

# Brief Communication: Better Ways To Question Patients about Adverse Medical Events

## A Randomized, Controlled Trial

Stephen Bent, MD; Amy Padula, MS; and Andrew L. Avins, MD, MPH

**Background:** There is no standard method of identifying adverse events in clinical trials.

**Objective:** To determine whether 3 different methods of questioning patients about adverse events in a clinical trial affect the frequency of reported events.

**Design:** Randomized, single-blind, controlled trial.

**Setting:** A Veterans Administration medical center, San Francisco, California.

**Participants:** Men 50 years of age or older who had benign prostatic hyperplasia.

**Measurement:** Frequency of self-reported medical problems.

**Intervention:** The authors randomly assigned 214 men who were undergoing a 1-month, single-blind, placebo run-in period during an existing clinical trial to 3 groups to test different self-administered methods of assessing medical problems at the end of the run-in period. The first group was asked an open-ended question; the second group was asked an open-ended, defined question; and the third group was given a checklist of 53 common side effects.

**Results:** All 214 patients completed the study. Patients assigned to the checklist group reported a total of 238 adverse events; in comparison, patients who were asked an open-ended question or an open-ended, defined question reported 11 and 14 adverse events, respectively ( $P < 0.001$ ). The percentage of patients reporting any adverse event was also much higher in the group assigned to the checklist (77%) than in the first group (14%) or second group (13%) ( $P < 0.001$ ).

**Limitations:** The study included only relatively healthy, well-educated, middle-aged men and assessed only self-reported medical problems after the participants had taken placebo for 1 month. All personnel overseeing the study were aware of the group assignments.

**Conclusions:** Different methods of collecting patient data regarding adverse events lead to large differences in the reported rates of adverse events in clinical trials, potentially reducing the validity of comparisons between the side effect profiles of drugs and other interventions.

*Ann Intern Med.* 2006;144:257-261.

For author affiliations, see end of text.

[www.annals.org](http://www.annals.org)

Currently, there is no standard method for identifying adverse events that occur during a clinical trial. Although regulatory agencies (such as the U.S. Food and Drug Administration) require that studies of new drugs report adverse events in a standard way, they do not specify a standard method for ascertaining these data (1). Consequently, how individual studies identify adverse events varies considerably. For example, early studies of nonsteroidal anti-inflammatory drug-induced gastric ulcers reported much lower frequencies of ulcers than more recent studies, mostly because researchers have recently made greater efforts to detect this side effect (2). The implications of this lack of consistent ascertainment methods are substantial; comparisons of rates of reported side effects from 2 or more drugs may not be valid if the methods of collecting adverse events differ. This could impair the ability of patients and physicians to compare the risk-benefit profile of drugs. We therefore conducted a randomized, controlled trial to determine whether different methods of identifying adverse events in a clinical trial would lead to different estimates of the frequency of these events.

## METHODS

### Study Design

The study protocol and all procedures were approved by the Committee on Human Research at the University

of California, San Francisco. The study, which took place between April 2002 and April 2005, was a randomized, single-blind, controlled trial that assigned patients to 3 groups to test self-administered methods of assessing medical problems that they experienced while taking a placebo for 1 month.

### Participants

We recruited participants from a larger study that was examining the safety and efficacy of the herb saw palmetto for treatment of benign prostatic hyperplasia (3). The trial, known as the STEP (Saw Palmetto Treatment for Enlarged Prostates) study, required that participants be 50 years of age or older, have moderate to severe symptoms of benign prostatic hyperplasia, and have no serious comorbid illness. All participants in the study gave informed consent; were

See also:

#### Print

Editors' Notes . . . . .	258
Editorial comment . . . . .	298
Related letter . . . . .	306

#### Web-Only

Conversion of figure and tables into slides

**Context**

Investigators use diverse methods to assess the adverse events experienced by study participants.

