Contemporary Clinical Trials Communications 18 (2020) 100554

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: http://www.elsevier.com/locate/conctc

Short communication

Mechanisms for the testing effect on patient-reported outcomes

Salene M.W. Jones [a, *], Lisa J. Shulman [b], Julie E. Richards [b, c], Evette J. Ludman [b]

[a] Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA, 98109, USA
[b] Kaiser Permanente Washington Health Research Institute, 1730 Minor Ave, Seattle, WA, 98101, USA
[c] University of Washington Health Services Department, Seattle, WA, USA

ARTICLE INFO

Keywords:
Measurement reactivity
Test-retest effect
Retest effect

ABSTRACT

The testing effect is when patient-reported outcomes (PRO) improve with repeated administration without intervention. The testing effect can confound interpretation of clinical trials using PROs as endpoints. This study investigated potential mechanisms. The parent study (n = 302) investigated a self-management intervention for depression. We qualitatively analyzed exit interview feedback from the 89 control group participants completing the last assessment. Participants reported several perceived benefits from control group participation including novel mechanisms (study participation was meaningful, emotional support, appreciating outreach), a possible negative testing effect and mechanisms previously identified (behavioral change).

1. Introduction

Randomized clinical trials (RCT) are increasingly using patient-reported outcomes (PROs) as endpoints. The testing effect is when participants report improvement after repeated PRO administration without any intervention. As most RCTs using PROs utilize a control group with repeated assessments, the testing effect may lead to underestimation of the actual treatment effect. The testing effect is small to medium and inconsistent across studies [1–3], including in assessment frequency needed to see the effect [1,3–5]. The testing effects occurs for mood, health behaviors, attitudes and beliefs [2], but not positive emotions [3,4,6].

Data supporting testing effect mechanisms are scant though several have been suggested. Repeated assessment may function like self-monitoring in psychotherapy, increasing awareness and triggering change [7–9]. Social desirability has not been strongly supported as a testing effect mechanism [2–4]. Response shift refers to changes in how people answer PRO questions over time, such as changing their baseline [10]. Mood-congruent processing occurs when people experience test anxiety from a research assessment and this leads to more negative mood at the first but not subsequent assessments [3]. Improving our understanding of the mechanism for the testing effect will help inform PRO use and RCT design. We conducted a qualitative analysis of exit interviews from an assessment-only control group of an RCT for a self-management depression program to identify testing effect mechanisms.

2. Methods

2.1. Participants and procedures

An RCT of a self-management support program recruited participants from two Seattle, WA healthcare systems from 2010 to 2013 [11]. Inclusion criteria included: age 18–79, history of recurrent depression or dysthymia, ≥10 on the Patient Health Questionnaire (PHQ9) [12] and willingness to attend in-person group sessions. The PHQ9 has nine items corresponding to the symptoms of a major depressive episode and scores range from 0 to 27. Exclusion criteria were bipolar disorder; cognitive impairment or serious medical illness; and plans to move out of state in the next 18 months. Participants were randomized to the intervention (n = 150) or a usual care control group (n = 152). The intervention included visits with a psychotherapist, a peer support specialist and weekly self-management skills training. Participants in the control group were only contacted by blinded interviewers. Study participants completed structured assessments of PROs at baseline, 3, 6, 12, and 18 months by phone or in-person. The assessors made up to 10 attempts to contact participants for assessments. Participants did not receive the results of their

* Corresponding author.
E-mail addresses: smjones3@fredhutch.org (S.M.W. Jones), Lisa.J.Shulman@kp.org (L.J. Shulman), Julie.E.Richards@kp.org (J.E. Richards), evette@dbandel.com (E.J. Ludman).

https://doi.org/10.1016/j.conctc.2020.100554
Received 25 November 2019; Received in revised form 2 March 2020; Accepted 14 March 2020
Available online 16 March 2020

assessments. After the 18-month assessment, participants were asked: "Is there anything you would like to share with the study team about your experience in this project?". Study procedures were approved by the institutional review boards at both sites.

2.2. Data collection & analysis

The participants' responses to the unstructured exit interview feedback question was examined. Interviewers wrote participants' verbatim responses on paper, which were later entered in the questionnaire database. The unstructured text from the final phone interview with the participants assigned to the control group was imported into Dedoose, a web-based application designed for qualitative and mixed-method analysis. One of the interviewers, trained in qualitative methods, coded the data using inductive content analysis and organized the coded data into themes [13,14]. The coded data and themes were reviewed by the senior author to ensure consistency and distinctiveness of the codes and subsequently reviewed and refined by all authors.

