



HHS Public Access

Author manuscript

Psychol Aging. Author manuscript; available in PMC 2023 March 17.

Published in final edited form as:

Psychol Aging. 2022 February ; 37(1): 43–50. doi:10.1037/pag0000653.

Good Clinical Practice Improves Rigor and Transparency: Lessons from the ACTIVE Trial

Brad P. Taylor¹, George W. Rebok^{2,3}, Michael Marsiske¹

¹Department of Clinical and Health Psychology, University of Florida, Gainesville, FL

²Department of Mental Health, Johns Hopkins University, Baltimore, MD

³Center on Aging and Health, Johns Hopkins University, Baltimore, MD

Abstract

Clinical trials are governed by principles of good clinical practice (GCP), which can strengthen the achievement of rigor, reproducibility, and transparency in scientific research. Rigor, reproducibility, and transparency are key for producing findings with greater certainty. Clinical trials are closely supervised, often by a clinical trial coordinating center, data safety and monitoring board, and a funding agency, with policies that are a manifestation of GCP and support rigor, reproducibility, and transparency. The multi-site Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE) study is an example clinical trial of relevance to a psychology and aging audience, that utilized many protocols that can be applied to single laboratory designs, including: a manualized protocol with accompanying scientific rationale, predefined analysis plans, standardization of procedures across field sites, assurance of competence of study staff in study procedures, transparent coding/entry/transmittal of data, regular quality assurance, and open publication of data. Despite substantial resource discrepancies between the two, single laboratory studies can model the GCP principles utilized in large clinical trials to provide an excellent foundation for rigor, reproducibility, and transparency.

Keywords

Good Clinical Practice; Rigor; Reproducibility; Transparency; ACTIVE Trial

Correspondence concerning this article should be addressed to Dr. Michael Marsiske, Department of Clinical and Health Psychology, University of Florida, P.O. Box 100165, Gainesville, FL 32610-0165, United States. marsiske@ufl.edu.

No data or pre-registered data were used in the preparation of this manuscript, but readers can access data and documentation at this link for the ACTIVE trial, the parent study: <https://www.icpsr.umich.edu/web/NACDA/studies/36036>. The parent study pre-registration (ClinicalTrials.gov Identifier: NCT00298558) can be found at <https://clinicaltrials.gov/ct2/show/NCT00298558> (National Institutes of Health, 2014). The content of this manuscript was briefly summarized at the University of Florida's Research Reproducibility Virtual Conference. The concepts of using a lab manual, pre-defined analysis plans, detailed data management protocol, certification protocols and checklists, making study information and data publicly available, separating the data analysis role, and instilling quality assurance were included as part of this virtual poster presentation.

Good Clinical Practice Improves Rigor and Transparency: Lessons from the ACTIVE Study

The hypothetico-deductive cycle, a core concept in scientific practice, is premised on the idea that there will be research replication attempts (Mertens & Recker, 2020); this in turn requires that research methods in any given study be rigorous (ruling out alternative plausible explanations of phenomena; e.g., Kazdin, 2016) and transparently reproducible. In psychological and clinical sciences, however, a number of attempts to reproduce seminal findings have been unsuccessful, leading some to express concern that there is a “reproducibility crisis” in psychology (Maxwell et al., 2015; Shrout & Rodgers, 2018). Because clinical science is bound by its focus on producing and improving treatment and prevention interventions (Onken et al., 2014), it is critical that our fields move toward a more rigorous scientific agenda that can lead to evidence-based treatments. Additionally, the costs associated with poor rigor may be substantial and can limit the ability to draw conclusions about the effectiveness of an intervention and to inform public policy; following best practices in research helps ensure the scientific community or public are not misled (Simons et al., 2016). Although the concepts of rigor, reproducibility, and transparency are still being empirically tested, there is some evidence that these concepts have increased the number of null findings (Kaplan & Irvin, 2015). In this paper, we consider how one long-standing model for conducting research (“good clinical practice,” GCP, Vijayanathan & Nawawi, 2008) is typically implemented in clinical trial designs. We further suggest that these principles are implemented in consistent ways in many clinical trials and might be considered a template for investigators in individual psychology and aging laboratories to improve the rigor and transparency of their work. We will draw on one multi-site clinical trial with which we have experience, the Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE) trial (Jobe et al., 2001; Ball et al., 2002; Willis et al., 2006), to demonstrate how GCP might be concretized in laboratory settings. Throughout, we also consider how single laboratories and junior investigators could increase study conformity with GCP principles, despite not having the resources of a large multi-site trial. It is important to note that we offer the clinical trial as a best-case scenario when resources to support rigor and reproducibility have been maximized. The expectation would not be for single investigators to replicate everything that a multi-site trial does, but to incorporate features (like manualization and documentation, pre-registration, adequate training of staff, ongoing quality control and fidelity observations, open materials, and transparent methods sections) that seem feasible within resource availability.

Rigor, Reproducibility, and Transparency

According to the NIH, “the application of rigor ensures robust and unbiased experimental design, methodology, analysis, interpretation, and reporting of results” (National Institutes of Health, 2019a, para. 1). When a rigorously obtained result is further reproduced, confidence in the result is increased, permitting theoretical refinements and follow-up research. Replication, in turn, is premised on the transparency of the original research methods; reproducibility is premised on high-quality data and project management (Sullivan et al., 2019). Yet there are constraints (e.g., journal article page lengths) and self-interest considerations that often make the method descriptions of a particular study incomplete or

opaque. Transparency refers to the reporting of experimental details in sufficient detail so that other researchers may assess the proposed research and may be capable of reproducing and extending the findings (National Institutes of Health, 2019b).

