Maintaining the blind in sham controlled interventional trials: lessons from the EPISOD study

Authors
Peter Cotton[1], Qi Pauls[2], April Wood[1], Valerie Durkalski-Mauldin[2]

Institutions
1   Digestive Disease Center, Medical University of South Carolina, Charleston, South Carolina, United States
2   Department of Public Health Science, Medical University of South Carolina, Charleston, South Carolina, United States

Corresponding author
Peter Cotton, Professor of Medicine, Digestive Disease Center, Medical University of SC, 114 Doughty Street, Suite 249, MSC 702, Charleston, SC, 29425, USA
Fax: +1-843-849-6862
Petercotto@gmail.com

ABSTRACT

Objective and study aims This study was designed to demonstrate the techniques used and the effectiveness of blinding in the EPISOD study (Evaluating Predictors and Interventions in Sphincter of Oddi Dysfunction). This was a large sham-controlled trial evaluating the efficacy of endoscopic sphincterotomy treatment for patients with suspected sphincter of Oddi dysfunction.

Patients and methods We describe the methods intended to ensure that the subjects, caregivers and those assessing the outcomes were blinded to the treatment allocation and systematically assess the success of subject blinding procedures using the Bang's Blinding Index (BI) for each treatment arm as an indicator of potential unblinding.

Results Blinding procedures proved to be acceptable and adhered to by the study team at each site. The BI indicated "wishful thinking" by the subjects regardless of treatment assignment, even when they were confident in their opinions.

Conclusion We conclude that it is possible to design and maintain a system for blinding the treatment allocation in a sham-controlled interventional study. Treatment guess plus confidence in the guess should be collected to examine the success of blinding procedures. The EPISOD study provides a blueprint for future sham-controlled trials in endoscopy.

Introduction

The randomized controlled study is the gold standard for definitive evaluation of therapeutic interventions. Central to the concept is that patients, caregivers, and those evaluating the outcomes are blinded to the treatment assignment. This is relatively easy to arrange when evaluating medications, because placebo pills can be provided that look identical to the active medication. The situation is much more complex and many practical issues arise when dealing with surgical or endoscopic interventions. The main difference is that, in order for them to be blinded to the treatment, patients in the sham arm of such studies are subjected to the inconvenience and potential risks (and possible scars) of an invasive procedure without any immediate benefit. The ethical issues in this dilemma have been

argued strongly, but most authorities call for more such studies, albeit with important safeguards [1 – 6]. Supporting that call is the fact that a review of 53 placebo-controlled surgical studies found that half of them showed no benefit for surgery over the sham procedure [7]. Similar conclusions have been drawn when examining the practicality of performing sham-controlled trials of endoscopic interventions [8].

The EPISOD study (Evaluating Predictors and Interventions in Sphincter of Oddi Dysfunction) was a large National Institutes of Health-funded multicenter sham-controlled clinical trial which showed that endoscopic sphincterotomy was not superior to sham treatment in terms of reducing pain in patients with suspected Sphincter of Oddi dysfunction [9]. We describe the steps taken to maintain the treatment blind when planning

and executing the trial, and report the success of the blinding procedures.

Patients and methods

Patients with burdensome biliary-type pain after cholecystectomy and no definite evidence for biliary pathology were invited to participate in the study involving seven medical centers in the United States. Those eligible and consenting (a total of 214 subjects) all underwent endoscopic retrograde cholangiopancreatography (ERCP) with sphincter of Oddi manometry, under standard sedation or anesthesia. After successful performance of manometry, they were randomized to sphincterotomy or to no therapeutic intervention (in a 2:1 allocation). All subjects received a temporary pancreatic stent to reduce risk of pancreatitis. Success was defined at the 12-month visit post-randomization as a subject reporting fewer than 6 days of disability due to their abdominal pain during a 90-day period (during months 10–12). This self-reported outcome was measured by the RAPID (Recurrent Abdominal Pain Intensity and Disability) score, an instrument initially developed and validated for the EPISOD study [10].