**Contribution**

During a 1-month placebo run-in period of a clinical trial, this single-blind substudy randomly assigned 214 men with benign prostatic hyperplasia to 3 groups to test different methods of asking about recent medical problems. Men who completed a checklist about 53 common side effects reported many more problems than participants in the 2 groups that were given different formats of open-ended questions. For example, 77% of the checklist group reported 1 or more medical problems, compared with 13% and 14% of the open-ended groups.

**Implications**

Varying the assessment method can cause large differences in reported rates of adverse events.

—The Editors

told that they would be taking placebo at some point during the study; and were assigned to a single-blind, 1-month placebo run-in period.

**Randomization and Intervention**

After taking the placebo (referred to as the “study medication”) for 1 month, patients were randomly assigned to 3 methods of collecting adverse events. All patients were given 1 of 3 self-administered paper forms. The form given to the first group asked an open-ended question: “Did you have any significant medical problem since the last study visit?” The form given to the second group asked an open-ended question that was more defined: “Since the last study visit, have you limited your usual daily activities for more than 1 day because of a medical problem?” A checklist accompanied the form given to the third group, which asked a more pointed question: “Since the last visit, have you experienced any of the following (checklist attached)?” The checklist contained 53 symptoms, grouped by anatomical region. Two of the authors developed the checklist after conducting an unpublished review of checklists that were used in earlier clinical trials performed at the same institution. The checklist did not ask patients to rate the frequency or severity of symptoms and did not ask patients to make a judgment about whether their medical problem was caused by the study medication. Patients in the open-ended question groups who answered “yes” were asked to identify their medical problem, which was recorded by a study assistant on the same checklist used in the third group.

**Outcomes and Analysis**

The primary outcome measure was the difference in the proportion of patients reporting 1 or more adverse events in each group. All patients in the STEP study were

included in the current study; therefore, the sample size was not calculated on the basis of the needs of this study. Participants were randomly assigned to the 3 groups in equal proportions by using a computer-generated, random allocation sequence that was prepared before the study began. Study personnel were blinded to the allocation sequence but were aware of group assignments after they were made. Patients were not informed of their group assignment. Persons performing the data analysis were blinded to group assignment. Baseline characteristics of the 3 intervention groups were compared by using analysis of variance for continuous variables and chi-square tests for categorical variables. We also used chi-square tests to compare the number and specific type of adverse events that occurred among groups. All analyses were performed by using Stata, version 8.0 (Stata Corp., College Station, Texas).

**Role of the Funding Sources**

The funding organizations had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

**RESULTS**

We randomly assigned 214 patients to 1 of 3 methods of collecting data on adverse effects. Patients were predominantly healthy, well-educated white men (mean age, 63 years) who were taking a mean of 2.5 medications (Table 1). Baseline characteristics of the patients were similar among the 3 groups. All patients completed the study and the outcome assessment (Figure).

The group that was assigned to a checklist method reported a significantly greater number of adverse events (238 events) than the first and second groups, which were asked open-ended questions (11 and 14 adverse events, respectively;  $P < 0.001$ ) (Table 2). A much higher percentage of patients in the checklist group reported 1 or more adverse events (77%) compared with the patients asked each of the 2 different open-ended questions (14% for the first group and 13% for the second group;  $P < 0.001$ ). For each of the 10 most commonly reported adverse events (Table 2), participants in the checklist group reported a greater number of adverse events ( $P < 0.001$ ). No serious adverse events occurred during the study period.

**DISCUSSION**

In this randomized, controlled trial, we found that a checklist method of identifying adverse events dramatically increased the number of reported events compared with 2 types of open-ended questions. Although this finding is intuitive, the magnitude of effect has important implications both for the conduct of clinical trials and for assess-