The antidote to the absence of transparency is openness. Openness is defined along several dimensions that may include detailed manuals, published protocols, open data, open code, pre-registration, rigorous training/certification for all roles, and involving outside statisticians and quality control. All help to mitigate these issues by improving scientific rigor, allowing for reproducible analyses, and transparently placing everything in the public eye to ensure that people with less investment in the results have a chance to critically review all aspects of the research (e.g., Kidwell et al, 2016). Major journals and funding agencies have launched initiatives to promote, track, and reward rigor and openness (e.g., Freedman et al., 2017; Open Science Framework, 2020).

Good Clinical Practice and Clinical Trials

Interestingly, although the open science concept is relatively new, many of the principles associated with open science have long constituted standard operating practice for clinical trials – in particular, multi-site clinical trials in which standardization of practices across sites and accountability to external monitors demands a high level of rigorous documentation and transparency. Clinical trials are governed by principles of GCP (Vijayanathan & Nawawi, 2008). Because many clinical trials are often high-stakes studies where the outcome can affect treatments, systems of rigor, transparency, and openness have been “baked in” to large-scale clinical trials. World Health Organization (WHO) guidelines have attempted to codify important ethical and rigor-related principles falling under the umbrella of GCP (World Health Organization, 2002). These principles are meant to “help assure the safety, integrity, and quality of clinical trials by addressing elements related to the design, conduct, and reporting of clinical trials” (National Institutes of Health, 2017, para. 2). The principles most related to rigor and transparency include: (1; WHO Principle 2) Research involving humans should be scientifically justified and described in a clear, detailed protocol; (2; WHO Principle 6) Research involving humans should be conducted in compliance with the approved protocol; (3; WHO Principle 10) Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his or her respective task(s) and currently licensed to do so, where required; (4; WHO Principle 11) All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation, and verification; and (5; WHO Principle 14) Systems with procedures that assure the quality of every aspect of the trial should be implemented.

All clinical trials are governed by GCP principles, but multi-site trials (used to assess the generalizability of a treatment across a variety of settings) often include a governing entity, the clinical trial coordinating center, which is charged with implementing policies and oversight that are a manifestation of GCP, and are designed to support rigor, reproducibility, and transparency of the trial. There are well-articulated roles for a clinical trial coordinating center (Biswas et al., 2012; National Heart, Lung, and Blood Institute, 2011), which include (a) maintaining blind (a firewall between the data and investigators); (b) coordinating routine

communication among study Principal Investigators (PIs) and field staff; (c) coordinating manuscript proposals; (d) maintaining data entry, data cleaning, data provisioning; (e) conducting primary analyses (and replicating others) independently of the PIs; (f) creating and updating a Manual of Procedures (MOP); (g) scheduling and overseeing a Data Safety and Monitoring Board (DSMB); (h) designing and conducting/monitoring quality assurance protocols for assessment/interventions/data/scoring, and having an intervention plan when there is drift; and (i) archiving data, protocols, measures, and code. To provide some concrete exemplars of how these were implemented, we consider the clinical trial in which we were involved, ACTIVE.

Example Clinical Trial: ACTIVE

The ACTIVE trial was an NIH-funded multi-site Phase 3 clinical trial, aimed at investigating the immediate and long-term impacts of cognitive training on older adults' independence and everyday functioning (Jobe et al., 2001; Ball et al., 2002; Willis et al., 2006). Between 1998-2000, 2,802 community-residing adults aged 65-94 were enrolled from six field sites around the United States: University of Alabama at Birmingham, the Boston Hebrew Rehabilitation Center for the Aged, Indiana University School of Medicine, Johns Hopkins University, Pennsylvania State University, and Wayne State University. Participants were randomized into one of three training arms (memory, reasoning, speed of processing) or a no-contact control. As is true for most NIH-funded clinical trials, great emphasis was put on standardization of procedures across field sites, assurance of competence of study staff in study procedures, transparent coding/entry/transmittal of data, predefined analysis plans, and open publication of data. Much of the transparency of a clinical trial is achieved by three oversight bodies that interact with the steering committee: the coordinating center (when independent from field sites), the DSMB, and the funding agency itself. Below we present an example of WHO principles of GCP implemented in ACTIVE; values in parentheses after each principle name reflect WHO numbering scheme.

1. Scientific Justification (WHO Principle 2)—ACTIVE, like all clinical trials, utilized a steering committee, which served as the main governing body and comprised the PI from each field site, the PI at ACTIVE's coordinating center (New England Research Institutes; NERI), and the funding agencies' (National Institute on Aging and National Institute of Nursing Research; NIA/NINR) scientific coordinators. The steering committee was responsible for designing the study and developing the study protocol included in the MOP (ACTIVE Steering Committee, 2008).

