical of drug-only testing. In rehabilitation research, the process may begin with treatment ideas derived from other patient populations, clinical observations, and/or natural history studies – or with the traditional generation of ideas based on studies in tissue or animal models – with the ultimate goal of developing a treatment that is efficacious on a selected outcome measure. However, there is not always one clearly defined outcome for all phases of translation. On the other hand, at any given translational step, medical rehabilitation investigators should remain cognizant of the International Classification of Functioning, Disability and Health (ICF) – established by the World Health Organization (WHO) and endorsed by all WHO member states in 2001 – as the international standard for describing and measuring health and disability.[26]

Rehabilitation study designs and the sequence in which different interventions are explored may differ substantially from the standard approaches in drug-only trials. For example, an early proof-of-concept rehabilitation trial in TBI might need an untreated control group because of participant variability and potential for natural recovery; behavioral treatments already in widespread clinical use may be studied in earlier phases to understand their mechanisms; or trials of behavioral treatments may require multiple iterations within "Phase II" since optimizing potency may not be guided by straightforward physiologic factors. Exploratory studies that are not scaled-down versions of an efficacy study are often very important, and they should explore key details that could derail a larger trial. Investigators conducting Phase II trials further need clear go/no go decision rules for a Phase III trial.

Clinical "effectiveness" in rehabilitation research depends not only on the efficacy of the treatment, but on the participant's constellation of impairments and abilities. Two classes of theory are relevant to rehabilitation research translation, as it grapples with the complexity of restoring functioning: **Treatment theory** – A class of theories that postulate how a therapy's active ingredients

impact a specific aspect of functioning, via a mechanism of action;[27-30] **Enablement theory** – Theories that postulate the distal or "downstream" functional changes that will result from change in a specific aspect of functioning, depending on the pattern of coexisting deficits and strengths.[28-30]

In many rehabilitation treatments, the mechanism of action is not precisely known and this renders treatment theory difficult to apply initially. As a result, some early studies might need to use multiple outcome measures to determine the changes produced by the active ingredients. Once the treatment theory is able to define the specific functional change that will result directly from the treatment, early proof-of-concept studies should use outcome measures of this target function that can be linked back to mechanism(s). Later phases of research that seek to explore the more macro impact of the treatment on downstream function may select larger or more distal measures of treatment outcome. But no matter how potent the treatment of interest, it will be predicted to have important downstream effects only if: (i) it is given to participants whose downstream deficits are solely or predominantly due to a deficit in the treated function; (ii) it is combined with treatments for other functional areas that also contribute to the downstream functional deficits; *or* (iii) a different treatment is selected that more directly targets the downstream entity (e.g., an assistive device rather than exercises and training contributing to improved walking; or a supported work program, rather than seeking to improve cognitive, motor, and behavioral skills contributing to employment). The challenge is to know when the question of interest can be best addressed by treatment theory vs. enablement theory. The panel recommends the investigative team give this due consideration in the earliest stages of trial design.

INTER-INDIVIDUAL RESPONSE HETEROGENEITY

The goal of any intervention trial is to induce favorable changes in participants – e.g., physiological adaptations, attenuated pathophysiology, improved symptom management – that result in meaningful health benefits and/or functional improvements. The success or failure of a trial is therefore

303 based on whether changes in the group mean of a primary outcome are statistically and clinically
304 significant (and different from control). However, no intervention impacts all participants equally, and
305 the inter-individual response heterogeneity can be informative. The traditional approach of focusing on
306 group means fails to recognize the value in exploring the range of low to high responders. This often
307 overlooked variance can reveal important predictors of differential responsiveness, lead to
308 improvements in intervention design, and facilitate development of precision rehabilitation strategies.
309 For example, the Bamman group leveraged inter-individual response heterogeneity during trials of
310 exercise rehabilitation to reverse muscle atrophy and compromised neuromuscular function in older
311 adults to: (i) identify cellular and molecular indices of responsiveness[31-35]; (ii) conduct a follow-up dose-
312 response trial aimed to optimize the intervention prescription by minimizing the poor responder rate[36];
313 and (iii) leverage this optimal intervention prescription – and the underlying potential mechanisms
314 inhibiting responsiveness – in a subsequent double-blind, placebo-controlled, exercise-drug interaction
315 trial with the goal of further minimizing poor responder rate[37].