**Table 1. Baseline Characteristics of Study Participants**

Characteristic	All Groups (n = 214)	Open-Ended Question Group (n = 70)	Open-Ended, Defined Question Group (n = 70)	Checklist Group (n = 74)
Mean age (SD), y	63.1 (7.9)	63.3 (8.5)	62.5 (7.5)	63.3 (7.7)
Ethnicity, n (%)				
White	173 (81)	53 (77)	59 (84)	61 (83)
Black	12 (6)	5 (7)	3 (4)	4 (5)
Asian	14 (6)	6 (9)	4 (6)	4 (5)
Hispanic	11 (5)	3 (4)	3 (4)	5 (7)
Other	4 (2)	2 (3)	1 (2)	0 (0)
Completed college, n (%)	145 (68)	43 (61)	52 (74)	50 (68)
Mean score on the American Urological Association Symptom Index (SD)*	15.2 (5.5)	15.0 (6.1)	15.3 (5.2)	15.1 (5.0)
Comorbid conditions, n (%)				
Hypertension	89 (42)	31 (44)	27 (39)	31 (42)
Lung disease	40 (19)	16 (23)	9 (13)	15 (20)
Heart disease	37 (17)	13 (19)	12 (17)	12 (16)
Diabetes	16 (7)	4 (6)	6 (9)	6 (8)
Medications†				
Mean medications used (SD), n	2.5 (2.4)	2.5 (2.6)	2.2 (2.4)	2.5 (2.5)
Aspirin, n (%)	65 (30)	22 (31)	22 (31)	21 (28)
Statin, n (%)	49 (23)	22 (31)	12 (17)	15 (20)
Angiotensin-converting enzyme inhibitor, n (%)	32 (15)	12 (17)	12 (17)	8 (11)
β-Blocker, n (%)	28 (13)	9 (13)	12 (17)	7 (9)
Thiazide diuretic, n (%)	25 (12)	8 (11)	9 (13)	8 (11)
Calcium-channel blocker, n (%)	23 (11)	11 (16)	6 (9)	6 (8)
Nonsteroidal anti-inflammatory drug, n (%)	21 (10)	5 (7)	7 (10)	9 (12)
Antidepressant, n (%)	18 (8)	5 (7)	5 (7)	8 (11)
Proton-pump inhibitor, n (%)	16 (7)	4 (6)	6 (9)	6 (8)
Hypoglycemic, n (%)	11 (5)	3 (4)	4 (6)	4 (5)

\* A patient's score reflects the severity of symptoms; a score of 0 to 7 indicates mild symptoms, a score of 8 to 19 indicates moderately severe symptoms, and a score of 20 to 35 indicates severe symptoms.

† The 10 most commonly used medicines taken by participants.

ment of the risk–benefit profile of drugs and other interventions.

It is common practice for physicians and patients to select drugs and other interventions on the basis of their reported rate of side effects. However, if different drugs used for the same indication are examined in clinical trials that use different methods of identifying adverse events, then it is not valid to compare the reported rate of side effects. For example, the reported rates of sexual side effects from selective serotonin reuptake inhibitors range from 2% to 73%; much of this difference is probably attributable to different methods of adverse event collection (4). Similarly, a recent systematic review found that published trials of pharmacologic treatments for rheumatoid arthritis were much more likely to report data on harm than trials of nonpharmacologic treatment (5), highlighting the difficulty of comparing the safety of different treatments for the same condition.