The MOP is necessary for transforming a study protocol into a handbook describing a study's conduct and operations (National Center for Complementary and Integrative Health, 2012). ACTIVE utilized a MOP for describing the scientific rationale, study protocol, study organization and administration, policies and procedures for publications and procedures, standardized data forms, data management protocol, quality assurance protocol with checklists, and standardized interviewing techniques. A design paper was published in *Controlled Clinical Trials*, describing the background and context of the trial; primary objective and hypotheses; design; study population; inclusion/exclusion criteria; recruitment; sample characteristics; outcome measures; field methods, treatments, training

of intervention trainers; the components of memory, reasoning, and speed training; and analytical approaches (Jobe et al., 2001).

The study pre-registration ([ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT00298558) Identifier: NCT00298558) can be found online (National Institutes of Health, 2014). Of note, this registration only concerned the primary and secondary outcomes of the trial, and only considered the planned analyses of intervention effects. As of this writing, a total of 122 manuscripts¹ have appeared in print, but most were not included in the original pre-registration. Since investigators have had access to the data prior to declaring their analysis intentions for subsequent manuscripts, the challenge with pre-registration of such papers is the inability to independently verify that investigators analysis plans were not informed by undeclared exploratory data analyses (Chambers, 2019). ACTIVE tried to mitigate against this concern by having a publication and presentation committee approve all manuscripts prior to analysis, logging the approval dates internally, and (for many) having the analyses independently verified by the coordinating center statisticians.

2. Compliance with Protocol (WHO Principle 6)—Formal oversight of ACTIVE fell under the control of the funding agencies (National Institute on Aging; National Institute of Nursing Research). External review for the funding agencies was conducted by the DSMB; internal monitoring and compliance support was provided by the coordinating center, NERI. The DSMB’s functions are to evaluate accumulated study data to monitor participants’ safety, and the study’s conduct, progress, and efficacy, and recommend the continuation, modification, or termination of the clinical trial (National Institute of Dental and Craniofacial Research, 2018). Regarding ACTIVE, the DSMB was also responsible for reviewing the scientific premise and evaluating the study protocol, measures, interventions, and preparation of the field sites, in addition to monitoring of study recruitment/retention progress and challenges and adverse events.

The study coordinating center at NERI functioned to ensure common practices across field sites, independent management of data, and overall methodological rigor. NERI created and utilized a web-based data management system called Advanced Data Entry and Protocol Tracking (ADEPT) that had many features, including: randomization of participants into specific intervention groups, complete overview of participants’ study status, and status variables incorporated into the study database which permitted constant tracking of forms. ACTIVE also used question-by-question manuals and scripts and standardized data forms to minimize data collection, recording, scoring, and entry errors.

NERI maintained a communication log book at each field site, which stored numbered study-wide memos between NERI and the sites. This served to provide a single, reliable location for all updates and changes to protocol during the study.

¹Electronic searches to determine the number of secondary analyses were performed in PubMed (<https://pubmed.ncbi.nlm.nih.gov/advanced/>, 2001-2020). The main search strategy used combinations of keywords AG014282[Grant Number] OR AG014263[Grant Number] OR AG056486[Grant Number] OR NR004508[Grant Number] OR AG014260[Grant Number] OR AG014276[Grant Number] OR AG014289[Grant Number] OR NR004507[Grant Number]. In addition, we cross-referenced our search with articles from NACDA (National Archive of Computerized Data on Aging, 2015, <https://www.icpsr.umich.edu/web/NACDA/studies/36036/publications>).

3. Qualifications (WHO Principle 10)—Rigorous training certification/recertification was required for all roles (assessment, intervention, data coordination, scoring, and special procedures), and standardized checklists aided in systematizing this process. For example, NERI conducted a 4-day training workshop for all the data collectors from all the field centers. This workshop included instructions, demonstrations, and practice sessions on each test or measurement procedure and presentations on study design, recruitment issues, and interviewing protocol (Jobe et al., 2001). Data collectors and scorers were required to be certified by NERI prior to data collection, and they were provided clear instructions, such as: remaining neutral, probing for more information, and scoring and recording answers properly. They were also required to receive special certification for clinical procedures such as BMI measurements, hand dynamometer usage, and physical performance tests.

PIs led training, certification, and quality-control monitoring for all trainers from all field sites at NERI during an intensive 6-day workshop. All three intervention conditions had uniform certification/recertification requirements for trainers including reviewing all training manuals and participant materials, attending central training, practicing ten training sessions, and being observed by a certifier. Trainers benefited from detailed manuals and scripts to aid in consistency. Trainers were observed by certified trainers at least every 6 months; a standardized assessment checklist was completed, and a debriefing session was held with every trainer (Jobe et al., 2001). PIs also led conference calls to monitor trainers and training activities.

4. Proper Information Handling (WHO Principle 11)—ACTIVE sought to ensure transparency by storing and making all data available on the National Archive of Computerized Data on Aging (NACDA) (National Archive of Computerized Data on Aging, 2015). The repository includes elements of the MOP, instruments/surveys (used as a codebook, along with providing descriptive information about each data element), and all deidentified study files (new random participant IDs were assigned, so that data could not be linked back to original IDs, and thus not back to original participants). From October 29, 2017 to October 29, 2020, there have been 8,071 unique downloads of one or more datasets or codebooks from NACDA (National Archive of Computerized Data on Aging, 2015). In addition to expanding the value of the original data (i.e., new investigators can identify additional research questions not posed by the original investigators), the availability of the data permits independent verification and replication of published study analyses. To date, code associated with specific manuscripts has not been archived, representing a future area of archiving.

