

Toward a Comprehensive and Accurate Measure of Clinical Trial Workload, Equity, Quality Assurance and Patient Safety-How Much Workload is Too Much? Commentary and Brief Research Report

Ralph Jay Johnson*

MD Anderson Cancer Center, University of Texas, Unit 439, 1515 Holcombe Blvd Houston, Texas 77030, USA

*Corresponding author: Ralph Jay Johnson, MD Anderson Cancer Center, University of Texas, Unit 439, 1515 Holcombe Blvd Houston, Texas 77030, USA, Tel: +832-372-3511; E-mail: rjohnson@mdanderson.org

Received date: May 09, 2017; Accepted date: May 23, 2017; Published date: May 29, 2017

Copyright: © 2017 Johnson RJ. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

This article provides a brief commentary on the methodology surrounding controlled clinical trials, the growing trend of centers conducting multiple controlled trials (i.e. "factory science"), trial workload measures, and possible relationships among workload, especially excessive workload, and mistakes, mishaps, deviations, violations, or just plain slippage. Findings are reported for each factor or measure included in an incremental algorithm designed to provide a numeric score for clinical trial workload. This algorithm was developed in the interest of quality assurance as part of program evaluation through an Oracle Delphi process by a study group of subject matter experts who work with a substantial number of clinical trials in an international cancer center in Houston, Texas (UT-MD Anderson). At a minimum, the algorithm also reflects the complexity of the issues surrounding the clinical trial workload and the conduct of clinical trials in general. Unlike previous measures reported in the literature, it may lack in simplicity for expedient use as a tool for informing management, although it provides comprehensiveness and accuracy and lends itself more to scientific testing. Future avenues of study are considered.

Keywords: Controlled clinical trials; Workload measures; Patient safety; Quality assurance; Research study management

Introduction

Controlled clinical trials, specifically research studies comparing the use and non-use of drugs and devices, are commonly considered the gold standard for evaluating new medical interventions [1-10]. They accurately test, validate, improve on, and advance the generalized body of scientific knowledge and inform and extend available medical treatments [1,3]. Thus, despite the inherent risks, such clinical trials are the keystones for treatment and scientific progress [1,3].

Nevertheless, the integrity and safety of this research has been questioned due to scandals regarding lapses in study implementation, neglect, non-observance of details, errors and oversights, non-compliance, and deviations and violations [11-20]. However, confidence in the integrity and safety of medical intervention research is crucial for continued sponsorship, funding, and participation and advancement of medical science [11,12,18-22].

A common industry-wide contention is that an upper threshold exists on the number of multiple and different responsibilities to which workers must attend [23-27]. Beyond that threshold, the potential and certainty increases in terms of mistakes, mishaps, deviations, violations, or just plain slippage [23,25,26,28]. This does not include staff turnover due to overwork and burnout. Put differently, the greater the workload, the more sophisticated and complicated the work, the greater the probability something will go wrong, or terribly wrong, and in medical care the unthinkable will happen with patients. The same is no less true with controlled clinical trials. (Note: concerns have even been voiced about mistakes and oversights in the scientific publications arena of peer review in that reviewers and editors are overwhelmed

and overworked with a plethora of complicated and sophisticated study reports leading to poor science slipping through [29]).

Increasingly, the trend in controlled trial research appears to be consolidation into centers specializing in particular disease fields and populations conducting multiple, similar controlled clinical trials. The same study staff is expected to support increasing numbers of studies simultaneously. A 2017 review of clinicaltrials.gov data revealed, for example, that approximately 8,500 active trials were conducted in the United States alone, with 2,750 or 33%, being Phase 1 or 2 trials (the more complicated, sophisticated, and risky). Simply put, the trend has been increasing in the direction of "factory science." About 169 centers conducted six or more trials, with a median of 10 and a maximum of 128. Clearly, there is the potential for a violation of the psychological maxim of the rule of 7s as applied to the conduct of science [30,31]. The general rule of 7s refers to a psychological "rule" or management principle suggesting that humans can successfully attend to no more than seven basic stimuli-for-response or responsibilities at any one time [30,31]. After that, they must juggle and will eventually experience errors and breakdown; the more stimuli-for-response or responsibilities added, the more the juggling increases, with a consequent rapid decrease in the time to errors and breakdown [30,31]. Based on the clinicaltrials.gov data, many center staffs are responsible for clinical trials exceeding the number recommended by the rule of 7s. Even when they do not exceed that number, the nature of clinical trials can be extremely demanding and this can only be compounded with reduced staff when austerity measures are instituted.