Numerous protocol-specified steps were taken to ensure that the subjects, their caregivers, and research staff remained blinded to the treatment allocation in the immediate and later follow-up periods. In summary, the research coordinators in the procedure room who supervised the randomization and documented the procedure were not involved in future assessments of the subject's progress; details of the actual treatment performed were sealed in the research records and the subject's routine medical records indicated that the patient had undergone ERCP, manometry and temporary stenting, and "may also have had biliary or dual sphincterotomy"; subjects were not billed for the procedures or any overnight hospital stay, these costs were funded by the grant; and, the research coordinators calling the subjects each month were blinded. It was anticipated that a small portion of subjects would need upper endoscopy to remove their temporary pancreatic stent if it had not passed spontaneously as planned. Endoscopists performing these procedures were asked not to comment on the appearance of the papilla. To reduce bias in assessment and management of subjects who were unhappy with their progress, the study included support for each site to have an independent "evaluating physician" who had not been involved in the initial therapy and was unaware of the treatment arm to as-

sess these returning subjects. Subjects and their outside caring physicians were told that they could be informed of treatment assignment in an emergency, and a telephone "hot line" was provided.

Subjects and the research coordinators were asked at months 1, 3, 6, 9, and 12 post-randomization to provide their "best guess" of the treatment allocation and to provide the confidence level of their guess (a five-point scale ranging from "Not at all" to "Extremely"). Repeated measures were collected to capture changes in the "guess" during the long-term follow up. The effectiveness of blinding was measured using a blinding index (BI) that ranges from –1 to 1 and measures the treatment-specific proportion of unblinded subjects taking into account the confidence in the guess [11, 12]. A value of 0 indicates "random guessing" and successful treatment blinding, a positive value indicates correct guessing of the treatment assignment and a negative value indicates incorrect guesses. When the BI values for each arm are symmetric around 0 ($BI_{Shpc} = -BI_{sham}$), the blinding can be considered "wishful Thinking." In addition to estimating the BI, potential predictors of subjects correctly guessing the assigned treatment were examined, including assigned treatment arm, confidence in the guess, change in RAPID score from baseline, the RAPID score at the specific visit, the treating site, specific visit (months 3, 6, 9, and 12) and the interaction between RAPID score and treatment arm. These potential predictors were examined using a generalized linear model for the binary outcome of correct guess, accounting for repeated measures within a subject. All analyses were conducted in SAS Version 9.3 (SAS, Cary NC).

The protocol for the EPISOD trial was approved by Institutional Review Boards at all participating sites, and all subjects gave informed consent.
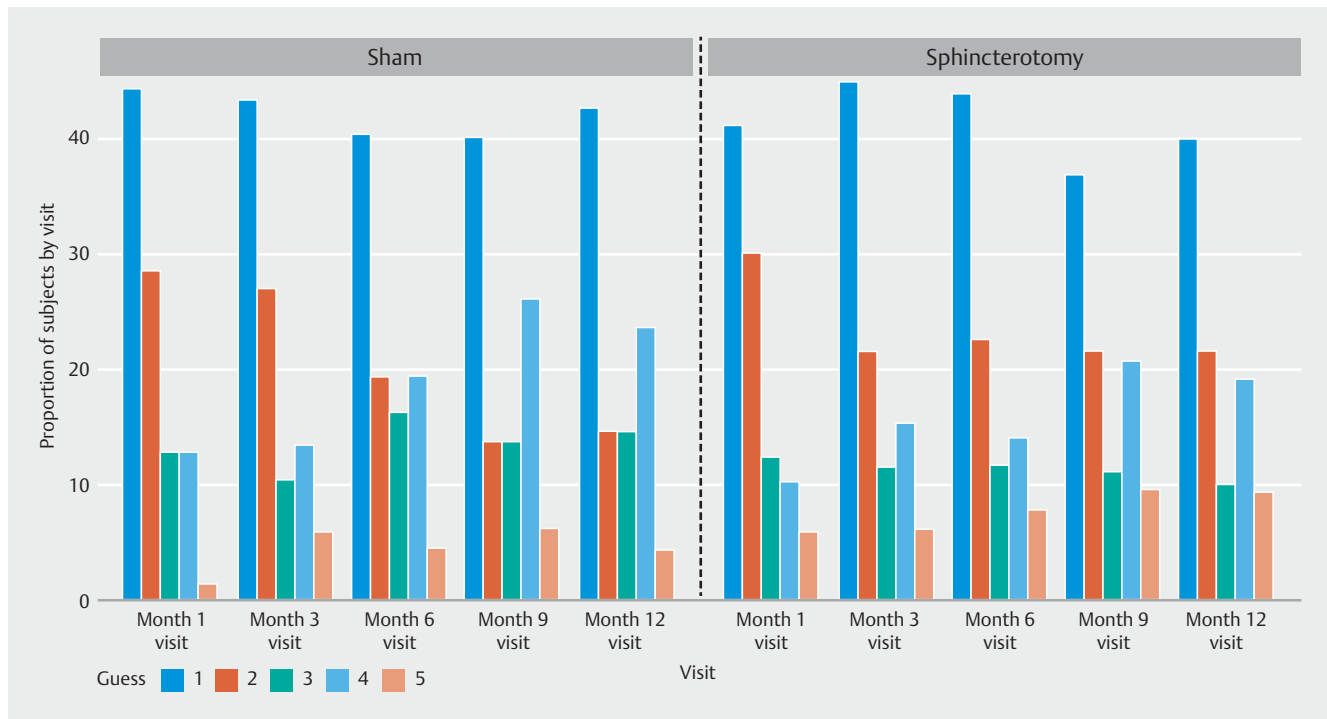
Results

