

Video Review of Baseline Performance on Global Ratings
in a Double-Blind Placebo Surgery Trial

Cynthia McRae, PhD¹, Jennifer Caspari, PhD², Daniel Russell, PhD³,

Heiner Ellgring, Dr. rer.nat., Dipl.Psych⁴,

Chris Bezzant, MA¹, Paul Greene, MD⁵, Stanley Fahn, MD⁶

¹University of Denver ²VA Portland Health Care System ³Iowa State University

⁴University of Wuerzburg ⁵Mt. Sinai Health System, NY

⁶Neurological Institute, Columbia University Medical Center

Corresponding author:

Cynthia McRae
Morgridge College of Education
University of Denver
Denver, CO 80208
Phone: 303.871.2475
Email: cynthia.mcrae@du.edu

Key words: Parkinson's disease, double-blind, placebo surgery, video review, global ratings

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/mdc3.12666

Abstract

BACKGROUND: A randomized double-blind sham surgery-controlled trial was conducted to determine the effectiveness of implantation of human embryonic dopamine neurons into the putamen of patients with advanced Parkinson's disease (PD). The present analyses determined whether patients viewing a video of themselves performing motor activities off medications at baseline would affect self-ratings 12 months later on the Global Rating Scale (GRS).

OBJECTIVES: To examine changes in GRS scores pre-/post-video review for the total sample; to examine differences in scores between *actual* implant and sham groups, as well as *perceived* groups pre- and post-video review; to examine differences among four subgroups of patients based on *actual* and *perceived* treatment (i.e., actual implant/perceived implant).

METHODS: Forty participants were recruited and randomly assigned to receive either neural implantation or sham surgery. The primary outcome variable was a one-item GRS ranging from -3 (much worse since surgery) to +3 (much improved since surgery). At 12 months (before the blind was lifted) patients rated themselves on the GRS before and after viewing the baseline video.

RESULTS: Total sample GRS scores improved after the video ($P = .001$). There were no differences between the *actual* implant and sham groups before or after the video, but there were differences between *perceived* groups at both times ($P < .001$). Among subgroups, improvement after the video was found only in the group receiving the implant but who *thought* sham ($P = .011$).

CONCLUSION: When self-ratings are an outcome variable, review of baseline videos is recommended before making comparative ratings.

Accepted Article

INTRODUCTION

A double-blind sham surgery-controlled trial was developed to determine the effectiveness of implantation of human embryonic dopamine neurons into the putamen of patients with advanced Parkinson's disease (PD), with half the patients receiving the implant (n = 20) and half receiving sham surgery (n = 20)¹. The protocol also specified recruiting half the participants to be 60 years of age or younger and half to be older than 60 because analyses for age were pre-specified. The double-blind was maintained for at least 12 months. The overall purpose of this study (hereafter referred to as the parent study) was to determine whether patients who received the implant improved more than patients who received sham surgery over the extended period of the double-blind. Recruitment, clinical assessments and videotaping were performed at the Irving Center for Clinical Research at Columbia University in New York City. Surgeries were performed at the University of Colorado Hospital. The primary outcome was the self-administered Global Ratings Scale (GRS) in which participants rated themselves from -3 (much worse since surgery) to +3 (much improved since surgery) at 12 months after surgery. The published results of the study found no significant differences in the primary outcome between the implant and sham groups¹. There were statistically significant improvements in some of the secondary outcomes of the trial among the younger participants (age \leq 60), a pre-determined outcome.

Quality of life (QoL) was assessed extensively in a related, concurrent study with 30 of the 40 parent study patients participating. In addition to determining whether patients who received the implant improved more than patients who received sham surgery in terms of QoL over the period of the double-blind, a second aim of the study

was to determine whether patients who *thought* they received the implant improved more than patients who *thought* they received the sham surgery². Results of this QoL study indicated there was a strong placebo effect in this group².

Taken together, the parent study and QoL study allowed investigators to determine not only the benefits of neural implantation versus sham surgery, but also provided opportunities to examine the placebo effect, or the effects of *perceived* treatment, or type of surgery patients *thought* they received. This unique project also allowed us to investigate the effects of a double-blind condition (specifically, sham surgery) over an extended period of time (12 months).