First, the median number of clinical trials conducted at centers is 10. Second, the workload involved with clinical trials can be highly demanding. Thus, the rule of 7s may even be insufficient for setting an upper limit; nor can it be relied on as a threshold for when

circumstances become overwhelming and risky in terms of patient safety, thereby degrading scientific integrity and ultimately sabotaging successful study accomplishment [23,24].

To control the trial workload from a management standpoint, Goode et al. [32] developed an acuity scale (scored between 1 and 4) that depends on a pooling of several factors representing trial complexities, multiplies the number of patients by the score, then assigns scores to variously skilled nurse study coordinators. They found this beneficial in terms of rebalancing workloads through routine monitoring. They noted that their acuity score was a beneficial but rudimentary and limited managerial tool for addressing some of the harmful effects of excessive workload stress. However, they also acknowledged that it might be an oversimplification, likely fails to include important and relevant factors that influence workload, and is not a good measure for evaluating the different amounts of influence particular factors have and the constellation of those factors on work-product quality. As a first but critical step toward a more accurate and comprehensive estimation of “how much is too much,” a more detailed algorithm was derived for estimating clinical trial workload. The purpose of this commentary and research brief is to report on that algorithm and elaborate on the reasoning behind the development of the different factors used in it. Once a more comprehensive and accurate measure of clinical trial workload is approximated, the next step can be broached, specifically, relating that measure to a threshold where too much workload impinges on quality assurance, study integrity, and patient safety.

Method

To devise an algorithm to approximate clinical trial workflow, the study employed the Oracle Delphi [33-35] process. The process was used among a study group of subject matter experts who were tasked with coordinating the conduct of approximately 30 complex, sophisticated, and complicated Phase I/II cancer treatment trials as part of a department in a major international cancer treatment hospital center. The crux of the method was including group judgment and response as more valid than individuals’ judgment and response alone. This also covers for shortfalls where precise prediction has yet to be established. In addition, it avoids the time-consuming expense of conducting large-scale surveys that only test rather than develop. The study group developed a preliminary algorithm based on that used by Goode et al. [32], a review of the literature on workload measures applied to controlled clinical trials, and their own individual experiences. This was then circulated for discussion and revision among the study group members until consensus was reached.

Findings

The Oracle Delphi process resulted in the development and refinement of the following additive algorithmic model for approximating controlled clinical trial workload per worker and the factors constituting workload. The model is additive in that each factor combines into a total workload score and the higher the score, the greater the workload. In some instances, factors are weighted according to how they ultimately affect workload (Figure 1). This also provides some insight into the dynamics of how clinical trials operate.

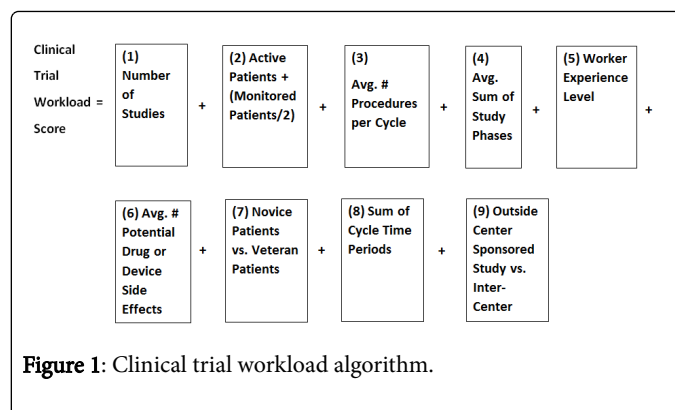


Figure 1: Clinical trial workload algorithm.

Number of Studies

The number of studies represents a crude measure of the amount of overall workload per worker which, as Goode et al. [32] noted, is present and primary in all measures of clinical trial workload. Nevertheless, for example, one worker might have six Phase 2 studies with two patients in them but only monitor the patients monthly. However, another worker might have only three studies, but these studies may be Phase I and have six very medically unstable patients in each study, with the patients receiving a drug with many serious side effects on a weekly cycle. Using only the number of studies, the workload in the former would sensibly be deemed higher but in actuality the latter’s workload would far exceed the former’s.