316 **Input Factors**

317 Numerous modifiable (e.g., comorbidities, functional capacity, diet, medications, physical
318 activity, sleep) or non-modifiable (e.g., age, gender, genotype, race/ethnicity, disease stage) factors can
319 influence response heterogeneity. For example, it is well-recognized that aging influences intervention
320 efficacy in a number of domains and, regarding pediatric medical rehabilitation approaches, there are
321 obvious biological and practical reasons that limit translatability of an intervention tested only in adults.
322 And for medical rehabilitation trials, there may be additional influential input factors to consider in
323 specific populations: **(i) Traumatic injury** (e.g., SCI, TBI, stroke, fracture) such as site of injury and
324 diagnosis, and duration of time since acute injury. For example, remarkable response heterogeneity was
325 noted recently in a stroke rehabilitation trial that was in part explained by the time elapsed between
326 each particiant's stroke event and the onset of the tested rehabilitation intervention.[38] **(ii) Chronic**

disease such as stage and duration of disease. For example, feasibility of, and individual responsiveness to, a rehabilitation intervention in Parkinson's disease may be dramatically influenced by Hoen and Yahr disease stage (1-5) of each participant, which is the impetus for investigators tightening inclusion criteria based on disease stage[39,40]. **(iii) Post-surgical rehabilitation** such as mode of surgery and structures affected. As an example, some orthopaedists perform total hip arthroplasty via an anterior surgical approach, while others take a posterior approach; the specific skeletal muscles and other support structures affected are entirely different, and may influence both rehabilitation strategy and responsiveness.

Rolling Factors

Several ongoing factors during a trial can substantially affect individual responsiveness, ranging from: **(i) Dynamic changes in molecular profiles** (e.g., transcriptome, epigenome, proteome, metabolome) through **(ii) Behaviors** (e.g., adherence/compliance to the treatment, or changes in behaviors external to the treatment such as free-living physical activity, diet, medications, etc.). Individual differences in molecular responses can be quite informative, and may help "personalize" treatments, whereas individual differences in behaviors can introduce significant layers of complexity, particularly in intent-to-treat designs, where variability in these behaviors may be wide-ranging.

Design and Analysis Considerations

Investigators are encouraged to embrace the inescapability of inter-individual response heterogeneity by: (i) considering its potential impact in trial design; (ii) maximizing data yield to better understand it; (iii) minimizing extraneous influential factors where appropriate; and (iv) controlling or monitoring behaviors and other influential factors during the trial to the degree possible. The latter requires a fine balance – and essential decision-making – in the trial design stage between treatment fidelity (e.g., efficacy) and real-world translatability (i.e. pragmatism). There are several statistical approaches one can apply to understand response heterogeneity (e.g., posthoc K-means cluster analysis

of a primary outcome with subsequent cluster comparisons of possible influential factors[31,35]). A priori stratification or posthoc covariate analysis can be leveraged for likely ***input factors***; however, there are risks in over-using both approaches, including reduced statistical power and the potential erroneous assumption that a given input factor influences all participants in each "bin" fairly equally. Regardless of the approach, modeling and exploring inter-individual response heterogeneity can substantially increase the innovation and impact of any rehabilitation trial, along with yielding invaluable data and resources that can be shared to advance the science of precision rehabilitation (e.g., NICHD Data and Specimen Hub https://dash.nichd.nih.gov/ as a centralized resource where researchers can store and access de-identified data from NICHD-funded research studies for secondary research).

OPTIMIZING RECRUITMENT AND RETENTION

While identifying and developing an optimal trial design is the important first step, the trial's success and potential impact ultimately hinge on participant recruitment and retention, which have proven to be among the greatest challenges in conducting successful clinical trials, and are therefore under intensified scrutiny by modern trial sponsors and monitoring boards. Recognizing key challenges up-front is essential, in order to proactively adopt strategies for success.

Barriers

Successful recruitment of participants is one of the most challenging aspects of conducting clinical trials[41-43] and there are several known barriers. **Sociocultural issues**. Health beliefs and life priorities, socioeconomic status, and level of fear or mistrust of research are the most cited barriers to participation, especially among minorities.[42] **Referral healthcare providers**. Providers often serve as the gate keepers of potential participants[41-43] and may fear: (i) loss of control over what happens if patients participate in a trial; (ii) the legal liability of referring patients to a study that might harm them; (iii) the uncertainty about how to explain a clinical trial to a potential participant; or (iv) lack of information about participant progress during a trial. **Study design.**[41,42] Complex consent forms, participant concerns

about being in a control group, and the time and complexity required for participation are all potential barriers. The costs of rehabilitation in RCTs may limit sample size, and/or require strict inclusion and exclusion criteria to increase control. Therefore, Investigators must be aware that more homogeneous samples can lead to less generalizable results; although in some cases (e.g. early, exploratory trials), maximizing homogeneity may be warranted. **Communication**. 90% of participants desire to know trial results, but only 7% receive that information.[44] This lack of communication reflects poorly on the research enterprise, and may persuade participants not to enroll in another trial or discourage others.

Strategies

Investigators often focus on recruitment, but overlook strategies to maximize retention. Investigators can prevent some loss-to-follow up by using appropriate exclusion criteria, but maintaining participation during the trial requires resources (e.g., non-monetary incentives, assistance with transportation or child care)[45]. Staff must attempt to recover participants who miss appointments through case management and an open door policy to encourage return.