NERI statisticians conducted all pre-specified primary data analyses, allowing for statistical analysis independent of the investigators (Ball et al., 2002), and they shared interim findings with the DSMB. The DSMB reviewed these interim statistical reports from NERI to evaluate merit and safety of continued data collection. To better ensure rigorous secondary analyses, guidelines were provided in the MOP for the pre-approval of all proposed presentations and posters. Further guidelines for reviewing drafts and reviews from journals were also described. Prior to 2010, NERI statisticians independently replicated all secondary data analyses.

5. Systems to Assure Quality (WHO Principle 14)—Annually, at a minimum, NERI (with members of the DSMB and NIA/NINR representatives) conducted quality assurance monitoring visits, which included data audits and observations of trainers and assessors to monitor for drift (Tennstedt & Unverzagt, 2013), in addition to providing feedback on assessment, intervention, and data management. The steering committee also regularly reviewed progress and adherence to the protocol and quality control. NERI selected a random 10% of tests to be re-scored by the specified scoring certifiers, and agreement was used to re-certify the original scorers or require they undergo a recertification process. Further, throughout testing, standardized equipment was used and calibrated equally across all six field sites to reduce the likelihood of equipment differences.

Generalizable GCP Principles for the Psychology and Aging Laboratory

The multi-site clinical trial is a unique entity that is likely to be well-funded to provide infrastructure to support GCP. Indeed, NERI received (between NIH fiscal years 1997 and 2010) approximately \$4,866,859 to support their operation². While a substantial portion of these funds were allocated to conduct central training, to support travel costs to the six ACTIVE field sites, and to pay for equipment and telecommunication's costs, these resources also allowed for a higher level of effort allocated to tasks of rigor and reproducibility than the typical single (and sometimes unfunded) laboratory could. But we can take away some generalizable principles that could be implemented, regardless of resources; many of these are consistent with current open science principles. In Table 1, we present the WHO principles of GCP that are specifically germane to rigor, transparency, and reproducibility; an example of how those principles were implemented in ACTIVE; and a proposed means of implementation for single-lab studies.

As a summary, the key generalizable principles of GCP that we emphasize are: (1) Manualization and documentation (of recruitment, study procedures, measure administration protocols and scripts, data coding and scoring, and protocol changes); (2) Pre-registration of study hypotheses and analysis plans; (3) Adequate training (possibly including certification) of assessment, intervention, and data staff; (4) Ongoing quality control and fidelity monitoring, possibly involving external oversight (i.e., by individuals who are not members of the lab team); (5) Open materials (data, code instruments); and (6) Methods sections making transparent the details of data collection and analysis. We consider how these might be implemented on a smaller scale.

1. Scientific Justification (WHO Principle 2)—Manualizing all aspects of the study in a “lab manual,” akin to a MOP used in clinical trials, prior to study initiation, aids in facilitating strict adherence to an a priori plan for the study's conduct, procedures, and analyses and provides valuable information regarding the scientific rationale of the study. Similar to how ACTIVE published a design paper (Jobe et al., 2001), the manuscript or protocol should be submitted as a Registered Report or pre-registered on a registry, such as the Open Science Framework or clinicaltrials.gov (in the case of an experiment/trial) to

²Total costs, as displayed at the NIH Reporter (Query URL: https://reporter.nih.gov/search/qx1-sI_wpE2Vsp_EFtArlg/projects/charts) were aggregated over all budget years funded to the coordinating center under grant number U01 AG014282.

reduce the likelihood of data dredging or other dubious statistical misjudgments by setting pre-defined analysis plans (Chambers, 2019). When submitting future manuscripts, reporting in the method section how the sample size, any data exclusions, all manipulations, and all measures were determined in the study planning phase is a good way to transparently disclose details of data collection and analysis (Simmons et al., 2012).

2. Compliance with Protocol (WHO Principle 6)—To ensure the study is conducted in compliance with the approved protocol, we recommend a detailed data management system, detailed manuals and scripts, standardized data forms, and implementing a communication channel or electronic lab notebook. Data management systems, such as Research Electronic Data Capture (REDCap; Harris et al., 2009; Harris et al., 2019), allow for researchers to track data entry accuracy and set data entry forms with variable labels and value limits. Data management is also aided by the inclusion of question-by-question manuals and scripts and standardized data forms, which may be employed to reduce data collection, recording, scoring, and entry errors. Additionally, uniformity of instructions to participants may be aided by implementing a tablet-based procedure incorporating prompts to lab personnel who must follow a scripted procedure. Additionally, computer-based instruction and training may help eliminate variability in implementation fidelity from less well-trained interventionists. Other tools like the NIH toolbox for cognitive testing (NIH Toolbox, 2021) may be a way to achieve high standardization, while reducing the demands for tester/interventionist training and monitoring.

An idea that is seldom implemented in the local laboratory, but which may have real value in retrospectively documenting design, implementation, and analysis decisions, is a version of ACTIVE's Communication Log. More recent instantiations of this include laboratory-specific communications channels (e.g., Slack or GroupMe; Slack Technologies Inc., 2020; Microsoft Corporation, 2020), which offer archivable threads of team communications, or electronic lab notebooks, which aid in data management by offering the ability to document analysis decisions, code, output, and initial interpretations in a single place (Dirnagl & Przesdzing, 2016). At a minimum, ensuring a study folder in which email correspondence and notes from meetings are saved, along with dating and numbering these documents, would be optimal to ensure updates and protocol changes are stored in a central, known location and protect against any electronic communication losses.

