

OPEN

Individualized Coaching After Stroke Does Not Work

How Much or Which One?

Luigi Tesio, MD, Stefano Scarano, MD, and Laura Perucca, MD

Key Words: Stroke Rehabilitation, Individualized Coaching, Trial Design, Tailoring of Interventions

(*Am J Phys Med Rehabil* 2020;99:e3–e6)

Preventing or slowing the long-term functional decline after stroke is a difficult and very relevant clinical challenge. An elegant Norwegian study recently tested the efficacy of an individualized coaching program, which started at 10–16 wks after stroke and continued for 18 mos.¹ The program comprised individualized monthly coaching by a physiotherapist, partly via phone meetings. Based on the patient's exercise preferences and subjective goals, an exercise schedule was defined for the following month. The schedule included physical activity to be performed, on average, 45–60 min/d, inclusive of 2–3 periods of vigorous activity (score of 15–17 on the 6–20 Borg exertion scale), once a week. Various settings were offered: individualized or group treatment, with exercise at home or in an outpatient clinic. A daily training diary had to be completed. This individualized program was superimposed on the “standard care” provided after discharge from the hospital. The standard care comprised 45-min exercise sessions at moderate intensity every week for at least 3 mos and up to 6 mos or longer in selected patients. Outpatient, home, or inpatient settings could be offered to the patients.

WHAT IS THE GOAL OF THE ARTICLE?

The study was a multicenter, single-blind randomized controlled trial: coaching + standard care ($n = 186$) versus standard care only ($n = 194$). The primary outcome was the score on the Motor Assessment Scale (MAS) at the 18-mo follow-up visit. The MAS is an eight-item six-level score; “supine to side lying onto intact side” and “hand movements” are

representative items. A series of secondary measures was collected (e.g., the Barthel independence index, 6-min walk test, etc.). Adverse events, compliance, and the International Physical Activity Questionnaires measure of physical activity were also recorded.

WHAT ARE THE CONCLUSIONS OF THE ARTICLE?

A Very Reasonable Hypothesis Failed

Although compliance and safety were excellent, the conclusions of the study were negative: “regular individualized coaching did not improve maintenance of motor function or the secondary outcomes compared with standard care.”

ARE THERE STRENGTHS OR LIMITATIONS TO THE STUDY THAT ARE IMPORTANT IN INTERPRETING THE RESULTS?

The study was very rigorous in both design and statistics. The authors were very correct in reporting the negative results, which are notoriously underreported.² Given the relevance and seriousness of the research, it may be helpful to reflect on the reasons for the failure to confirm an otherwise reasonable hypothesis; that is, the addition of a careful and customized coaching program should add efficacy to a mild “standard” program in terms of the maintenance in functional status after stroke.

The following considerations do not detract from the quality of the article, which followed the best standards of controlled studies in the field; rather, they aim at offering a critical viewpoint on the application of these standards.

A Hidden Drift Toward No Evidence

Perhaps the negative results could have been anticipated because of four biases implicit in the design, all of which diminished the statistical power of the study.

1. A ceiling effect in the outcome measures was troublesome, although a functional decline, not an improvement, was expected. For example, the Barthel index of independence, which ranges from 0 to 100, was 96/100 at baseline. This bias is easy to explain. In the endeavor to sustain the “pragmatic” nature of the study, the inclusion criteria were rather loose (e.g., age ≥ 18 yrs, first-ever or recurrent stroke; ischemic or

From the Department of Biomedical Sciences for Health, Università degli Studi di Milano, Milano, Italy (LT, LP); and Department of Neurorehabilitation Sciences, Istituto Auxologico Italiano, IRCCS, Milano, Italy (LT, SS, LP).

All correspondence should be addressed to: Luigi Tesio, MD, Dipartimento di Scienze Neuroriabilitative, Istituto Auxologico Italiano, IRCCS, via Giuseppe Mercuri 30, 20122 Milano, Italy.

The study was supported by Ministero della Salute, Ricerca Corrente 2019.

Financial disclosure statements have been obtained, and no conflicts of interest have been reported by the authors or by any individuals in control of the content of this article.

Copyright © 2019 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ISSN: 0894-9115

DOI: 10.1097/PHM.0000000000001153

hemorrhagic). Nevertheless, the trial flowchart shows that 944 (71%) of the 1324 screened patients were not eligible or declined participation. Institutionalized patients, who were obviously not recruited, represented 22.4% of the screened patients. Twenty-three percent of the patients denied participation: were they the most impaired patients? One can also assume that the remaining 21% of the excluded patients comprised only more impaired patients (however, this information is lacking). In summary, these randomized patients were high functioning and thus, presumably, rather stable.³ One may assume that they were also rather active (however, the International Physical Activity Questionnaires of physical activity was not recorded at baseline). This selection bias challenged the claimed “pragmatic” approach of the study.