Navigators. An innovative approach for enhancing recruitment and retention is the use of navigators. Navigators could assist under-resourced and minority participants who are reluctant to enroll by addressing barriers to enrollment such as fear and mistrust. Navigators can be effective in enhancing retention by providing essential social support[46]. For example, in oncology therapeutic trials, minority enrollment and retention rates are higher among participants who receive navigator support.[47]

In summary, recruitment and retention should be approached in a scientific fashion leveraging evidence-based design, methodology, regulations and ethical principles. Importantly, recruitment/retention efforts should not introduce bias into the study. This concern is reduced by a well-designed recruitment and retention plan developed with rigor equal to the design of the clinical trial.

SUSTAINABILITY

One of the primary goals of rehabilitation research is the development, design, and delivery of interventions that have lasting effects and placement (i.e., sustainability); this is a key premise underlying effectiveness. Sustainability involves the maintenance or durability of intervention effects and programs over time. For example, one may study sustainability of changes in a behavior (e.g., diet) and the durability of its consequences (e.g., blood glucose regulation) over a prolonged time period, particularly after the formal cessation of an intervention. One might further test maintenance and durability of an intervention program itself upon cessation of a focal research study (e.g., community-based exercise rehabilitation program).

Sustainability of Intervention Effects (on an Individual Basis). There are many considerations when designing interventions that target sustainability. For example, one would not expect that outcomes of behavior change would be maintained, if the behavior change itself were not maintained. This requires behavioral interventions wherein participants acquire the skills and techniques necessary for sustained behavior change over time. Such behavioral interventions often are based on theory (e.g., Social Cognitive Theory (SCT)[48] or Theory of Planned Behavior[49]), and integrate educational models and principles for teaching persons self-regulatory skills for action regulation that align with assumptions and principles of theory.[50] For example, based on SCT, one might develop modules around self-monitoring and goal setting. For intervention effects to be stable and durable, the interventions must be initiated with sustainable behavior change in mind. Such designs capture formative information on strengths and weaknesses of study design, including the intervention itself, from the perspectives of the participant and research team, and are critical for refining and improving Phase II through Phase IV trials. Sustainable intervention effects on behaviors and outcomes must further leverage the input and expertise of behavioral medicine. Motl and colleagues recently published on the design and evaluation

of a feasibility study involving an exercise-training intervention based on SCT that integrates expertise in behavioral medicine for long-term change in multiple sclerosis.[51,52]

Sustainability of an Intervention Program. The other aspect of sustainability involves the durability of an intervention program after cessation of a research study. This involves identifying the challenges and adopting the strategies for moving from Phase I through Phase IV trials. One central feature involves the integration of clinical and community partners (e.g., health systems; community facilities) in the initial development, design, and delivery of the intervention; this is key for lasting placement of the intervention within the proper clinical and community context. Other important aspects for maximizing durability involve incorporation of participant options and leveraging payment/reimbursement plans from payers. One example of an evidence based rehabilitative exercise program that adopted this approach is the Strength After Breast Cancer program.[53] This physical therapy-based intervention is translated from the community-based Physical Activity and Lymphedema Trial.[54,55] The program itself is now covered by third party payers, and training to prepare physical therapists for program delivery is commercially available (http://klosetraining.com/course/online/strength-abc/). Overall, the panel encourages investigative teams to consider strategies for studying and promoting long-term sustainability – for both the individual and overall program itself – in order to maximize the ultimate impact of any given clinical trial or series of trials across the stages of development.

LEVERAGING ADVANCES IN TECHNOLOGY

Emerging technologies may be utilized in rehabilitation clinical trials as interventions to be evaluated or as facilitators of clinical trials. Advances in the basic sciences and engineering have led to tremendous technological innovations directed at ameliorating the health and disability burden of persons with musculoskeletal and neurological conditions. However, the benefits of these technologies at the level of activities and participation are far from clear. Moreover, new advanced technologies can

be costly and without a clear societal benefit over usual care. The emergence of population health and global risk models present both opportunities and challenges. Comparative effectiveness and pragmatic trials of new technologies relative to usual care are encouraged. Such technologies may translate to outcomes comparable or superior to usual care, but avoid the need for costly healthcare infrastructure and personnel. Such trials should take into account the full burden of societal cost rather than just the cost of the technology. Nevertheless, because the changing healthcare landscape places such a high premium on cost containment, the escalating cost of technology based interventions for a relatively small population is a serious threat to leveraging advanced technologies in rehabilitation clinical trials.

Advances in monitoring, communications and sensor technology are poised to have significant impact on the conduct of rehabilitation trials by reducing costs, improving treatment fidelity and facilitating effectiveness trials that measure high-level outcomes. Costs may be reduced through the implementation of electronic data collection systems that collect data faster and without the need for research staff intervention and travel. At the same time, accuracy and adherence are likely to improve because these electronic systems are less intrusive and can implement quality safeguards. Perhaps the most interesting and promising benefits of integrating technology into rehabilitation trials concern the types of outcomes that might be measured and the advanced trial designs that may be implemented. These technologies may allow us to evaluate clinical outcomes at the level of capacity, performance and participation. Or, a trial could be designed so that the intervention is responsive to the participant's performance and context for ecological momentary assessments. Still, caution should be exercised in using new technology. Study procedures should be carefully designed and beta tested, and sensors and systems need to be validated prior to their implementation.