As part of the baseline assessment in the parent study, a video was made of each patient (on and off medications) performing the Motor component of the Unified Parkinson's Disease Rating Scale (UPDRS)³, which was then archived for review 12 months after surgery. One of the Principal Investigators of the parent study (SF) theorized that when participants made their final rating on the GRS prior to lifting the blind, they would not be able to clearly remember their condition prior to surgery. It was hypothesized that if patients saw the video, they would be reminded of their previous condition and be able to make a more accurate assessment of how much their functioning had changed since surgery.

The initial assessment design was approved by the NIH reviewers, the sponsor, but the Data and Safety Monitoring Board (DSMB) appointed by the NIH was concerned that such a self-assessment based on viewing videotapes of themselves had never been carried out before, and did not approve this analysis. Instead, the DSMB required that the primary outcome be the self-administered GRS without a prior review of what they

looked like on the videos taken 12 months earlier (at baseline). In addition, the DSMB encouraged the investigators to investigate whether such post-video GRS assessments have merit in future clinical trials. The investigators were encouraged to obtain the baseline videos and have the subjects review them and re-score themselves on the GRS after they had already scored the GRS without reviewing their videos. Only the results of the DSMB-approved GRS analysis (without video review) were included in the publication of the parent study report¹.

The primary aim of this present study was to determine whether there were changes in GRS scores at 12 months between pre- to post-review of the video made at baseline. The second aim was to determine whether there were differences in scores between the *actual* implant and sham surgery groups, as well as differences between the *perceived* implant and sham groups before or after viewing the video. If changes or differences between groups were found, the third aim of the study was to examine differences among the four subgroups of patients based on *actual* and *perceived* treatment (i.e., actual implant/perceived implant, actual implant/perceived sham, etc.). This is the first study known to the authors which has assessed the influence of viewing videotapes taken at baseline on self-rated global functioning at 12 months in a double-blind trial.

PATIENTS AND METHODS

Patients

Forty participants with advanced PD were recruited from across the United States and Canada and were randomly assigned to receive either the neural implant or sham surgery. The mean age at baseline was 57.8 years, with an average disease duration of

15.7 years. The mean age at disease onset was 42.1 years. Nineteen women and 21 men participated in the study; the average level of education was 16.4 years.

Methods

Neurological assessments were performed by medical staff during the four day inpatient evaluations at baseline and at 4, 8, and 12 months after surgery. Evaluation was done both on and off medications. The primary outcome variable of the parent study was the Global Rating Scale¹, which was completed by patients seven days after surgery and one week before the 4, 8, and 12 month neurological evaluations while the participants were at home and had not yet been examined by the investigators. At the same times, patients also indicated which surgery they thought they had received. On the basis of their perception at 12 months, participants were assigned to one of four subgroups; actual implant/perceived implant, actual implant/perceived sham, etc. All results were mailed to the biostatistician, with the postmarks serving as authentication of the date. The clinical investigators remained blind to the mailed GRS results. Appropriate institutional review board approvals were received from the universities involved in the trial.

At the baseline evaluation, a video recording was made while patients performed standard UPDRS Motor “off” (off medications) activities. When patients entered the hospital at the 12 month evaluation, they were immediately asked to rate themselves using the GRS. Patients then watched the video recorded prior to surgery and rated themselves on the GRS a second time.

Outcome Variable

Global Rating Scale¹: The GRS is a one item scale with scores ranging from -3 (“parkinsonism much worse since surgery”) to 0 (“no change”) to +3 (“parkinsonism

much improved since surgery”). No reliability or validity information is available for this scale. Absolute blindness as to which surgical procedure was performed is essential for the GRS to be a reliable scale. To make sure how effective the blindness of the surgical procedure had been, all participants had to indicate which procedure they thought s/he had received—tissue implants or sham procedure within 7 days of the surgery (after they were home) and then quarterly preceding the evaluations by the investigators. The investigators were not given knowledge of the results of the GRS evaluations, which were mailed to the unblinded biostatistician.