One subject who suffered retroduodenal perforation after the ERCP procedure was thereby unblinded. No additional cases of unblinding of subjects, caregivers or research staff in the immediate post-procedure period were reported. Treatment allocation was requested during follow-up in one case by a treating physician, and was provided. There were no calls to the hotline.

The blinding questionnaire was captured on 213 of the 214 randomized EPISOD subjects. Each of the follow-up visits had a minimum of 190 completed questionnaires. ▶ **Table 1** illus-

▶ **Table 1** Number of subjects by treatment assignment and subject's guess for the 1-month post-randomization visit and the 12-month visit.

	Guess			
	1-month		**12-month**	
Assignment	Sphincterotomy	Sham	Sphincterotomy	Sham
Sphincterotomy	101	35	87	43
Sham	52	18	40	28
Total	153	53	127	71

Cotton Peter et al. Maintaining the blind … Endoscopy International Open 2019; 07: E1322–E1326

E1323

▶ **Fig. 1** The proportions of subjects and confidence in the guess by visit. Guess 1: Strongly believe the treatment is sphincterotomy, 2: Somewhat believe the treatment is sphincterotomy, 3: Somewhat believe the treatment is sham, 4: Strongly believe the treatment is sham, 5: Don't Know.

trates the assignment and subject's guess for the 1-month and 12-month visits. Overall, both the subjects and the research coordinators more often made the determination that they were assigned to the sphincterotomy arm at each visit. ▶ **Fig. 1** illustrates the number of subjects and confidence in the guess by visit for each treatment arm. Regardless of the visit and treatment arm, it is shown that the majority of subjects strongly believed they were assigned to the sphincterotomy arm. Subjects responded as "extremely" confident in their determination an average of 28% over the visits. However, the accuracy of these extremely confident cases was only 60%. For the coordinators, fewer than 5% of all responses were rated "extremely" confident; of these 76% were accurate. For those less than "extremely confident" in their guess, they were accurate 57% of the time.

Site and visit did not have any association with correctly guessing treatment assignment. Regardless of treatment received, subject's responses were strongly influenced by how many days of disability they reported at that time, as shown in ▶ **Fig. 2**. When their pain-related disability (RAPID score) was high, subjects more frequently guessed that the sham arm was their treatment assignment. When their RAPID scores were low, subjects more often responded that they had undergone sphincterotomy. This trend did not change when basing it on the change in the days of disability from baseline.