SUMMARY

With continual advancement in medical and surgical care extending long-term survival, the medical rehabilitation field finds itself in a unique and challenging position of advancing interventions to extend healthspan for both chronic and traumatic disease. This mission requires rigorous clinical trials that are sufficiently innovative to exert a meaningful, long-term impact on the field by yielding effective and translatable models and approaches to patient care. The 2016 NCMRR / REACT workshop and this accompanying summary are intended to provide some guidance to researchers seeking to help achieve the mission – from design considerations, to day-to-day rigor in trial operations, to innovations. Together it is our goal to maximize the impact of future trials in medical rehabilitation.

REFERENCES

1. Simon R, Maitournam A. Evaluating the Efficiency of Targeted Designs for Randomized Clinical Trials. *Clinical Cancer Research.* 2004;10(20):6759-6763.
2. Cutter G, Kappos L. Clinical trials in multiple sclerosis. *Handbook of clinical neurology.* 2014;122:445-453.
3. Zhang X, Cutter G. Bayesian interim analysis in clinical trials. *Contemp Clin Trials.* 2008;29(5):751-755.
4. Bhatt DL, Mehta C. Adaptive Designs for Clinical Trials. *New England Journal of Medicine.* 2016;375(1):65-74.
5. Gomes-Osman J, Cortes M, Guest J, Pascual-Leone A. A Systematic Review of Experimental Strategies Aimed at Improving Motor Function after Acute and Chronic Spinal Cord Injury. *Journal of neurotrauma.* 2016;33(5):425-438.
6. Constantini S, Young W. The effects of methylprednisolone and the ganglioside GM1 on acute spinal cord injury in rats. *Journal of neurosurgery.* 1994;80(1):97-111.
7. Geisler FH, Dorsey FC, Coleman WP. Recovery of motor function after spinal-cord injury--a randomized, placebo-controlled trial with GM-1 ganglioside. *N Engl J Med.* 1991;324(26):1829-1838.
8. Backman CL, Harris SR. Case studies, single-subject research, and N of 1 randomized trials: comparisons and contrasts. *American journal of physical medicine & rehabilitation.* 1999;78(2):170-176.
9. Senn S. Are placebo run ins justified? *BMJ.* 1997;314(7088):1191.
10. Senn S, Lee S. The analysis of the AB/BA cross-over trial in the medical literature. *Pharmaceutical Statistics.* 2004;3(2):123-131.
11. Sackett DL HR, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine.* 2nd ed: Little, Brown & Co; 1991.
12. P. O. Single case designs. In: Everitt BS HD, ed. *Encyclopedia of Statistics in Behavioral Science*. Chichester: John Wiley & Sons Ltd; 2005.
13. Hawkins NG, Sanson-Fisher RW, Shakeshaft A, D'Este C, Green LW. The multiple baseline design for evaluating population-based research. *American journal of preventive medicine.* 2007;33(2):162-168.
14. Center for Information and Study on Clinical Research. www.ciscrp.org. Accessed February 8, 2017.
15. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of health economics.* 2016;47:20-33.
16. Institute of M. *Envisioning a Transformed Clinical Trials Enterprise in the United States: Establishing an Agenda for 2020: Workshop Summary.* Washington, DC: The National Academies Press; 2012.
17. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence — What Is It and What Can It Tell Us? *New England Journal of Medicine.* 2016;375(23):2293-2297.
18. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis.* 1967;20(8):637-648.
19. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology - the Essentials.* 2nd ed. Baltimore, MD: Williams & Wilkins; 1988.
20. Pincus T, Sokka T. Clinical trials in rheumatic diseases: designs and limitations. *Rheum Dis Clin North Am.* 2004;30(4):701-724, v-vi.
21. Eapen ZJ, Lauer MS, Temple RJ. The imperative of overcoming barriers to the conduct of large, simple trials. *JAMA.* 2014;311(14):1397-1398.

22. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Determining the Extent of Safety Data Collection Needed in Late-Stage Premarket and Postapproval Clinical Investigations - Guidance for Industry. 2016; http://www.fda.gov/downloads/drugs/guidances/ucm291158.pdf. Accessed February 8, 2017.
23. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol.* 2009;62(5):464-475.
24. 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.
25. Saag KG, Mohr PE, Esmail L, et al. Improving the efficiency and effectiveness of pragmatic clinical trials in older adults in the United States. *Contemp Clin Trials.* 2012;33(6):1211-1216.
26. Organization WH. International Classification of Functioning, Disability and Health. 2001; http://www.who.int/classifications/icf/en/.
27. Hart T, Ehde DM. Defining the treatment targets and active ingredients of rehabilitation: Implications for rehabilitation psychology. *Rehabil Psychol.* 2015;60(2):126-135.
28. VanHiel LR. Treatment and enablement in rehabilitation research. *Archives of physical medicine and rehabilitation.* 2014;95(1 Suppl):S88-90.
29. Whyte J. Contributions of treatment theory and enablement theory to rehabilitation research and practice. *Archives of physical medicine and rehabilitation.* 2014;95(1 Suppl):S17-23 e12.
30. Whyte J, Dijkers MP, Hart T, et al. Development of a theory-driven rehabilitation treatment taxonomy: conceptual issues. *Archives of physical medicine and rehabilitation.* 2014;95(1 Suppl):S24-32 e22.
31. Bamman MM, Petrella JK, Kim JS, Mayhew DL, Cross JM. Cluster analysis tests the importance of myogenic gene expression during myofiber hypertrophy in humans. *J Appl Physiol (1985).* 2007;102(6):2232-2239.
32. Mayhew DL, Hornberger TA, Lincoln HC, Bamman MM. Eukaryotic initiation factor 2B epsilon induces cap-dependent translation and skeletal muscle hypertrophy. *The Journal of physiology.* 2011;589(Pt 12):3023-3037.
33. Petrella JK, Kim JS, Mayhew DL, Cross JM, Bamman MM. Potent myofiber hypertrophy during resistance training in humans is associated with satellite cell-mediated myonuclear addition: a cluster analysis. *Journal of applied physiology.* 2008;104(6):1736-1742.
34. Stec MJ, Kelly NA, Many GM, Windham ST, Tuggle SC, Bamman MM. Ribosome biogenesis may augment resistance training-induced myofiber hypertrophy and is required for myotube growth in vitro. *American journal of physiology Endocrinology and metabolism.* 2016:ajpendo 00486 02015.
35. Thalacker-Mercer A, Stec M, Cui X, Cross J, Windham S, Bamman M. Cluster analysis reveals differential transcript profiles associated with resistance training-induced human skeletal muscle hypertrophy. *Physiol Genomics.* 2013;45(12):499-507.
36. Stec MJ, Thalacker-Mercer A, Mayhew DL, et al. Randomized, four-arm, dose-response clinical trial to optimize resistance exercise training for older adults with age-related muscle atrophy. *Exp Gerontol.* 2017;99:98-109.
37. Long DE, Peck BD, Martz JL, et al. Metformin to Augment Strength Training Effective Response in Seniors (MASTERS): study protocol for a randomized controlled trial. *Trials.* 2017;18(1):192.
38. Knutson JS, Gunzler DD, Wilson RD, Chae J. Contralaterally Controlled Functional Electrical Stimulation Improves Hand Dexterity in Chronic Hemiparesis: A Randomized Trial. *Stroke.* 2016;47(10):2596-2602.
39. Kelly NA, Ford MP, Standaert DG, et al. Novel, high-intensity exercise prescription improves muscle mass, mitochondrial function, and physical capacity in individuals with Parkinson's

disease. *Journal of applied physiology.* 2014;116(5):582-592.
40. Corcos DM, Robichaud JA, David FJ, et al. A two-year randomized controlled trial of progressive resistance exercise for Parkinson's disease. *Mov Disord.* 2013;28(9):1230-1240.
41. McDonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials.* 2006;7:9.
42. Kadam RA, Borde SU, Madas SA, Salvi SS, Limaye SS. Challenges in recruitment and retention of clinical trial subjects. *Perspect Clin Res.* 2016;7(3):137-143.
43. Frank G. Current Challenges In Clinical Trial Patient Recruitment and Enrollment. *SOCRA Source.* 2004;2(February):9.
44. Shalowitz DI, Miller FG. Communicating the results of clinical research to participants: attitudes, practices, and future directions. *PLoS Med.* 2008;5(5):e91.
45. Caldwell PH, Hamilton S, Tan A, Craig JC. Strategies for increasing recruitment to randomised controlled trials: systematic review. *PLoS Med.* 2010;7(11):e1000368.
46. Fouad MN, Acemgil A, Bae S, et al. Patient Navigation As a Model to Increase Participation of African Americans in Cancer Clinical Trials. *Journal of Oncology Practice.* 2016;12(6):556-563.
47. Fouad MN, Acemgil A, Bae S, et al. Patient Navigation As a Model to Increase Participation of African Americans in Cancer Clinical Trials. *J Oncol Pract.* 2016;12(6):556-563.
48. Lai B, Young HJ, Bickel CS, Motl RW, Rimmer JH. Current Trends in Exercise Intervention Research, Technology, and Behavioral Change Strategies for People With Disabilities: A Scoping Review. *American journal of physical medicine & rehabilitation.* 2017;96(10):748-761.
49. Larkin L, Kennedy N, Gallagher S. Promoting physical activity in rheumatoid arthritis: a narrative review of behaviour change theories. *Disabil Rehabil.* 2015;37(25):2359-2366.
50. Ellis T, Motl RW. Physical activity behavior change in persons with neurologic disorders: overview and examples from Parkinson disease and multiple sclerosis. *Journal of neurologic physical therapy : JNPT.* 2013;37(2):85-90.
51. Adamson BC, Learmonth YC, Kinnett-Hopkins D, Bohri M, Motl RW. Feasibility study design and methods for Project GEMS: Guidelines for Exercise in Multiple Sclerosis. *Contemp Clin Trials.* 2016;47:32-39.
52. Learmonth YC, Adamson, B. C., Kinnett-Hopkins, D., Bohri, M., & Motl, R. W. . Results of a feasibility randomized controlled study of the guidelines for exercise in multiple sclerosis project. *Contemporary Clinical Trials.* In Press.
53. Beidas RS, Paciotti B, Barg F, et al. A hybrid effectiveness-implementation trial of an evidence-based exercise intervention for breast cancer survivors. *Journal of the National Cancer Institute Monographs.* 2014;2014(50):338-345.
54. Schmitz KH, Ahmed RL, Troxel A, et al. Weight lifting in women with breast-cancer-related lymphedema. *N Engl J Med.* 2009;361(7):664-673.
55. Schmitz KH, Ahmed RL, Troxel AB, et al. Weight lifting for women at risk for breast cancer-related lymphedema: a randomized trial. *Jama.* 2010;304(24):2699-2705.