Statistical Analyses

Data were analyzed using paired-samples t tests to examine differences between GRS ratings before and after viewing the video. Independent samples t tests were used to investigate group differences between scores in the actual and perceived treatment groups both before and after viewing the video.

RESULTS

For the total sample, results indicated there was a change in GRS ratings from pre- to post-video ($t = -3.47$; $P = .001$). The mean rating before the video was $-.08$ ($SD = 1.88$) and after the video the mean was $.46$ ($SD = 1.71$), indicating improved ratings after viewing the video. Among the 39 responses (one person died in an automobile accident several months after the surgery), 16 (41%) rated themselves higher after the video, 20 (51%) remained the same, and three (8%) reported lower ratings.

Analyses were conducted to determine whether there were differences in GRS ratings between those who received the *actual* implant and sham surgery. Results

indicated there were no differences in scores between the two groups before or after viewing the video, nor in change scores (Table 1).

Based on *perceived* treatment, or type of surgery patients *thought* they received at 12 months, there were differences in the average GRS scores of the perceived treatment groups both before and after the video; there was a marginally significant difference in the change scores (Table 2). Patients who *thought* they received the implant reported higher scores at both time points than those who *thought* they received the sham surgery, regardless of the actual type of surgery they received.

In order to gain a clearer understanding of which patients' scores changed as a result of watching the video, the sample was divided into the following subgroups: 1) received implant/perceived implant, 2) received implant/perceived sham, 3) received sham/perceived implant, and 4) received sham/perceived sham. Results shown in Table 3 indicate the number of participants in each subgroup whose scores improved, remained the same, or declined as a result of watching the video. Group 2 (received implant/perceived sham) clearly demonstrated more positive revision after viewing the video than the other three groups ($t = -2.07$; $P = .046$).

Figure 1 presents a plot of scores for the four subgroups before and after the video. Although scores for Groups 1, 2, and 4 all improved, significant improvement was found only in Group 2 ($t = -3.07$; $P = .011$), who received the implant but *thought* they received sham surgery at 12 months. Viewing the baseline video apparently affected the group's perception of change in their condition since surgery, with the average score improving from -1.33 before the video to -.33 after the video. Eight of 12 persons revised their GRS scores upward and one person revised it downward. Scores for Group 3, who

received sham surgery but *thought* they received the implant, were the same before and after the video; one person revised the GRS score upward and one revised it downward. Because they had received sham surgery, it is probable their condition would have declined over the one year period of the double-blind, yet their average GRS score remained the same. To gain a better understanding of what contributed to the stable scores of Group 3, we examined perceived treatment across time (7 days, 4, 8, and 12 months) for this group (Figure 2). Unlike the other three groups, all patients in Group 3 thought they received the implant at each time, indicating they believed they received the actual implant from the very beginning of the study.

DISCUSSION

The primary aim of this study was to determine whether there were changes in patient GRS scores at 12 months following review of a videotape of themselves performing Motor “off” activities at baseline. Results indicated that scores for the sample were significantly higher after the video. As no physical changes could have occurred during the brief time patients watched the video, we considered several explanations for the improvement in scores among 16 of 39 participants, six of whom received sham surgery (see Table 3). It is possible that patients were influenced by comments or social cues from others in the room when viewing the video. In addition to nurse coordinators, who were specifically instructed to not respond to the video, family members were sometimes present during the tape review. Alternatively, patients may have been influenced by their own desire to have improved since surgery, or to be seen as “successful patients” for the medical staff.

Cognitive theory suggests that many individuals have a tendency to distort their abilities and performance compared to perceptions of observers, and watching a videotape may have allowed patients to take on the perspective of an observer^{4,5}. Theoretically, with increased information related to baseline performance, self-perceptions of some participants apparently changed and ratings were adjusted.

It is also possible that “response shift,” or recalibration of one’s own internal standards of well-being over time, may have occurred⁶. Among persons experiencing chronic illness, perceptions of quality of life often gradually change as individuals adapt to progressive symptoms. Viewing the video may have provided patients with objective evidence relative to perceptions of their physical status before surgery, and thus, ratings of some individuals were revised.