2. The outcome measures did not seem to be adequately sensitive to change for two reasons. The first reason concerns the ceiling effect (which was honestly acknowledged by the authors in their discussion). Presumably, in these high-functioning, ambulating patients, the primary outcome measure (the MAS score) could only detect changes in the upper arm and hand items (three of the eight available items), which are the most difficult ones to change.⁴ This also raises a validity concern. It is hard to imagine how the “vigorous activity” foreseen by the treatment could influence the functioning of the paretic upper limb. The second reason concerns the error variance in the measurements. One source of variance comprised the therapist-patient interaction. Coaching implies communication skills in the recipient, but aphasia was only considered as a criterion for lowering the recruitment threshold for the Mini-Mental State Examination (the threshold was really a very low one: 21/30 reduced to 17/30 in cases of aphasia). Low Mini-Mental State Examination scores may indicate deficits in language, attention, spatial orientation, and motor programming, which would impose constraints on any tailored exercise program and require careful monitoring. However, the monitoring was mostly based on a training diary.

3. A minor, yet conceptually relevant, source of variance across the secondary outcome measures was the extraction of single items, used as a 1-item scale, from the multi-item questionnaires. This was the case for the “standing on one leg” item in the 14-item Berg Balance Scale and for one fatigue-related item from the Helseundersøkelsen i Nord-Trøndelag-HUNT3 questionnaire (which was hard to identify among the set of 21 questionnaires and “sections” of the HUNT3 battery).⁵ In traditional test theory the error measurement decreases with the number of scale items (the random error averages out). Even more important, 1-item scales are suspect for low validity, given that they do not necessarily reflect the same latent trait across different subjects or raters.⁶

4. The exercise regimen and clinical conditions of the individual patients were ill-defined, thus introducing error variance in the assessment of both groups. The selection of the treatment goals was based on an interview detailing the patient's goals and physical activity preferences. However, the adopted Exercise Preference Questionnaire poses very generic questions, which in the original article lead to the demonstration that “... relative to controls, stroke survivors preferred exercise to be more structured, in a group, at a gym or fitness centre, and for exercises to be demonstrated.”⁷ In real life, no physiotherapists should base their exercise prescription only on these contextual, nonclinical

factors. Again, in an endeavor to ensure the pragmatic nature of the design, exercise was defined generically and on a purely subjective basis.

Paradoxically, three methodological flaws worked in the other direction, in that they tended to inflate the statistical power of the comparisons in outcomes between treated and control samples. First, the “coaching” program was added to “standard care”; this favored the additive approach (e.g., via an increased placebo effect). Second, the authors themselves admit in the discussion that “People may overestimate their activity levels when self-reported measures are used.” Third, no corrections in the significance level for multiple comparisons were applied across the series of secondary outcomes. Nevertheless, no significant differences emerged. Thus, by greater force, the existence of a drift of the design toward a low power is supported.

HOW DOES THIS HELP IN CLINICAL PRACTICE?

Four Lessons Can Be Learned From This Study

a) *Rehabilitation is more than a black box.*

Negative results are quite common in “pragmatic” studies that aim to compare the efficacy of new exercise programs *versus* “standard” or “usual” care (e.g., neurodevelopmental [Bobath] exercises *versus* standard care in poststroke rehabilitation inpatients).⁸ In particular, this holds true for studies comparing forms of health service organizations in the real world of chronic, home-based care. A recent example of negative results, for instance, is provided by a meta-analytic study on home care after hip fracture.⁹ An underestimated problem in the rehabilitation literature is the temptation to consider a treatment package as a black box. In a seminal article published in 2003, this metaphor was elegantly substituted by a new metaphor of a Russian nesting doll.¹⁰ The outer doll of, say, a form of service organization (e.g., an in- or outpatient service) may conceal many inner dolls containing “treatments” of increasing granularity (e.g., which type of exercise prescription, administered by which professional, with which schedule, etc.). The degree of detail in the description of the treatment is a critical choice for the researcher and must match the target of the research (e.g., the level of participation, activity limitations, and impairments, as per the International Classification of Functioning- World Health Organization glossary) and the trial design. Proponents of the new metaphor acknowledge that “Research designed to compare forms of health service organization...will probably not profit from unpacking of the contents of treatment at a detailed level because the goal of such research is to optimize the deployment of large-scale components of health care.” However, this holds for studies comparing services already in place, onto which no experimental intervention is applied. In experimental studies, where the relevant variables are manipulated, not just observed, the search for the deepest appropriate doll, based on theory-driven hypotheses, is recommended. The same authors admit that “it is currently the case that most rehabilitation treatments are not based on specific theories of functional recovery but rather on traditions and administrative conveniences. Such theories that do guide practice tend to be very general, such as, ‘Work on it and it will improve.’” Perhaps the ambiguity between epidemiological and experimental designs was the subtle fault in