Title: Medical Rehabilitation: Guidelines to Advance the Field with High-Impact Clinical Trials

ONLINE APPENDIX

A Practical Guide: Recommended Operational Procedures to Maximize Trial Rigor and Impact

SUMMARY

The purpose of this appendix is to summarize key recommendations for day-to-day trial conduct, with the ultimate goal of maximizing trial rigor and impact. Investigators are encouraged to consider these recommendations when: (i) identifying key staff (coordinator, interventionists, assessors); (ii) establishing operating procedures and the optimal plan for data and safety monitoring; (iii) demonstrating compliance with NIH requirements for rigor, reproducibility, and transparency; and (iv), in the case of multi-site trials, standardizing human participant protections and oversight via a shared IRB.

KEY STAFF AND OPERATING PROCEDURES

Key Staff

Coordinators. Clinical Research Coordinators (CRCs) are trained research professionals working under the leadership of the Principal Investigator (PI). They are responsible for the ethical conduct of clinical trials using Good Clinical Practices (GCPs) and International Council on Harmonization (ICH) Guidelines.[1] The primary responsibility is the protection of human subjects. CRCs may be responsible for subject recruitment and retention, as well as maintaining regulatory files, writing informed consents, managing IRB submissions, negotiating and preparing budgets/contracts, creating case report forms (CRFs), reporting AEs, coordinating monitoring activities and study close-out.[2] Qualifications of a CRC should include GCPs, ICH guidelines, Human Subjects, Research Integrity, and Sponsor Training. Due to the ongoing flux of regulations and guidelines, CRCs should obtain clinical research certification that requires continued education to maintain certification.[3]

Outcomes Assessor(s). An outcome assessor is responsible for the administration of reliable and accurate outcome assessments according to protocol as described in the Manual of Procedures (MOP). Optimally, the same assessor should administer assessments per scheduled time point for any given participant. Coordination between the CRC and assessor is important for trial efficiency (e.g., scheduling,

tracking). Assessors who work on a FFS (fee for service) basis for which payment is linked to delivery of data (e.g., CRF) is optimal for data completeness and budgeting purposes. Investigators should include the direct cost for pre-enrollment training and standardization of assessors in the budget, especially if adopting a FFS payment plan. **Standardization across multiple assessors**. An important part of quality control is a rigorous standardization procedure for test/evaluation administration. Prior to assessment of an enrolled participant, assessors should submit material (e.g., written CRF, knowledge test, visual/video) that provides evidence of didactic knowledge and correct administration (e.g., 90% criterion) of all primary and in some cases secondary outcomes. This should be staged with pilot participants who meet the same (or similar) inclusion criteria as those targeted for the trial. Constructive feedback and status (i.e., meets/does not meet 90% criteria) should be provided by the clinical research team/administrator in a timely manner. Only those having met standardization criteria with demonstrated competence should be eligible to administer outcome tests. Re-standardization should be performed more frequently in the beginning (e.g., monthly) and at least every 6 months thereafter to maintain consistency. New assessors who join the team after trial initiation must provide the same level of evidence of competency before being eligible to serve as outcome assessors.

Interventionists. Rigorous standardization of intervention administration is important for quality control. Qualifications for interventionists should be specified in accord with the specific skills needed to carry out the intervention. For example, is board or specialty certification required? How much experience is required? Qualifications should align with the nature of the clinical trial design (e.g., efficacy vs effectiveness). **Efficacy trials (research staff) vs. effectiveness trials (clinic staff).** Who should deliver the intervention? If efficacy is the goal (i.e. does this work under optimal, controlled conditions with expert, trained and standardized clinicians), research staff should deliver the intervention. If effectiveness is the goal (i.e. does this work in the actual clinical environment with all the noise/variability in expertise, scheduling, etc.), clinical staff should deliver the intervention.[4]

Operating Procedures

Masking or Blinding. For single and double-blind trials, steps should be taken to assure that assessors are masked to group assignment. The CRC should keep the randomization secure so assignments are not known to the assessor. Every effort should be made to ensure the evaluations and interventions are conducted in different locations to help prevent accidental unblinding. The purpose of blinding the assessor is explained to the participant verbally and in writing. The CRC will be responsible to remind the participant and family members to help keep the assessor blinded before each encounter. Hiring assessors as independent contractors, without affiliation with the recruitment site, can be an effective strategy. There should be a mechanism for documenting and tracking unblinding which might occur during an evaluation session.