The second aim of the study was to determine whether there were differences in GRS scores between those who received the *actual* and *perceived* implant and sham surgeries. Similar to previous results of comparisons of *actual* implant and sham groups with this sample^{1, 2, 7-11}, there were no differences between the *actual* groups on GRS scores either before or after the video (Table 1). Although scores of both groups improved slightly after the video review, effects of surgery were apparently not clear enough at 12 months to allow patients to determine how different they were as a result of surgery.

Results based on *perceived* treatment revealed differences between those who *thought* they received the implant and those who *thought* they received the sham both before and after the video (Table 2). These results offer another example of the strong

placebo effect found in previous results with this sample wherein more differences were found between the *perceived* treatment groups than among *actual* treatment groups^{2, 8-11}.

Because pre- and post-video changes and differences between groups were found, the third aim of the study was to examine the data based on four subgroups of the sample as described above. As shown in Figure 1, those who *thought* they received the implant (Groups 1 and 3) reported higher scores both before and after the video than those who *thought* they received the sham surgery (Groups 2 and 4), regardless of the actual treatment they received. Although scores for Groups 1, 2 and 4 improved, the score for Group 2 (received implant/perceived sham) was the only one that improved significantly. Of the two groups who received the implant (Groups 1 and 2) and may legitimately have experienced some improvement from the surgery, the significant increase in scores for Group 2 may be explained in part by the relatively low (-1.33) pre-video score, which allowed a broad range on the scale (-3 to +3) for improvement. By contrast, the average score for Group 1 was on the upper end of the GRS (+1.86) before the video, leaving little room for improvement. Scores for Group 4 (received sham/perceived sham) improved slightly from pre- to post-video. Although they received sham surgery, viewing the video apparently countered some of the negative perceptions that contributed to the initial low ratings of the group; 39%, or 5 of 13 scores improved after the video (Table 3).

Scores of Group 3 (received sham/perceived implant) were very similar to scores of Group 1 (received implant/perceived implant) and did not change from pre- to post-video (see Figure 1). Thus, the placebo effect was very evident in Group 3 as they were clearly more influenced by perceived treatment at 12 months than by objective data

(baseline video). Results in Figure 2 showing perceived treatment across time also suggest the placebo effect was very strong in this group.

A major premise of this video review study was that viewing a baseline video at 12 months would help patients recall their pre-surgical status and contribute to a more accurate rating of change since surgery. Although we know that ratings on the GRS improved after the video, we do not know that they became “more accurate.” Previous research has suggested that showing patients a video of their current status and then viewing the baseline video might result in a more accurate perception. This has been shown for general or specific self-image; e.g., with regard to psychopathological symptoms^{12, 13} or social behavior¹⁴.

Having a current video as an objective reference point potentially balances the self-perceptions of current functioning, which may be influenced by distorted cognitions⁵ or the response shift⁶. This is an important area to explore in order to establish the value or influence of video review on patient self-ratings. It also raises the question of whether having an accurate sense of one’s functioning is as important as what one perceives.

Limitations

The most obvious limitation of this study is sample size, which becomes more challenging when dividing the sample into smaller subgroups. However, the opportunity to examine such a unique sample in a study which resulted in such a strong placebo effect is theoretically important even if we can only derive results that are suggestive of profitable further investigation in larger clinical trials.

Conclusions

Video review of baseline status appears to be most valuable to participants who received the implant but had perceived they received the sham surgery. Actually observing their baseline state via the video may have put a more realistic perspective of the changes encountered 12 months after implant surgery. We conclude that in future trials where self-ratings are an outcome variable that participants be allowed to review baseline videos of themselves before making a comparative rating of improvement or worsening.

Results confirmed that the placebo effect was very strong in this experimental study. Because of the small sample size, it was particularly impressive to find statistically significant results, underscoring the power of the placebo in this group. Results indicated the added value of analyzing data in placebo studies based on a paradigm of four subgroups (received treatment/perceived treatment, received treatment/perceived sham, etc.). It is possible that our knowledge about the placebo effect can be broadened by examining these smaller, more specific groups over the period of the double-blind.