3. Qualifications (WHO Principle 10)—It would be good practice for single laboratory studies to include rigorous training and certification/recertification procedures for all roles. Developing standardized checklists for certification of interviewers, testers, trainers, and data analysts can help regulate these processes; four example certification checklists are provided in Supplemental Figure S1 and may also be found in the MOP (ACTIVE Steering Committee, 2008). Implementing thorough training protocols, including mandating complete review of study protocol and organization, cultural issues that may relate to the study population, and informed consent procedures helps ensure study members adhere to the established protocol.

4. Proper Information Handling (WHO Principle 11)—Chambers (2019) has argued that it is important to implement policies within one's laboratory for standardizing the

public archiving of data (i.e., open data), materials, and code, as well as local archiving of investigator communication (e.g., laboratory manual, notebook, and communication logs). Open data and statistical analysis scripts enable accountability and reproducibility of analysis findings (Jomier, 2017). Open data may also allow for independent researchers to perform alternative analyses or aggregate data into important meta-analyses (Chambers, 2019). It is essential, however, to ensure that sharing of data is done under IRB oversight, and is consistent with applicable privacy laws (e.g., Health Information Privacy, 2000).

Some external oversight of data analysis can be accomplished by pre-registering planned analyses and publishing analysis code. Other strategies might include (a) allocating a study statistician from within the study team who has exclusive access to the data until the study is complete; (b) constructing “buddy” systems with colleagues from external laboratories (either within or outside of one’s institution), in which data analysis roles are housed outside one’s own laboratory; and (c) identifying a within-institution informal data management and safety oversight group, even if one is not required to do so.

Alternatively, if these options are unrealistic, researchers of small laboratories can better account for their own biases via data analytic decisions. Since conclusions can change because of arbitrary choices in data construction, a multiverse analysis, which involves performing analyses across the whole group of datasets that arise from different reasonable choices for data processing, can offer an estimate of how much conclusions vary due to differential decisions in data construction (Steege et al., 2016). This is a means of reducing the likelihood of selective reporting by making the results’ degree of robustness transparent through reporting the range of p -values across all alternatively created datasets. Similarly, displaying a graphical illustration (specification curve) that allows readers to see the statistical consequences of various, reasonable specifications (such as types of regression model, exclusion criteria, and dichotomizing vs. leaving variables continuous) can aid in reducing selective reporting (Simonsohn et al., 2020). Multiverse analyses and specification ‘curves’ are emerging open science concepts and may represent next steps to ensure more transparent data analysis in accordance with GCP principles.

5. Systems to Assure Quality (WHO Principle 14)—As was the case in ACTIVE, quality control should be implemented at regular intervals. Checklists are a standardized means of quality control and can be designed for use among lab members to monitor each other for protocol drift. For example, checklists with questions assessing test or questionnaire scoring accuracy could be employed. Checklists can also be developed for certifying data scorers and enterers, or trainers/interventionists.

Certification of study personnel is perhaps the most resource-intensive task because it involves the development of training materials, trainers, and evaluation of staff proficiency. Single labs can approximate staff credentialing in several suggested ways. One way would be to use measures for which online training and certification are available; for example, the Montreal Cognitive Assessment (MoCA) test (Nasreddine et al., 2005) for cognitive screening offers an online certification process, and this is also true for the Clinical Dementia Rating scale (Morris, 1997), and for many of the measures given as part of the Uniform Data Set of the US Alzheimer’s Disease Research Centers (Besser et al., 2018).

Additionally, single labs could cultivate the open sharing of staff training procedures. For example, if a single laboratory develops a procedure manual for administering and scoring a particular test, putting this on a public site (like the Open Science Framework) could facilitate other laboratories using the same measures/procedures.

Scaling GCP to Single Labs: Limitations and Generalizability Concerns

We appreciate that the typical single lab often does not have access to the resources to help with rigor and reproducibility and that rigor and reproducibility often come with a price. We suggest that our proposed implementations in single-lab studies each help to move the needle in favor of increasing GCP and rigor and reproducibility, but we do not anticipate/expect all labs to implement all these strategies, especially considering the substantial financial, personnel, and time limitations single labs face. It will be up to the individual lab to determine whether the benefits of GCP outweigh the costs, though some of our proposed strategies are likely more feasible and applicable to most research studies and smaller research operations in the domains of psychology and aging. These include manualization and documentation, pre-registration of study hypotheses and analysis plans, adequate training of staff, ongoing quality control and fidelity monitoring, open materials, and transparent methods sections.

There is also the counterintuitive possibility that less rigorous and standardized studies may be more generalizable; for example, a phenomenon that does not survive slight variability in instructions may not be worth understanding since it is unlikely to replicate. Thus, balancing internal and external validity is another important consideration for labs hoping to improve their rigor and reproducibility. While the clinical trial setting clearly seeks to optimize internal validity, it would be inaccurate to say that generalizability is not a chief concern; the GCP principles described here are not intended to increase the artificiality of the research environment, but to increase the consistency of researcher behavior, the documentation of procedures and data, and the sharing of data and code and other study resources.