the Norwegian study. In this and many other rehabilitation studies, the theory guiding the selection of exercise treatment and the degree of its detail are simply limited to macro categories of exercise, such as aerobic and strength training, with little customization. For instance, these categories are overemphasized in official guidelines concerning chronic stroke patients, whereas a much larger variety of approaches, including motor, cognitive, sphincter, and pain treatments, is listed for acute stroke inpatient rehabilitation.¹¹ Consistently enough, there is a tendency to import and apply trial designs coming from the pharmacological and/or the epidemiologic tradition.

b) *Rehabilitation is not a pill.*

Too often rehabilitation programs are assimilated to drug therapies so that their generic type (e.g., neuromuscular vs. aerobic, etc) and dose and schedule (e.g., number of sessions, minutes/day) are considered as sufficient specifications. In behavioral research, where the outcome may be heavily predicted by individual interactions, at least at the same level as that by “main effects,” explicit decision-tree tailored algorithms should substitute the “dosage” criterion. The if/then nodes should be based on the pathophysiological knowledge of impairments (e.g., unbalance, spasticity, weakness, pain, etc.) and a theory linking the treatment components to the outcome variable (here, the MAS score). Advanced efforts to open the Russian dolls by defining a taxonomy of treatment components and targets, which would allow one to channel the therapeutic decision within a theory-driven frame, exist.^{12,13}

c) *Persons are not populations.*

Individual subjects may differ in too many variables, rendering the matching of treated and control samples difficult. Fortunately, the group-cohort approach (with uphill randomization and blinding) is not the only game in town. Refined quasi-experimental designs can accommodate for individual uniqueness.¹⁴ A further way to mediate between the need for “mean” and individual outcomes is to count the number of patients who achieve a pre-set threshold of change, rather than measure the average change across a sample.¹⁵ In the article reviewed here, a main (and sensible) concern of the authors was the patients' compliance. This was sought after by privileging the patient's preferences. Although this approach is a form of tailoring, it works at the expense of a clear, theory-driven diagnostic process based on the motor impairments and a prognostic definition of susceptibility to specific exercises. Biological (e.g., pharmaceutical) treatments act rather deterministically on body cells. By contrast, the more the treatment is based on whole-person psychological and behavioral features, the more the results are also sustained by clinical and psychological characteristics and by unpredictable interactions within both the patients and patient-therapist pairs, with these interactions being potentially unbalanced between groups. Given that the MAS score was preferred as a primary outcome in the reviewed study, which had an average score at baseline of approximately 40 of 48, perhaps a pilot study should have been performed, with patients selected on the basis of their sensitivity to the scale (e.g., excluding subjects at both the floor and ceiling of the instrument). The training program should have been focused on the patient's individual motor impairment (the principal outcome scale is centered on the *impairment*

World Health Organization domain), not on subjective preferences expressed in terms of exercise context (home vs. gym; self-selected vs. demonstrated, etc). For instance, if unbalance emerges as a main problem from the viewpoint of both the patient and therapist, balance exercises should be customized and emphasized.^{16,17} Little or no effect can be anticipated, for instance, from strenuous aerobic exercises, regardless of whether these are conducted at home or outdoor, etc. Indeed, caution should be also required in following preferences expressed in more detail with respect to motor impairments. For instance, if a poststroke patient is already walking autonomously (as a 90/100 Barthel score would suggest), yet has an awkward sickling gait, and one of his or her upper limbs is irreversibly impaired, the patient may nonetheless express a preference for the treatment of the upper limb although the largest margin for recovery can be foreseen, by the therapist, for gait. In summary, with a more theory-based, tailored program, the power of the study should have been higher, rendering positive results more likely and negative results (if any) more convincing.

d) *“Fishing” across outcomes does not substitute for a priori hypotheses.*

The inclusion of many “secondary” endpoints imitates epidemiological designs based on wide surveys. The population, or the “mean” individual, not individual persons, is the target of such studies. Regardless of their design, studies with many secondary endpoints aim at (risky) inferences regarding the cause-effect nature of associations across many variables. The cost of these “fishing expeditions” is, of course, the inflation of the false positive rate for “significant” associations, and the risk for hazardous cause-effect interpretations.¹⁸ By contrast, behavioral studies can be driven by strong a priori knowledge and allow the experimental manipulation of variables. Behavioral studies allow one to work on precise targets with limited samples; at an extreme, trial designs exist, providing evidence from single cases.¹⁹