Acknowledgements:

The authors would like to acknowledge the courage and perseverance of all participants in this extraordinary experimental sham surgery controlled trial.

Ethical Compliance Statement:

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this work is consistent with those guidelines. This study was

approved by the IRB at the University of Denver, #560812-3. Informed consent was obtained.

Authors' Roles

McRae: Specific Research Project: Conception, Organization, Execution

Statistical Analysis: Design, Execution

Manuscript: Writing first draft, Review and Critique

Caspari: Research Project: Conception

Manuscript: Review and Critique

Russell: Research Project: Conception

Statistical Analysis: Design, Review and Critique

Manuscript: Review and Critique

Ellgring: Research Project: Conception,

Manuscript: Writing first draft, Review and Critique

Bezzant: Research Project: Execution

Manuscript: Review and Critique

Greene: Research Project: Execution

Manuscript: Review and Critique

Fahn: Research Project: Conception, Organization, Execution

Manuscript: Review and Critique

Funding Sources, Conflicts of Interest and Financial Disclosures for preceding 12 months:

McRae: Funding as PI for this research was received from NINDS for the original study from 1995-2000. This author declares that there are no conflicts of interest relevant to this work. Funding sources in the previous 12 months not related to the current research are Parkinson Foundation and Project Spark (consultant).

Caspari: No specific funding was received for this work. This author declares that there are no conflicts of interest relevant to this work. The author declares that there are no additional disclosures to report.

Russell: Funding as a consultant for this research was received from NINDS for the original study from 1995-2000. This author declares that there are no conflicts of interest relevant to this work. Funding source in the previous 12 months not related to the current research is United Health Care (consultant).

Ellgring: Funding as a consultant for this research was received from NINDS for the original study from 1995-2000. This author declares that there are no conflicts of interest relevant to this work. The author declares that there are no additional disclosures to report.

Bezzant: A summer stipend as a research assistant was received for this work. This author declares that there are no conflicts of interest relevant to this work. The author declares that there are no additional disclosures to report.

Greene: Funding for this research was received from NINDS for the original study from 1994-1999. This author declares that there are no conflicts of interest relevant to this work. A funding source in the previous 12 months not related to the current research was an honorarium from Johns Hopkins University Medical Center.

Fahn: Funding for this research was received from NINDS for the original study from 1994-1999. This author declares that there are no conflicts of interest relevant to this work. Funding sources in the previous 12 months not related to the current research include:

Commercial Consulting and Advisory Board Membership with honoraria: Adamas Pharmaceuticals; Retrophin Inc (DMC member); PhotoPharmics; Sun Pharma Advanced Research Co., Ltd; Kashiv Pharma

Grants/Research Support: Smart Family Foundation

Lecture Honoraria: Movement Disorder Society (Aug 2017, Oct 2017), Univ of Cincinnati (Feb 2018); Colombian Association of Neurology (Mar 2018), Univ of Utah (Mar 2018), Houston Methodist Hospital (April 2018); Parkinson Voice Project in Dallas (April 2018)

Editor Honoraria: Springer Publishers for serving as co-editor of Current Neurology and Neuroscience Reports (annual)

Author Royalties: Elsevier Publishers for co-authorship of book *Principles and Practice of Movement Disorders*.