Recommended Infrastructural Changes

The NIH wishes to enhance rigor and reproducibility in scientific research (National Institutes of Health, 2019a). Recently in 2014, the NIH expanded the clinical trial definition to be a research study with human subjects prospectively assigned to one or more interventions to evaluate its effect on health-related biomedical or behavioral outcomes (National Institutes of Health, 2017b). This definition requires many investigators to adopt many of the procedures and practices used in clinical trials. If the resources required to support GCP and enhance rigor and reproducibility (manualization, quality control observations, etc...) were construed as necessary research infrastructure available to all investigators, then perhaps the cost of such infrastructure could become part of negotiated indirect cost recovery agreements. Since these principles of GCP apply to any study seeking to collect and analyze data from human or non-human studies, we view these strategies as appropriate across the continuum of research. However, at universities with fewer grant funding resources, we would recommend professional organizations at the university-level work together to provide resources to support GCP.

Additionally, professional/scientific organizations could offer measure training and quality assurance materials to their members by crowdsourcing protocols from members/laboratories, or organizing their communities to produce protocols, and then providing organized, publicly-accessible lists of protocols that could be a major service to members. For example, the Society for Behavioral Medicine (SBM) offers free GCP training to anyone with a free registration. NIH-supported GCP training (U.S. Department of Health and Human Services, 2017) could be an excellent model for more widespread training and certification. Many NIH-funding institutions have a Clinical Translational Science Institute (CTSI), which is charged with supporting institutional research through cores (e.g., Data Management, Statistical Data Analysis, Study Design and Grant Proposal Development, etc...), and it would be reasonable to add to the charters of these centers the creation of a measure repository with training and credentialing support. For smaller institutions without a CTSI, the charter of institutions with such centers could be expanded to provide regional support to other institutions in their catchment area, or the Division of General Medical Sciences of the NIH could provide libraries of protocols for researchers at all institutions. There is already a precedent for this kind of broad protocol sharing through REDCap, a database building and management web application (Harris et al., 2009; Harris et al., 2019). REDCap is a shared resource among all institutions with an NIH-funded CTSI. In the REDCap system, individuals have developed data entry templates for many common measures. These are then put in a public library, and anyone can use them to facilitate building their own data entry templates. Finally, the NIH is currently launching a Common Data Elements Repository, which will standardize data collection (Mendoza-Puccini & Wilkins, 2021). The CDE repository will provide background information about each measure (including psychometrics) and provide clear variable definitions and machine/human readable forms. Providing certification services for these forms could be rolled into the mandate of the CTSI system. Notably, the CDE repository would be available even to institutions without substantial grant portfolios or a CTSI. This suggests that, building on the CTSI model, NIH (or professional associations) could develop a general “Good Clinical Practice” institute, with certification checklists, guidelines, archiving and data sharing “how-tos” that anyone from any setting could use.

Conclusion

The growing proliferation of Open Science suggests a set of practices at the level of the individual psychology laboratory that are consistent with decades of GCP implementation in clinical trials. To that end, there are regularized practices (MOP, independent analysis, quality control, archiving) that can be used as models for implementation in single-lab studies. GCP tools, themselves, can become part of the Open Science tradition. For example, as laboratories increasingly utilize MOPs, standardized instruments, data management tools, and archiving solutions, these can be made openly available, so that others can adopt them in their own laboratories. Although it is doubtful that single laboratories will be able to match the level of coordinated rigor and reproducibility resources seen in a multi-site clinical trial, we believe that GCP principles generally, and their implementation in clinical trials, can serve as excellent models for strengthening procedure standardization.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants from the National Institute on Aging: U01 AG014260 (Dr. George W. Rebok), U01 AG014282 (Dr. Sharon L. Tennstedt), U01 AG014263 (Dr. Sherry L. Willis), U01 AG014289 (Dr. Karlene K. Ball), U01 AG014276 (Dr. Michael Marsiske), R01 AG056486 (Dr. Sherry L. Willis, Dr. George W. Rebok), T32 AG020499 (Dr. Michael Marsiske, Dr. Glenn E. Smith, Dr. Adam J. Woods). This work was also supported by grants from the National Institute of Nursing Research: U01 NR004508 (Dr. Frederick W. Unverzagt) and U01 NR004507 (Dr. John N. Morris).