Rigorous treatment fidelity assessment. Both participants and investigators play key roles in determining treatment fidelity. **Participant adherence**. Adherence to treatment is often studied on the side of the participant and in recent years there have been many attempts to involve participants at all levels to maximize the success of rehabilitation.[5] Participant adherence must be rigorously monitored and tested throughout all stages of implementation. **Researcher/clinician fidelity (to prevent drift)**. In addition to participant adherence, it is essential that investigators establish and monitor adherence to key components of interventions by those delivering interventions (i.e., fidelity for treatment). Cost-efficient methods to document treatment fidelity should be established, including built-in warnings when deviation from the original model could potentially render the intervention ineffective (or iatrogenic). To the extent that rigorous clinical trials are required to include methods for monitoring intervention delivery, these methods should be published as part of the trial's protocol so they can be used in real-world rehabilitation settings. Carefully constructed exit interview/surveys or valid and reliable instruments (e.g., Health Care Climate Questionnaire) can be used with participants and/or clinicians to confirm intended effects of the intervention.[6,7]

Standard Operating Procedures (SOPs). The ICH defines a SOP as "Detailed, written instructions to achieve uniformity of the performance of a specific function." (ICH GCP 1.55). Each SOP details how a specific test, measure, or function is to be performed, and can be applied to any research study. The Manual of Procedures (MOP), on the other hand, is specifically written for a particular study which incorporates elements of SOPs. The MOP is a dynamic document updated throughout the study to record/track the impact of protocol amendments on the study procedures, and to document refinement of procedures; with each new update previous versions should be archived. A typical table of contents includes the following: Protocol, Staff Roster, Study Organization and Responsibilities, Training Plan, Communication Plan, Recruitment Plan, Study Flow, Screening and Eligibility Criteria, Informed Consent and HIPAA, Randomization, Study Intervention, Blinding and Unblinding, Participant Retention, Concomitant Medications/ Treatments, Safety Reporting, Data and Safety Monitoring, Study Compliance, Data Collection and Study Forms, Data Management, Quality Control Procedures, Study Completion and Closeout Procedures, Policies, and MOP Maintenance. For multi-site coordination, the administrative core/data management center is usually responsible for setting up a centralized secure website where the MOP can be accessed, updates posted, and alerts issued.

DATA AND SAFETY MONITORING

Quality by design can optimize the veracity and completeness of data collected during a trial. First principles include: collecting only the data needed to answer the research objectives and protect human subjects; defining standards for the intervention and outcome measures in a MOP; using standardized data elements when available; providing detailed instructions for completing forms and submitting data; monitoring data quality; and training staff on procedures and protocol implementation. Increasing standardization across studies helps move the field forward and facilitates the important goal of making trials data public and easy to share.

Data Collection. Data collection systems range from paper source to electronic source. Research data captured from electronic health records can facilitate efficiency with the caveat that these data may be collected for purposes, such as billing, that may be inconsistent with those of a research study. A system that keeps research data entry closest to the source of data acquisition and can be evaluated for quality in real-time is preferred; although institution or sponsor-specific security and privacy policies are important considerations.

Data Quality. Data quality and research execution monitoring can include performing on-site monitoring to assess the rigor of the informed consent process, execution of testing, evaluation of treatment fidelity (a substantial threat to study validity in rehabilitation trials), and verification of research data against source data. Monitoring can be used to flag issues of low enrollment, retention, incomplete data, or distributions inconsistent with expected data. Evaluations should include data consistency checks at data field, form, and cross form levels. Monitoring and assessment of data quality should be conducted by senior members of the research team on a frequent, regular basis, and by an experienced, independent auditor at least semi-annually. Local and/or sponsor requirements may differ and clearly take precedence.

Data and Safety Monitoring (DSM). Trials need a DSM plan (DSMP) in place prior to study initiation. The complexity of the plan can range from the PI reviewing accumulating safety events to convening an independent DSM Board (DSMB) to review the accumulating data and advise the sponsor whether or not the study should continue as designed or terminate early due to futility, benefit, or safety. The NIH and other funding agencies require that the plan specify the: information to be monitored, frequency of monitoring, interim analysis plans, early termination guidelines, and processes for monitoring and reporting adverse events and unanticipated problems. For all Phase III and for multi-site and/or high risk Phase I or II trials, a full, independent DSMB is required. Board members may include clinician(s), a biostatistician, basic scientist and a participant advocate or medical ethicist. For

small, lower-risk trials, the PI or a DSM Committee composed of investigators may be sufficient. Recommendations related to DSM are as follows: (i) PIs and staff new to clinical trials should seek consultation from those with clinical trial expertise when constructing a plan and should capitalize on the availability of standing DSM Committees; (ii) Investigators should identify potential important adverse events in advance, systematically query for and collect these events on data forms, and routinely query for all adverse events at appropriate times over the course of the study. If participants are not asked about potential events, the events may go unreported; (iii) DSMB members must fully disclose all potential conflicts of interest; (iv) DSMBs should review accumulating data unmasked so as to appropriately balance benefit versus harm throughout the study; (v) The DSMB has a responsibility to document and report to local IRBs, the occurrence of a meeting, the attendees, and high level decisions regarding safety and trial continuation; and (vi) DSMBs serve in an advisory capacity to the Sponsor.