Number of patients actively receiving the treatment+(Number of patients being monitored/2)

The study group’s reasoning was that actively treated patients must receive far more (double) attention due to sequels related to their experimental treatment and possible instability. In contrast, monitored patients not receiving treatment and those minimally maintained or at the point of receiving standard of care only need half as much time and attention.

Average number of procedures per cycle

The measure for average number of procedures per cycle is for all studies assigned to a worker. This can be easily assessed in clinical trial protocols in the schedule of events and orders for procedures. The reasoning here is to distinguish, for example, a worker who has 12 studies with monthly cycles with an average of three procedures from another worker who has only two studies but an average number of 24 procedures and weekly cycles. The latter’s workload far exceeds the former’s though the former’s might appear to be higher.

Average sum of study phases

The average sum of study phases uses the study phase metric shown in Table 1; the number scored per trial is inversely proportional to the number phase type for a study trial. The study group’s logic was to show that the lower phase study trials are more complicated and demanding in that they have more sophisticated procedures with more unstable patients. Thus, for example, a worker who has three Phase 1 trials, two Phase 2 trials, and one Phase 3 trial would have an average of 4.6.

This average provides an incrementally weighted score that is proportionately inverse based on combining and averaging study phases. The problem with this measure is that the vast majority of centers conducting multiple trials conducts mostly Phase I and II trials. So, this measure would discriminate between a worker conducting mostly Phase I trials and another conducting mostly Phase II trials.

Phase I	6
Phase II	4
Phase III	2
Phase IV	1

Table 1: Clinical trial phase type metric.

Worker experience level*

The study group's reasoning was that a more seasoned worker can accomplish more and this ability should be factored into the composite picture of workload. Put differently, experienced workers' workloads, though substantial, would be considered far less because their degree of work competence is much higher and performance tasks are automatic for them (i.e. "they have been drilled"). This measure uses the metric in Table 2. The score is proportionately inverse to the amount of workers' experience. So, for example, a worker with little or no experience would receive a higher workload score of 8, whereas a veteran worker with 14 years of experience would receive a workload score of 0-that worker knows the job. (Note*: Remarkably, the study group never considered education level as a factor in experience. In terms of workload, what counted was actual length of experience doing the work).

8+ years	0
5-7 years	2
1-2 years	4
0-1 years	8

Table 2: Worker experience level.

Average number of potential drug /device side effects

The measure for average number of potential drug/device side effects applies to all studies assigned to a worker. This can easily be assessed in clinical study trial protocols published in the investigator brochure or the study protocols. The study group's logic was that the more the side effects spread over studies, the more complicated the studies are and the greater the workload.

Novice patients vs. veteran patients

The study group's consensus was that a weight should be added for each novice patient. Specifically, a patient new to the medical care/hospital system in which the trial is conducted should be counted as two patients as opposed to a veteran patient. The reasoning was that new patients need far more attention and shepherding.

Sum of cycle types weight

The study group noted that the measure for sum of cycle type's weight is a reflection of the increasing complexity and oversight of trials, especially those that are not initiated by the center sponsor. For example, some trials have daily cycles whereas others have weekly or monthly cycles. This weights the entire algorithm score for the sum of those cycle values, as shown in Table 3.

Daily	4
Weekly	3
Monthly	2
Greater than monthly	0

Table 3: Cycle time periods.

For example, a worker can have one study with daily cycles, one with weekly cycles, and four with monthly cycles for a total score of 15. The study group recognized that sometimes the time period of cycles does not fit neatly; for example, some studies have cycles on the 1st, 8th, and 15th days or the 1st and 3rd days of 28-day cycles and those are considered weekly studies. Thus, sometimes approximations or "force fitting" and judgment calls are made using the metric table. Studies that merely monitor patients every several months would be assigned no numeric value using this metric. This measure was of one the study group probably struggled with the most and it may need further refinement.

Institution or investigator initiated vs. outside sponsor initiated studies

Finally, the study group's reasoning was that outside sponsor-initiated studies must receive far more (double) attention time and effort due to coordinating logistics and the sponsors' unfamiliarity with the center's systems, dynamics, operations, and even organizational culture. These studies are assigned a numeric value of 1 and, counter intuitively, inter-center investigator-initiated studies are assigned no numeric value.