References

- ACTIVE Steering Committee. (2008). Manual of Procedures. <https://osf.io/nyad9/>
- Ball K, Berch DB, Helmers KF, Jobe JB, Leveck MD, Marsiske M, ... & Unverzagt FW. (2002). Effects of cognitive training interventions with older adults: a randomized controlled trial. *JAMA*, 288(18), 2271–2281. [PubMed: 12425704]
- Besser L, Kukull W, Knopman DS, Chui H, Galasko D, Weintraub S, Jicha G, Carlsson C, Burns J, Quinn J, Sweet RA, Rascovsky K, Teylan M, Beekly D, Thomas G, Bollenbeck M, Monsell S, Mock C, Zhou XH, Thomas N, Robichaud E, Dean M, Hubbard J, Jacka M, Schwabe-Fry K, Wu J, Phelps C, Morris JC Clinical Core leaders of the National Institute on Aging. (2018). Version 3 of the National Alzheimer's Coordinating Center's Uniform Data Set. *Alzheimer Disease and Associated Disorders*, 32(4), 351–358. [PubMed: 30376508]
- Biswas K, Carty C, Horney R, Nasrin D, Farag TH, Kotloff KL, & Levine MM (2012). Data management and other logistical challenges for the GEMS: the data coordinating center perspective. *Clinical Infectious Diseases*, 55(suppl_4), S254–S261. [PubMed: 23169938]
- Chambers C (2019). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton University Press.
- Dirnagl U, & Przesdzing I (2016). A pocket guide to electronic laboratory notebooks in the academic life sciences. *F1000Research*, 5, 2. 10.12688/f1000research.7628.1 [PubMed: 26835004]
- Freedman LP, Venugopalan G, & Wisman R (2017). Reproducibility2020: progress and priorities. *F1000Research*, 6, 604. 10.12688/f1000research.11334.1 [PubMed: 28620458]
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, & Conde JG (2009). A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377–81. [PubMed: 18929686]
- Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, McLeod L, Delacqua G, Delacqua F, Kirby J, & Duda SN (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, 95, 103208. [PubMed: 31078660]
- Health Information Privacy. (2000). Summary of the HIPAA Privacy Rule. U.S. Department of Health & Human Services. Retrieved October 23, 2020 from <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>
- Jobe JB, Smith DM, Ball K, Tennstedt SL, Marsiske M, Willis SL, ... & Kleinman K. (2001). ACTIVE: A cognitive intervention trial to promote independence in older adults. *Controlled Clinical Trials*, 22(4), 453–479. [PubMed: 11514044]
- Jomier J (2017). Open science—towards reproducible research. *Information Services & Use*, 37(3), 361–367.
- Kaplan RM, & Irvin VL (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PloS One*, 10(8), e0132382. [PubMed: 26244868]
- Kazdin AE (2016). *Methodological issues and strategies in clinical research*. Washington, DC: American Psychological Association.

- Kidwell MC, Lazarevi LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg LS, ... & Errington TM (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, 14(5), e1002456. [PubMed: 27171007]
- Maxwell SE, Lau MY, & Howard GS (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, 70(6), 487. [PubMed: 26348332]
- Mendoza-Puccini C, & Wilkins KJ (2021, June 24). Common data elements: Increasing fair data sharing. National Institutes of Health. <https://nexus.od.nih.gov/all/2021/06/24/common-data-elements-increasing-fair-data-sharing/>
- Mertens W, & Recker J (2020). New guidelines for null hypothesis significance testing in hypothetico-deductive IS research. *Journal of the Association for Information Systems*, 21(4), 1.
- Microsoft Corporation (2020). GroupMe. Retrieved October 23, 2020, from <https://groupme.com/en-US/>
- Morris JC (1997). Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International Psychogeriatrics*, 9(S1), 173–176. [PubMed: 9447441]
- Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings J, & Chertkow H (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695–699. [PubMed: 15817019]
- National Archive of Computerized Data on Aging. (2015, July 29). Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE), United States, 1999-2008 (ICPSR 36036). National Institute on Aging. Retrieved September 23, 2020, from <https://www.icpsr.umich.edu/web/NACDA/studies/36036/summary>
- National Center for Complementary and Integrative Health. (2012). Guidelines for Developing a Manual of Operations and Procedures (MOP). Retrieved September 23, 2020, from https://files.nccih.nih.gov/s3fs-public/CR-Toolbox/MOP_NCCIH_ver1_07-17-2015.pdf
- National Heart, Lung, and Blood Institute. (2011, October). Compendium of Best Practices for Data Coordinating Centers. U.S. Department of Health & Human Services. Retrieved September 23, 2020, from <https://www.nhlbi.nih.gov/events/2011/compendium-best-practices-data-coordinating-centers>
- National Institute of Dental and Craniofacial Research. (2018, July). Data and Safety Monitoring Board (DSMB) Guidelines. U.S. Department of Health & Human Services. Retrieved September 23, 2020, from <https://www.nidcr.nih.gov/research/human-subjects-research/toolkit-and-education-materials/interventional-studies/data-and-safety-monitoring-board-guidelines>
- National Institutes of Health Reporter. Research Portfolio Online Reporting Tools (RePORT). Retrieved October, 23, 2020, from https://projectreporter.nih.gov/reporter_SearchResults.cfm?icde=52279311
- National Institutes of Health. (2014, April 16). ACTIVE: Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE). *Clinicaltrials.gov*. <https://clinicaltrials.gov/ct2/show/NCT00298558>
- National Institutes of Health. (2017, May 16). Good Clinical Practice Training. U.S. Department of Health & Human Services. Retrieved September 23, 2020, from <https://grants.nih.gov/policy/clinical-trials/good-clinical-training.htm>
- National Institutes of Health. (2017b, August 8). NIH’s Definition of a Clinical Trial. U.S. Department of Health & Human Services. Retrieved February 13, 2021, from <https://grants.nih.gov/policy/clinical-trials/definition.htm>
- National Institutes of Health. (2019a, Dec. 12). Rigor and reproducibility. U.S. Department of Health & Human Services. Retrieved September 23, 2020, from <https://www.nih.gov/research-training/rigor-reproducibility>.
- National Institutes of Health. (2019b, Dec. 12). Guidance: Rigor and reproducibility in grant applications. U.S. Department of Health & Human Services. Retrieved September 23, 2020, from <https://grants.nih.gov/policy/reproducibility/guidance.htm>.
- NIH toolbox. (2021). <https://www.healthmeasures.net/explore-measurement-systems/nih-toolbox>