Employment: Columbia University

REFERENCES

1. Freed CR, Greene PE, Breeze RE et al. Transplantation of embryonic dopamine neurons for severe Parkinson's disease. *N Engl J Med.* 2001; 344: 710-719.
2. McRae C, Cherin E, Yamazaki G et al. Effects of perceived treatment on quality of life and medical outcomes in a double-blind placebo surgery trial. *Arch Gen Psych.* 2004; 61:1-9.
3. Fahn S, Elton RL, Members of the UPDRS Development Committee. The Unified Parkinson's Disease Rating Scale. In Fahn S, Marsden CD, Calne DB, Goldstein M, eds: Recent Developments in Parkinson's Disease, Vol. 2. Florham Park, NJ: Macmillan Healthcare Information, 1987. pp. 153-163, 293-304.
4. Beck AT, Rush AJ, Shaw BF, & Emery G. *Cognitive Therapy of Depression.* New York: Guilford; 1979.
5. Beck JS. *Cognitive Therapy: Basics and Beyond.* New York: The Guilford Press; 1995.
6. Sprangers MA & Schwartz CE. Integrating response shift into health-related quality of life research: A theoretical model. *Soc Sci Med.* 1999; 48(11):1507-15.
7. Trott CT, Fahn S, Greene P et al. Cognition following bilateral implants of embryonic dopamine neurons in PD: A double blind study. *Neurology.* 2003; 60: 1938–1943.
8. McRae C, Fields A, Russell D, et al. Contributors to the placebo effect in a 12 month double-blind surgery trial. *Invited address: Aberdeen Royal Infirmary.* Scotland: March, 2005.

9. McRae C, Fazio E, Pinarowicz J et al. Predictors of the placebo effect at 4, 8, and 12 months in the double-blind sham surgery trial for the treatment of Parkinson's disease. *Poster presentation: Twenty Second Annual Symposium on Etiology, Pathogenesis, and Treatment of Parkinson's Disease and Other Movement Disorders.* Salt Lake City: September, 2008.
10. McRae C, Fazio E, Hartsock G, et al. Does the placebo effect extend to caregivers? *Paper presentation: Twenty Second Annual Symposium on Etiology, Pathogenesis, and Treatment of Parkinson's Disease and Other Movement Disorders.* Salt Lake City: September 2008.
11. McRae C. Longitudinal study of neural tissue implantation for treatment of Parkinson's disease: Effects on quality of life from 1995 to 2013. *Invited address: The 20th meeting of the German Society of "Psychology and Parkinson's Disease."* Munich, Germany: April, 2015.
12. Biggs S, Rosen B, & Summerfield, A. Video-feedback and personal attribution in anorexic, depressed and normal viewers. *Br J Med Psychol.* 1980; 53:249-259.
13. Griffiths R, & Gillingham P. The influence of videotape feedback on the self-assessments of psychiatric patients. *Br J Psychiatry.* 1978; 133:156-161.
14. Harve A, Clark D, Ehlers A, & Rapee R. Social anxiety and self-impression: cognitive preparation enhances the beneficial effects of video feedback following a stressful social task. *Behav Res Ther.* 2000; 38(12):1183-1192.

Figure 1. Plots of pre- and post-video GRS scores for four groups.

Figure 2. Plot of perceived treatment at 7 days, 4, 8, and 12 months for four groups

Accepted Article

TABLE 1. Differences in Global Rating Scores in Implant and Sham Groups (n = 39)

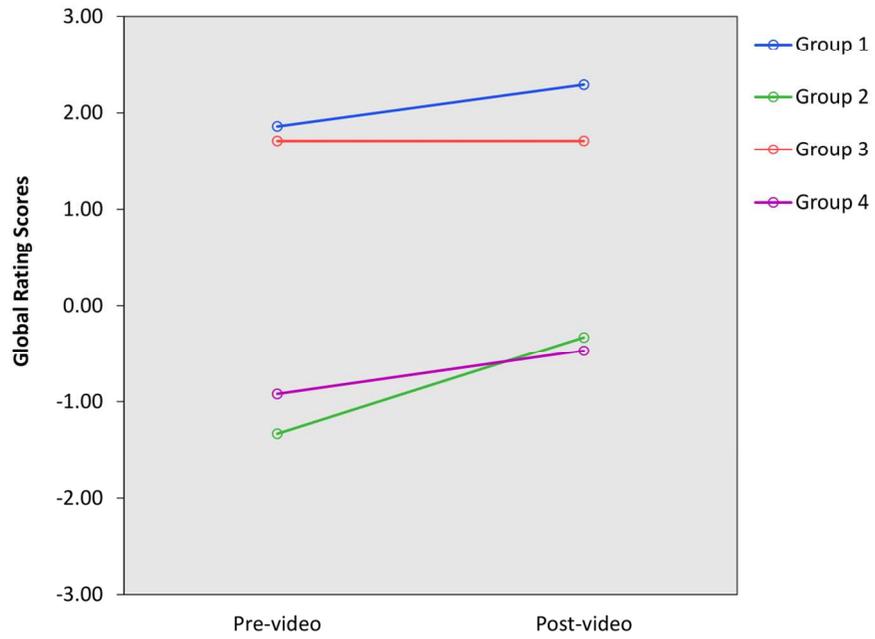
	Implant (n = 19)		Sham (n = 20)		
	M	SD	M	SD	P
Before	-.16	2.09	.00	1.72	.80
After	.63	1.86	.30	1.59	.55
Δ Score	.79	1.03	.30	.86	.12