Discussion

Although the algorithm is more detailed and complicated than simple acuity scores based on accumulated factors [32], it incrementally incorporates the major factors identified as contributing to controlled trial workload in a way that would be expected. One issue is that it is a comparative score at this point. However, without more data points, it lacks a range with upper and lower parameter values. Nevertheless, the algorithm as a social artefact alone also reflects the complexity, breadth, and depth of a rapidly evolving field, the range of the issues surrounding clinical trial workload, and the conduct of clinical trials in general. Put differently, the score paints a comprehensive picture of the different factors and their additive effect (i.e. tangling up and piling up), even with a low number of studies and/or patients. This is at least worth considering in light of the trend toward factory science, namely, a substantial shift to centers and continuous flow production involving conducting more trials along the same lines and using roughly the same staff.

What the algorithm reported herein lacks in simplicity for expediently informing management, it gains in accuracy and lends

itself more to scientific testing. The algorithm reported is preliminary and not the be all and end all; it is reported to stimulate discourse on how much eventually is too much. Its features can be easily incorporated into electronic spread sheets for comparative measures of individual clinical trial workloads to achieve re-balancing and fair and equitable workload distribution as well as inform and advance quality work product, trial integrity, successful trial accomplishment, and patient safety.

More importantly, this is a first step toward eventual correlation between workloads and factors representing clinical trial deviations and violations that can degrade patient safety-though this is an extremely sensitive subject. The problem in studying and reporting risks to patients is a tacit admission that patient safety is in some way compromised, which might not necessarily be the case. Nevertheless, to achieve this eventual objective will involve conducting a statistical modelling analysis of the factors in the algorithm and the resulting scores to determine which factors statistically significantly align with and how they predict measures of patient risk (as well as which ones drop out). This will also provide some notion about a metric or threshold (i.e. "how much is too much") above which a high probability of work quality being sacrificed is and patient safety being compromised. Simply put, there is much more work to be done.

Acknowledgements

The Author wishes to acknowledge Ms. Lore Lagrone for material, in-kind support, encouragement, and review of the concept for this manuscript. The Author gratefully appreciates the in-kind support of UT-MDACC Department of Myeloma. The Author also gratefully thanks the following UT-MDACC personnel for their involvement: Ms. Ashley Morphey, Ms. Sylvia Munoz, Ms. Kathleen Walls and Mr. Jasper Olsem. The Author also thanks Ms. Jacqueline Ramey for proofing and copyediting.

Conflicts of Interest

The author declares no conflicts of interest.