- Onken LS, Carroll KM, Shoham V, Cuthbert BN, & Riddle M (2014). Reenvisioning clinical science: unifying the discipline to improve the public health. *Clinical Psychological Science*, 2(1), 22–34. [PubMed: 25821658]
- Open Science Framework. (2020, Aug. 30). Open Science Badges enhance openness, a core value of scientific practice. Retrieved October 23, 2020, from <https://www.cos.io/initiatives/badges>
- Shrout PE, & Rodgers JL (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, 69, 487–510.
- Simmons JP, Nelson LD, & Simonsohn U (2012). A 21word solution. SSRN. 10.2139/ssrn.2160588
- Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, & Stine-Morrow EA (2016). Do “brain-training” programs work?. *Psychological Science in the Public Interest*, 17(3), 103–186. [PubMed: 27697851]
- Simonsohn U, Simmons JP & Nelson LD (2020). Specification curve analysis. *Nature Human Behavior*. 10.1038/s41562-020-0912-z
- Slack Technologies Inc. (2020). Slack. Retrieved October 23, 2020, from <https://slack.com/>
- Steegeen S, Tuerlinckx F, Gelman A, & Vanpaemel W (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702–712. [PubMed: 27694465]
- Sullivan I, DeHaven A, & Mellor D (2019). Open and reproducible research on open science framework. *Current Protocols Essential Laboratory Techniques*, 18(1), e32.
- Tennstedt SL, & Unverzagt FW (2013). The ACTIVE study: Study overview and major findings. *Journal of Aging and Health*, 25(8 0), 3S–20S. [PubMed: 24385638]
- U.S. Department of Health and Human Services. (2017, May 16). Good clinical practice training. National Institutes of Health. <https://grants.nih.gov/policy/clinical-trials/good-clinical-training.htm>
- Vijayanathan A, & Nawawi O (2008). The importance of Good Clinical Practice guidelines and its role in clinical trials. *Biomedical Imaging and Intervention Journal*, 4(1), e5. 10.2349/bijj.4.1.e5 [PubMed: 21614316]
- Willis SL, Tennstedt SL, Marsiske M, Ball K, Elias J, Koepke KM, ... & Wright E. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *JAMA*, 296(23), 2805–2814. [PubMed: 17179457]
- World Health Organization. (2002). Handbook for good clinical research practice (GCP): guidance for implementation. Retrieved September 23, 2020, from https://www.who.int/medicines/areas/quality_safety/safety_efficacy/gcp1.pdf

Table 1

World Health Organization “Good Clinical Practice” Principles: Implementation in ACTIVE and Single-Lab Studies

WHO Principle # and Description	Example Implementation in ACTIVE	Proposed Implementation in Single-Lab Studies
2. Research involving humans should be scientifically justified and described in a clear, detailed protocol	<ul style="list-style-type: none"> - Manual of Procedures (MOP) - Scientific rationale in MOP and introduction of design paper - Predefined analysis plans 	<ul style="list-style-type: none"> - Lab Manual - Scientific rationale included in Lab Manual and/or Registered Report - Pre-defined analysis plans - Methods sections make transparent details of data collection and analysis
6. Research involving humans should be conducted in compliance with the approved protocol;	<ul style="list-style-type: none"> - Full protocol conducted in compliance with approved protocol - Detailed data management protocol, detailed manuals and scripts, standardized data forms - Communication log book 	<ul style="list-style-type: none"> - Conduct full protocol in compliance with approved protocol - Detailed data management protocol, detailed manuals and scripts, standardized data forms - Communication channels, electronic lab notebooks, or study folder - Tablet and Computer-based Instruction
10. Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his or her respective task(s) and currently licensed to do so, where required;	<ul style="list-style-type: none"> - Rigorous training and certification/recertification procedures for all roles (assessment, intervention, data, coordination, test scoring; special certification for clinical procedures such as BML, hand dynamometer, physical performance tests), checklists 	<ul style="list-style-type: none"> - Rigorous training and certification/recertification procedures for all roles, checklists
11. All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation, and verification;	<ul style="list-style-type: none"> - Publicly available MOP; instruments, and data (via NACDA) - Separation of data analysis role (coordinating center analysis) from PIs - DSMB reviews interim data - Pre-approval and post-review of all presentations and papers 	<ul style="list-style-type: none"> - Make Lab Manual, instruments, data, and code publicly available - Separation of data analysis role (independent statistician); “buddy system” to check data analyses; or within-institution informal data management and oversight - Multiverse analyses, specification curves
14. Systems with procedures that assure the quality of every aspect of the trial should be implemented.	<ul style="list-style-type: none"> - Regular quality assurance observations (assessment, intervention, data management), including by outside parties (coordinating center, DSMB, funding agency) - 10% quality checks of all test scoring - Standardized equipment and calibration 	<ul style="list-style-type: none"> - Regular quality assurance (utilizing checklists) - Quality checks of test and/or questionnaire scoring - Use measures for which online training and certification are available

Note. DSMB = Data safety and monitoring board; MOP = Manual of procedures; NACDA = National Archive of Computerized Data in Aging; PI = Principal Investigator; WHO = World Health Organization