3. Results

In total, 89 participants in the control group provided feedback and were included in the qualitative analysis, although 37 provided responses that were generic ("I like [this clinic]") and not coded. Participants in the intervention group were excluded from this analysis. The participants were 63% female (n = 56), an average age of 51 years old (SD = 12.9) and 72% were Caucasian. At study entry, participants had an average PHQ9 score of 14.8 (SD = 3.76, range 10–27) meaning all participants had moderate to severe depressive symptoms at baseline. As reported in the original outcomes paper, depressive symptoms declined for both the intervention and control groups but more so for the intervention group [11]. Inductive content analysis resulted in two major themes and several subthemes (see Fig. 1).

3.1. Positive aspects of usual care: appreciated outreach

Usual care participants appreciated the study team's outreach and were thankful that someone was calling to check-in. Being part of mental health research, specifically depression research, felt important. One participant said,

> "I am glad that so many studies are focused on depression, especially an issue for women I think. I appreciated how professional everyone was on the phone. I am hard to get a hold of because I work full time during the day so I appreciate everyone's perseverance and continuing to follow up with me."

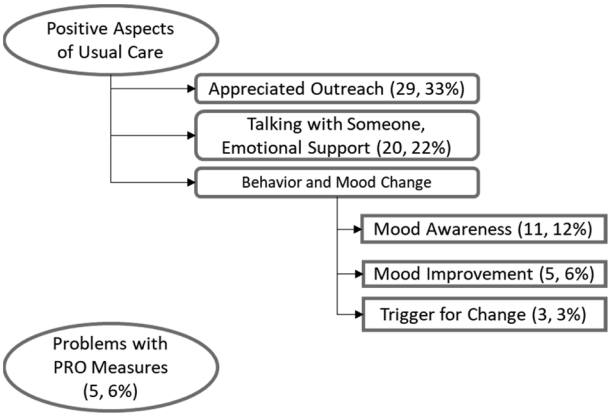

Fig. 1. Qualitative Codes. Parent codes are in ovals, child codes in rounded rectangles and grandchild codes in rectangles. Comments that could not be coded were generic ("Thank you for doing this study"; "I like [this clinic]").

Some participants appreciated the interviewers' persistence in accommodating their schedules.

3.2. Positive aspects of usual care: talking with someone, emotional support

Usual care participants commented that talking with the interviewers provided emotional support. Participants responded, "*It was helpful. You called in the nick of time with the surveys. Just answering questions was helpful, just having someone to talk to.*" One participant expressed appreciation for nonjudgmental support, "*It has been really helpful and supportive just having that neutral party to check in with.*" Participants commented that the interviewers were professional, patient, respectful, and courteous.

3.3. Positive aspects of usual care: behavior and mood change

Mood Awareness. Control group participants mentioned being made more aware of one's mental health status after completing the follow-up phone surveys. A participant noted, "*It helps me recall where I am at and how am I doing and where I want to be. Otherwise, I get caught up in the day to day of life and forget about recovery and that I am actually doing really well.*" Another replied, "*I was in the control group, but I think these surveys are a good check-in. They helped me notice specific things I was struggling with or working on. It shows me areas I need to hone in on.*"

Mood Improvement. Participants shared that their mood had improved from being part of the control group. One participant stated, "*Before this study and the other study I felt doomed, but being in these studies has helped me feel I can get through. I still have bad times but it is better. I hope my being in this study can help someone else.*" Another participant stated, "*I think it has helped a great deal to know that there is this anonymous group out there who cares.*"

Trigger for Change. Participants stated the interview process was a prompt to change behavior such as regularly taking medications or doing more yoga and gardening. One respondent commented that the interview,

> "… helped motivate me to get more care. When you would call I would realize I needed to get more help and I would reach out to my providers. Like just now you asking about the last 6 mos. I realized that I've been feeling low and I haven't been seeing anyone; that's not good."

3.4. Problems with PRO measures

Participants suggested improvements to the study. Respondents commented that certain survey questions were open to interpretation, making them difficult to answer. Answering questions over the phone was challenging. Others mentioned that the questions could trigger emotional distress, indicating a possible negative testing effect. Another respondent commented that the timeframes on some questions made them difficult to answer.