▶ **Table 2** illustrates the blinding index (BI) by treatment arm for subjects. The BI estimate for the sphincterotomy arm at 1 month is 0.397 (95%CI: 0.29, 0.50), indicating a significant amount of correct guesses. The sham arm BI at 1-month is −0.396 (95%CI: 0.-0.54, −0.27), indicating a significant amount of incorrect guesses. Because the indices are approximately symmetric around 0, the BI values indicate "wishful thinking." When contrasting the 1-month BIs to the 12-month BIs, the BIs at the end of the trial indicate that subjects are making more random guesses at treatment assignment particularly in the sham arm.

Discussion

Because of the strong placebo effect of surgical and endoscopic therapeutic interventions, it is clear that maintaining the treatment blind is essential if the results of a sham-controlled trial are to be accepted as valid and reliable. This applies especially in studies with soft endpoints, such as pain control [5, 8]. Blinding is not difficult to achieve when two treatments are being compared through the same entry point, for example when comparing two endoscopic hemostatic techniques in patients with gastrointestinal bleeding. Then it is necessary only for the endoscopist and staff involved in the procedure not to disclose the specific treatment, not to be involved in the subsequent assessments, and for the patient not to be unblinded by clinical reports or bills for specific instruments.

The situation is more challenging when an accepted treatment is being compared to a sham intervention, because the intervention must appear identical to the patients and to those doing the outcome assessments. This adds ethical issues, which have been widely and strongly argued [1–6]. Blinding can be more difficult to achieve or be compromised if the active treat-
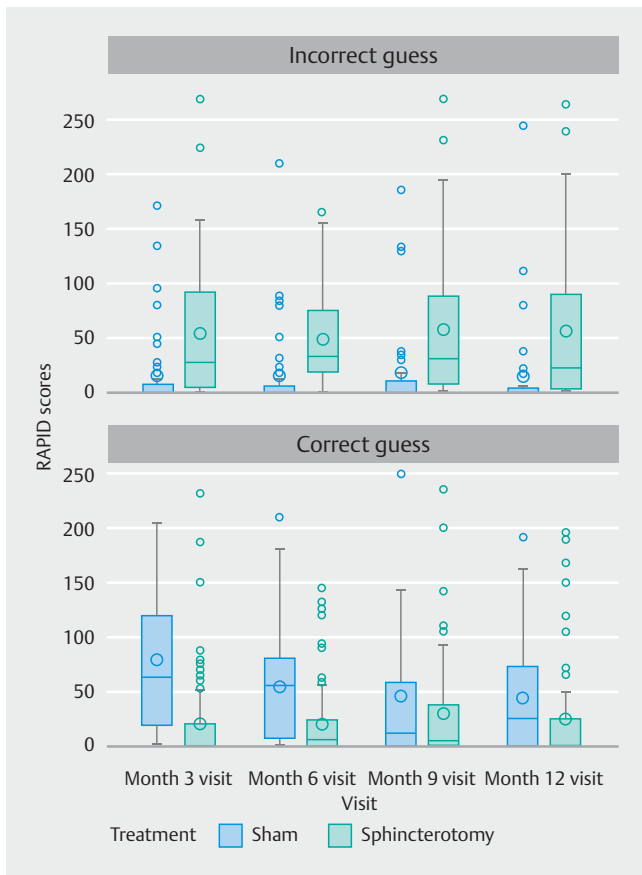
▶ **Fig. 2** Disability days (RAPID Score) by accuracy of treatment determination by the subject, by assigned treatment arm and study visit

ment causes pain (such as after mucosal ablation), or has common predictable effects (such as early satiety after bariatric procedures, or dysphagia after fundoplication) [13]. Pancreatitis was a common adverse event in the EPISOD study, but the incidence was the same in both treatment arms. However, the patient who suffered retro-duodenal perforation (and her caregivers) obviously realized that she had undergone sphincterotomy.

Blinding has been an essential element in dozens of sham-controlled surgical and endoscopic trials [8–11], but we have not found reports of the precise details of how this has been designed, or the rates of success. Furthermore, few trials have queried patients and research staff about their best guesses. Randomization and blinding are acceptable only if approved by Institutional Review Boards, and applicable only if patients understand and consent.