References

- Bothwell LE, Greene JA, Podolsky SH, Jones DS (2016) Assessing the Gold Standard-Lessons from the history of RCTs. *N Engl J Med* 374: 2175-2181.
- Nardini C (2014) The ethics of clinical trials. *Ecancermedalscience* 8: 387.
- DeSerres G, Skowronski DM, Wu XM, Ambrose CS (2013) The test-negative design: validity, accuracy and precision of vaccine efficacy estimates compared to the gold standard of randomized placebo-controlled clinical trials. *Euro Surveill* 18: 20585.
- Khokar MA, Rathbone J (2016) Droperidol for psychosis-induced aggression or agitation. *Cochrane Database Syst Rev* 12: CD002830.
- Koopmeiners JS, Hobbs BP (2016) Detecting and accounting for violations of the constancy assumption in non-inferiority clinical trials. *Stat Methods Med Res* [Epub ahead of print].
- Cerritelli F, Verzella M, Cicchitti L, D'Alessandro G, Vanacore N (2016) The paradox of sham therapy and placebo effect in osteopathy: A systematic review. *Medicine (Baltimore)* 95: e4728.
- Vlieg-Boerstra BJ, Bijleveld CM, van der Heide S, Beusekamp BJ, Wolt-Plompen SA, et al. (2004) Development and validation of challenge materials for double-blind, placebo-controlled food challenges in children. *J Allergy Clin Immunol* 113: 341-346.
- Chang C, Lin CH (2003) Hormone replacement therapy and menopause: A review of randomized, double-blind, placebo-controlled trials. *Kaohsiung J Med Sci* 19: 257-270.
- Johnson SR, Dunn BK, Anthony M (2001) Defining benefits and risks for SEMs in clinical trials and clinical practice. *Ann NY Acad Sci* 949: 304-314.
- Ellis SJ, Adams RF (1997) The cult of the double-blind placebo-controlled trial. *Br J Clin Pract* 51: 36-39.
- Unger CA, Barber MD (2005) Studying surgical innovations: Challenges of the randomized controlled trial. *J Minim Invasive Gynecol* 22: 573-582.
- Altman DJ (1994) The Scandal of poor medical research. *BMJ* 308: 283-284.
- <http://www.apa.org/monitor/2010/09/trials.aspx>
- <http://trialsjournal.biomedcentral.com/articles/a0.1186/cvm-1-1-003>
- www.sciencedirect.com/article/pii/S074920819190068Z
- Marusic A, Wager E, Utrobicic A, Rothstein HR, Sambunjak D (2016) Interventions to prevent misconduct and promote integrity in research and publication. *Cochrane Database Syst Rev* 4: MR000038.
- Satar O, Hlali S (2015) Independent data monitoring committees: An update and overview. *Urol Oncol* 33: 143-148.
- Ball G, Piller LB, Silverman MH (2011) Continuous safety monitoring for randomized controlled clinical trials with blinded treatment information. Part 1: Ethical considerations. *Contemp Clin Trials* 32: S2-S4.
- Wells RJ (2008) Secrecy and integrity in clinical trials. *J Clin Oncol* 26: 680-682.
- Seigel D (2003) Clinical trials, epidemiology, and public confidence. *Stat Med* 22: 3419-3425.
- Board on Health Sciences Policy (2001) Public Confidence and Involvement in Clinical Research: Symposium Summary, Clinical Roundtable, Sep 2000. Institute of Medicine: Washington D.C.
- Fügener A, Edenharter GM, Kiefer P, Mayr U, Schiele J, et al. (2016) Improving intensive care unit ward utilization by adapting master surgery schedules. *Periop Svcs* 6: 172-180.
- Ehrenfeld JM, Dexter F, Rothman BS, Minton BS, Johnson D, et al. (2013) Lack of utility of a Decision Support System to mitigate delays in admission from the operating room to the post-anesthesia care unit. *Economic, Education, and Policy* 117: 1444-1452.
- Dexter F, Abouleish AE, Epstein RH, Whitten CW, Lubarsky DA (2003) Use of operating room information system data to predict the impact of reducing turnover times on staff costs. *Anesth Analg* 97: 119-126.
- Bjorklund G, Petterson S, Schagtay E (2007) Performance predicting factors in prolonged exhausting exercise of varying intensity. *Eur J Appl Physiol* 99: 423-429.
- Wahlström J, Mathiassen SE, Liv P, Hedlund P, Ahlgren C, et al. (2010) Upper arm postures and movements in female hairdressers across four full working days. *Ann Occup Hyg* 54: 584-594.
- Cohen D, Wherry RJ Jr, Glenn F (1996) Analysis of workload generated by multiple resource theory. *Aviat Space Environ Med* 67: 139-145.
- Taekman JM, Stafford-Smith M, Velazquez EJ, Wright MC, Phillips-Bute BG, et al. (2010) Departures from the protocol of a clinical trial: A pattern from the data record consistent with a learning curve. *Qual Saf Health Care* 19: 405-410.
- Matwin S, Kouznetsov A, Inkpen D, Frunza O, O'Brien P (2010) A new algorithm for reducing the workload of experts in performing systematic reviews. *J Am Med Inform Assoc* 17: 446-453.
- Miller GA (1994) The magical number of seven, plus or minus two: Some limits on our capacity for processing information 1956. *Psychol Rev* 101: 343-352.
- Cowan N (2015) George Miller's Magical Number of immediate memory in retrospect: Observation on the faltering progression of science. *Psychol Rev* 122: 536-541.
- Goode M, Lubejko B, Humphries K, Medders A (2013) Measuring Clinical Trial-Associated workload in community clinical oncology program. *J Oncol Pract* 9: 211-215.

-
33. Rowe G, Wright G (1999) The Delphi technique as a forecasting tool: issues and analysis. *Int J Forecast* 15: 353-375.
34. Rowe G, Wright G (2001) Expert opinion in forecasting: role of Delphi technique. In: Armstrong JS, editor. *Principles of forecasting: a handbook of researchers and practitioners*. Boston: Kluwer Academic Publishers.
35. Dalkey N, Helmer O (1963) An experimental application of the Delphi method to the use of experts. *Manag Sci* 9: 458-467.