4. Discussion

This study examined exit interviews from a depression treatment RCT control group and identified mechanisms of the testing effect. Two novel mechanisms were identified: perceived social and emotional support from talking with the assessors [15] and finding meaning from research study participation [16]. Many participants reported the assessments triggered changes including increased mood awareness, improved mood and positive behavioral changes, consistent with previously suggested mechanisms [1]. Although one previous study of the testing effect found no relationship of the effect to behavior change [4], this was likely because their measure of behav-

ior change was unvalidated. We also found some evidence of a "negative testing effect" where a subset of participants reported the questions could make them feel worse and some participants reported the questions were difficult to answer. Some testing effect mechanisms were not mentioned in the interviews, specifically regression to the mean and mood-congruent processing.

The limitations of the study should be considered. As with all qualitative research, study participants' views may not be generalizable. This was a clinical sample. The exit interviews did not specifically ask about completing the measures repeatedly. We likely missed any causes of the testing effect that would not be identifiable to participants. Qualitative analysis typically has two coders and only one person coded the data in this study.

The results have implications for RCT design and future research. For trials using PROs, particularly as primary endpoints, the testing effect should be considered in power calculations and interpreting results. The Solomon four-group design or positive valence PROs may be solutions for the testing effect [2–4,6]. Research is needed to quantitatively test the mechanisms of the testing effect, including new mechanisms identified here.

Declaration of competing interest

The authors do not have any conflicts to disclose.

Acknowledgments

The authors would like to thank the study participants. The National Institute of Mental Health funded the parent study (MH065530) and the clinical trial registration number is NCT01139060.

References

[1] H. Hesser, et al., The effect of waiting: a meta-analysis of wait-list control groups in trials for tinnitus distress, J. Psychosom. Res. 70 (4) (2011) 378–384.
[2] D.P. French, S. Sutton, Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it?, Br. J. Health Psychol. 15 (Pt 3) (2010) 453–468.
[3] J.P. Sharpe, D.G. Gilbert, Effects of repeated administration of the beck depression inventory and other measures of negative mood states, Pers. Indiv. Differ. 24 (4) (1998) 457–463.
[4] B.T. Longwell, P. Truax, The differential effects of weekly, monthly and bimonthly administrations of the Beck Depression Inventory-II: psychometric properties and clinical implications, Behav. Ther. 36 (2005) 265–275.
[5] W.A. Arrindell, Changes in waiting-list patients over time: data on some commonly-used measures. Beware!, Behav. Res. Ther. 39 (10) (2001) 1227–1247.
[6] A.F. Jorm, P. Duncan-Jones, R. Scott, An analysis of the re-test artefact in longitudinal studies of psychiatric symptoms and personality, Psychol. Med. 19 (2) (1989) 487–493.
[7] K. Scott, C.C. Lewis, Using measurement-based care to enhance any treatment, Cognit. Behav. Pract. 22 (1) (2015) 49–59.
[8] S. Coster, et al., Self-monitoring in Type 2 diabetes mellitus: a meta-analysis, Diabet. Med. 17 (11) (2000) 755–761.
[9] P. Rozin, E.B. Royzman, Negativity bias, negativity dominance and contagion, Pers. Soc. Psychol. Rev. 5 (4) (2001) 296–320.
[10] F.J. Oort, M.R. Visser, M.A. Sprangers, Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift, J. Clin. Epidemiol. 62 (11) (2009) 1126–1137.
[11] E.J. Ludman, et al., Organized self-management support services for chronic depressive symptoms: a randomized controlled Trial, Psychiatr Serv 67 (1) (2016) 29–36 appips201400295.
[12] K. Kroenke, R.L. Spitzer, J.B. Williams, The PHQ-9: validity of a brief depression severity measure, J. Gen. Intern. Med. 16 (9) (2001) 606–613.
[13] U.H. Graneheim, B. Lundman, Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness, Nurse Educ. Today 24 (2) (2004) 105–112.
[14] R. Whitley, M. Crawford, Qualitative research in psychiatry, Can. J. Psychiatr. 50 (2) (2005) 108–114.
[15] M.A. Reger, et al., Implementation methods for the caring contacts suicide prevention intervention, Prof. Psychol. Res. Pract. 48 (5) (2017) 369.
[16] F. Riessman, How does self-help work?, Soc. Pol. 7 (2) (1976) 41–45.