TABLE 2. Differences in Global Rating Scores in Perceived Treatment Groups (n = 39)

	Implant (n = 14)		Sham (n = 25)		
	M	SD	M	SD	P
Before	1.79	1.37	-1.12	1.20	<.001
After	2.00	1.11	-.40	1.35	<.001
Δ Score	.21	.70	.72	1.06	.082

TABLE 3. Change in scores on GRS from pre- to post-video at 12 months

Groups	Positive Revision of Judgment (+)	No Change (0)	Negative Revision of Judgment (-)	N	% Who Improved
Implant/Perceived implant	2	5	0	7	29%
Implant/Perceived sham	8	3	1	12	67%
Sham/Perceived implant	1	5	1	7	14%
Sham/Perceived Sham	5	7	1	13	38%
Total	16	20	3	39	



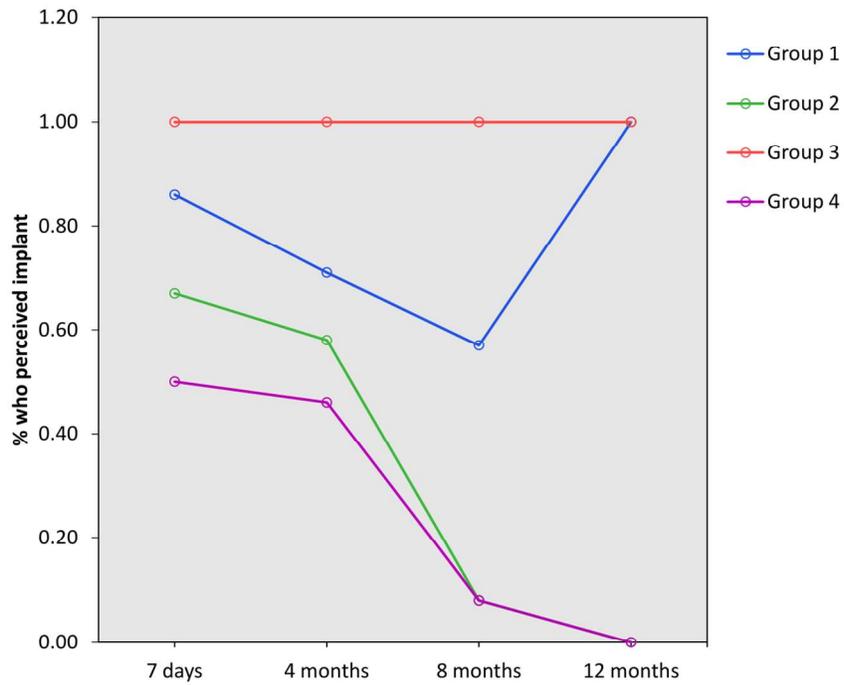
Group 1 = Received implant, thought implant at 12 months

Group 2 = Received implant, thought sham at 12 months

Group 3 = Received sham, thought implant at 12 months

Group 4 = Received sham, thought sham at 12 months

Figure 1. Plots of pre- and post-video GRS scores for four groups



Group 1 = Received implant, thought implant at 12 months

Group 2 = Received implant, thought sham at 12 months

Group 3 = Received sham, thought implant at 12 months

Group 4 = Received sham, thought sham at 12 months

Figure 2. Plot of perceived treatment at 7 days, 4, 8, and 12 months for four groups