Blinding proved to be acceptable to subjects, as illustrated by the fact that only 5% of those potentially eligible declined enrollment for that specific reason. Equally important, the methods proved to be effective, in that there were no unexplained cases of unblinding. Only one subject was unblinded due to a perforation. The 2:1 allocation (sphincterotomy: sham) may have improved acceptance rates but this is difficult to prove. The BI estimates support the conclusion of successful blinding in the trial and indicate that subjects had "wishful thinking." Because of the symmetric nature of the BI estimates between the two treatment arms and the 2:1 allocation, it is reasonable to conclude that the EPISOD trial may have experienced response bias such that randomized subjects more often believed that they received a sphincterotomy. A potential response bias is also evident in the accuracy of the guess and the RAPID score. It is not surprising that subjects who reported more disability guessed that they received the sham treatment arm, or that subjects that reported less disability guessed the sphincterotomy arm. As Bang and others have pointed out in their theoretical work on the BI, an important point is that "wishful thinking" or "random guess" are both "ideal blinding scenarios" that incur minimal bias associated with belief about allocation [11, 12].

Conclusion

This report shows that it is possible to design and maintain a system for blinding the treatment allocation in a sham-controlled interventional study, and provides a blueprint for future trials.

▶ **Table 2** Blinding Index for subjects.

Visit	BI for TX	95% CI (Lower)	95% CI (Upper)	BI for Sham	95% CI (Lower)	95% CI (Upper)	Total # sham arm	Total # TX arm
Month 1 visit	0.397	0.295	0.499	−0.396	−0.537	−0.256	70	136
Month 3 visit	0.341	0.243	0.440	−0.381	−0.524	−0.238	67	129
Month 6 visit	0.352	0.251	0.452	−0.228	−0.361	−0.095	67	128
Month 9 visit	0.220	0.120	0.320	−0.131	−0.258	−0.003	65	125
Month 12 visit	0.269	0.171	0.367	−0.188	−0.313	−0.062	68	130

BI, blinding index; TX, treatment; CI confidence interval.

Cotton Peter et al. Maintaining the blind … Endoscopy International Open 2019; 07: E1322–E1326

E1325

Competing interests

None

References

[1] Horng S, Miller FG. Is placebo surgery unethical? N Engl J Med 2002; 347: 137 – 139

[2] George AJT, Collett C, Carr AJ et al. When should placebo surgery as a control in clinical trials be carried out? Bull R Coll Surg Eng 2016; 98: 75 – 79

[3] Polgar S, Ng J. Ethics, methodology and the use of placebo controls in surgical trials. Brain Res Bull 2005; 67: 290 – 297

[4] Miller FG. Sham surgery: an ethical analysis. Sci Eng Ethics 2004; 10: 157 – 166

[5] Miller FG, Wendler D. The ethics of sham invasive intervention trials. Clinical Trials 2009; 6: 401 – 402

[6] Gelijns AC, Ascheim DD, Parides MK et al. Randomized trials in surgery. Surgery 2009; 145: 581 – 587

[7] Tenery R, Rakatansky H, Riddick FA et al. Surgical "placebo" controls. Ann Surg 2002; 235: 303 – 307

[8] Wilcox CM. Exploring the use of sham design: implications for endoscopic research. Gastrointest Endosc 2008; 67: 123 – 127

[9] Cotton PB, Durkalski V, Romagnuolo J et al. A multicenter, randomized trial of endoscopic sphincterotomy for suspected sphincter of Oddi dysfunction in patients with pain after cholecystectomy – the EPISOD trial. JAMA 2014; 311: 2101 – 2109

[10] Durkalski V, Stewart W, MacDougall P et al. measuring episodic abdominal pain and disability in suspected sphincter of Oddi Dysfunction. World J Gastroenterol 2010; 21: 4416 – 4421

[11] Bang H, Ni L, Davis C. Assessment of blinding in clinical trials. Controlled Clin Trials 2004; 25: 143 – 146

[12] Bang H. Random guess and wishful thinking are the best blinding scenarios. Contemp Clin Trials Comm 2016; 3: 117 – 121

[13] Schwartz MP, Wellink H, Gooszen HG et al. Endoscopic gastroplication for the treatment of gastro-oesophageal reflux disease: a randomized, sham-controlled trial. Gut 2007; 56: 20 – 28