Take-Home Message

The negative results from this elegant article demonstrate that rehabilitation research must be highly flexible in its trial designs²⁰ and not act subordinate to either biological or epidemiological designs. Otherwise, under such blurred lenses, the efficacy of rehabilitation medicine is doomed to fade. Along the gradient from body parts to the individual person and to populations, specific research paradigms must be applied for each level of observation.²¹ When a behavior or a perception (i.e., a whole-person latent trait) is the issue, specific quasi-experimental designs,¹⁴ pathophysiological a priori knowledge, and specific statistics (such as a Rasch analysis for questionnaires²² and the use of individual minimal detectable changes²³) must concur in the search for “evidence,” which, in essence, provides plausibility of a cause-effect inference. In the case of the person, evidence-based medicine, which is population oriented, should be always complemented by person-oriented, medically based evidence.

REFERENCES

1. Askim T, Langhammer B, Ihle-Hansen H, et al: Efficacy and safety of individualized coaching after stroke: the LAST Study (Life After Stroke). A pragmatic randomized controlled trial. *Stroke* 2018;49:426–32

2. Fanelli D: Negative results are disappearing from most disciplines and countries. *Scientometrics* 2012;90:891–904
3. Sennfalt S, Norrving B, Petersson J, et al: Long-term survival and function after stroke. *Stroke* 2018; STROKEAHA118022913
4. Lima E, Teixeira-Salmela LF, Magalhães LC, et al: Measurement properties of the Brazilian version of the Motor Assessment Scale, based on Rasch analysis. *Disabil Rehabil* 2018;2:1–6
5. Hunt3 Survey: Available at: <https://www.ntnu.edu/hunt/hunt3>. Accessed January 14, 2019
6. Franchignoni F, Salaffi F, Tesio L: How should we use the visual analogue scale (VAS) in rehabilitation outcomes? I: how much of what? The seductive VAS numbers are not true measures. *J Rehabil Med* 2012;44:798–9
7. Banks G, Bernhardt J, Churilov L, et al: Exercise preferences are different after stroke. *Stroke Res Treat* 2012;2012:890946
8. Hafsteinsdóttir TB, Algra A, Kappelle LJ, et al: Neurodevelopmental treatment after stroke: a comparative study. *J Neurol Neurosurg Psychiatry* 2005;76:788–92
9. Kuijlaars IAR, Sweerts L, Nijhuis-van der Sanden MWG, et al: Effectiveness of supervised home-based exercise therapy compared to a control intervention on functions, activities, and participation in older patients after hip fracture: a systematic review and meta-analysis. *Arch Phys Med Rehabil* 2019;100:101–14.e6
10. Whyte J, Hart T: It's more than a black box; it's a Russian doll: defining rehabilitation treatments. *Am J Phys Med Rehabil* 2003;82:639–52
11. Winstein CJ, Stein J, Arena R, et al: Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2016;47:e98–169
12. Zanca JM, Turkstra LS, Chen C, et al: Advancing rehabilitation practice through improved specification of interventions. *Arch Phys Med Rehabil* 2019;100:164–71
13. Hart T, Dijkers MP, Whyte J, et al: A theory-driven system for the specification of rehabilitation treatments. *Arch Phys Med Rehabil* 2019;100:172–80
14. Shadish WR, Cook TD, Campbell DT: *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA, Houghton Mifflin Co., 2002
15. Zamboni P, Tesio L, Galimberti S, et al: Efficacy and safety of extracranial vein angioplasty in multiple sclerosis: a randomized clinical trial. *JAMA Neurol* 2018;75:35–43
16. Tinetti ME, Baker DI, McAvay G, et al: A multifactorial intervention to reduce the risk of falling among elderly people living in the community. *N Engl J Med* 1994;331:821–7
17. Bricchetto G, Piccardo E, Pedullà L, et al: Tailored balance exercises on people with multiple sclerosis: a pilot randomized, controlled study. *Mult Scler* 2015;21:1055–63
18. Ioannidis JP: Exposure-wide epidemiology: revisiting Bradford Hill. *Stat Med* 2016;35:1749–62
19. Kazdin AE: *Single-Case Research Designs. Methods for Clinical and Applied Settings*. Oxford, Oxford University Press, 1982
20. Frontera WR: Clinical trials in physical medicine and rehabilitation. *Am J Phys Med Rehabil* 2015;94:829
21. Tesio L: The good-hearted and the brave. Clinical medicine at the bottom of the barrel of science. *J Med Pers* 2010;8:103–11
22. Tesio L: Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *J Rehabil Med* 2003;35:1–11
23. Tesio L: Outcome measurement in behavioural sciences: a view on how to shift attention from means to individuals and why. *Int J Rehabil Res* 2012;35:1–12