In medical rehabilitation, it is important to recognize that clinical trials often enroll participants with substantially compromised function coupled with heightened risk of comorbidities and other complications (e.g., SCI, stroke, neurodegenerative diseases, progressive musculoskeletal disorders). It may therefore be prudent to enlist an experienced DSMB even in small, early phase studies for which a DSMB is not required by the Sponsor.

RIGOR, REPRODUCIBILITY, AND TRANSPARENCY (R2T)

In 2014, NIH announced plans to support greater rigor, reproducibility, and transparency (R2T) in biomedical research.[8] New requirements for extramural research proposals were introduced, asking applicants to describe the premise, robust and unbiased approaches, identification of biological variables, and authentication of key biological or chemical resources. While R2T experimental design may be second nature to experienced RCT investigators, specific details (e.g., sample size estimation, randomization, blinding, participant recruitment and retention, etc) may be overlooked or under-appreciated by junior investigators and are therefore incorporated into many training programs. Other

R2T initiatives by the NIH included the development of training modules on good experimental design, guidance and resources for grant proposal reviewers, a big data initiative (the Data Discovery Index) for handling unpublished, primary data, and additional functionality in PubMed Commons to allow open discourse about published articles.

SHARED INSTITUTIONAL REVIEW BOARDS FOR MULTI-SITE TRIALS

A common obstacle faced by multi-site clinical trials is the need for multiple IRBs across the participating sites. One way to overcome this barrier is the use of a shared IRB, which can help streamline the IRB review process thereby improving efficiency and consistency. A shared IRB is different from a central IRB in that the latter provides review for multiple studies conducted by a consortium or network of institutions, whereas a shared IRB provides review for a single multi-site study. In 2015, the DHHS Notice of Proposed Rulemaking to revise the Common Rule on the Protection of Human Subjects mandated that all U.S. institutions engaged in multi-site research must have a single (shared) IRB in place.[9] This rule went into effect on May 25, 2017. The shared IRB comes with some clear practical challenges, including the need to coordinate the efforts of multiple institutions to develop reliance agreements, delineate institution-specific versus shared IRB responsibilities, standardize policies and procedures, accommodate variations in state laws, and develop a communication and cost-sharing plan.

To help facilitate the use of shared IRB review the National Center for Advancing Translational Sciences (NCATS) developed a shared IRB platform, called the Streamlined, Multisite, Accelerated Resources for Trials (SMART) IRB.[10] SMART IRB has been designed to ease common challenges and burdens associated with initiating multisite research. This includes the development of SOPs and informatics support, and master reliance agreements, all of which are currently underway. To help ensure and encourage research collaboration and harmonization of review, NCATS has developed a set of minimal requirements for authorization of institutional participation in the SMART IRB program,

168 including being a federal-wide assurance (FWA) or IRB registration institution, and having the ability to
169 meet Human Research Protection Program standards and the capacity to follow standard operation
170 procedures in support of the IRB reliance agreement.
171

REFERENCES

1. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). www.ich.org/.
2. Society of Clinical Research Associates (SOCRA). https://www.socra.org/.
3. Association of Clnical Research Professionals. https://www.acrpnet.org/.
4. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials.* 2009;10(1):37.
5. Ehde DM, Wegener ST, Williams RM, et al. Developing, testing, and sustaining rehabilitation interventions via participatory action research. *Archives of physical medicine and rehabilitation.* 2013;94(1 Suppl):S30-42.
6. Williams GC, Grow VM, Freedman ZR, Ryan RM, Deci EL. Motivational predictors of weight loss and weight-loss maintenance. *Journal of personality and social psychology.* 1996;70(1):115-126.
7. Chan DK, Lonsdale C, Ho PY, Yung PS, Chan KM. Patient Motivation and Adherence to Postsurgery Rehabilitation Exercise Recommendations: The Influence of Physiotherapists' Autonomy-Supportive Behaviors. *Archives of physical medicine and rehabilitation.* 2009;90(12):1977-1982.
8. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature.* 2014;505(7485):612-613.
9. Department of Health and Human Services. Notice of Proposed Rulemaking. *Federal Register.* 2015;80(173):53933-54061.
10. National Center for Advancing Translational Sciences. NCATS SMART IRB Reliance Platform. https://ncats.nih.gov/expertise/clinical/smartirb. Accessed February 8, 2017.