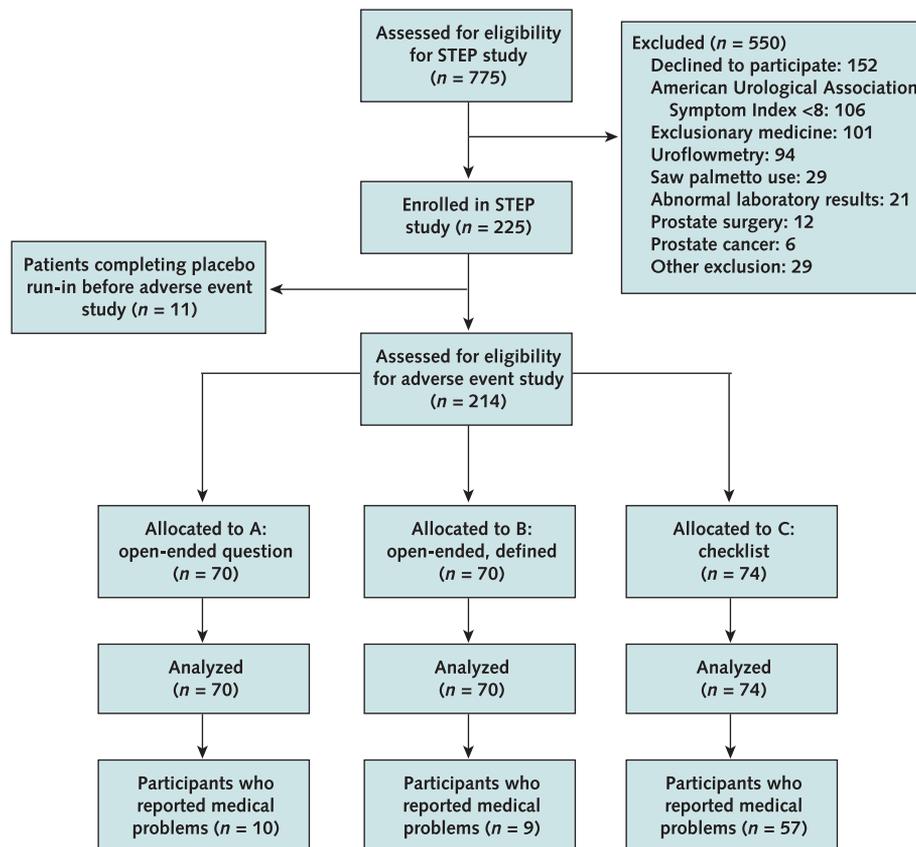
The 3 self-administered questions that we used to assess the frequency of adverse events in this study were, by design, limited in scope. The self-administered forms did not ask patients to describe the timing, severity, or frequency of their medical problems, nor did they ask participants or investigators to make a judgment of causality. Other techniques to assess adverse events, such as changes in vital signs, laboratory tests, physical examinations, or

more detailed searches for expected adverse events, were not included. The purpose of this simplified approach was to isolate and contrast 3 different initial screening methods of identifying medical problems occurring among participants in a clinical trial.

Because all patients in the current study were taking placebo, probably none of the reported adverse events were true side effects of the “study medication” but were merely symptoms that commonly occur in adults. For example, a previous survey of university students and hospital staff found that 81% of respondents reported some symptom within the past 3 days when prompted by a checklist (6). This highlights the problem that most study participants are likely to have a high prevalence of symptoms that are unrelated to a study drug or intervention, and a checklist method is therefore likely to have very low specificity for detecting true side effects.

The wording of the 3 self-administered questions that we used in this study asked about 3 different thresholds of medical problems. One question asked participants if they experienced a “significant medical problem,” one asked if they “limited their usual daily activities for more than 1 day because of a medical problem,” and one merely asked if participants had anything shown on a checklist. It is possible that the observed differences in reported medical problems with the open-ended questions versus the check-

Figure. Flow diagram showing the distribution of participants at each stage of the study.



STEP = Saw Palmetto Treatment for Enlarged Prostates.

list might have been at least partially attributable to the different wording of these questions. Also, our study included only relatively healthy, well-educated men; therefore, the findings may not apply to other populations.

Several previous studies have compared the use of open-ended questions and checklists to determine the frequency of reported adverse events in a clinical trial (7–11).

All of these studies assigned patients to receive both methods of adverse event collection, with the open-ended question being administered first; no randomized comparisons were performed. This method has the potential to bias the participants’ response to the checklist because they are initially “primed” by the open-ended question. These studies had several other methodologic shortcomings, including a

Table 2. 10 Most Frequently Reported Adverse Events\*

Adverse Event	Open-Ended Question Group (n = 70), n	Open-Ended, Defined Question Group (n = 70), n	Checklist Group (n = 74), n
Total events reported	11	14	238
Back pain	0	5	13
Upper respiratory infection, flu-like symptoms	1	2	15
Dry mouth	0	0	13
Fatigue	0	0	12
Cough	0	0	11
Nocturia	0	0	11
Joint pain or swelling	0	1	9
Impotency	0	0	10
Headache	0	0	10
Heartburn or indigestion	1	0	9

\* All P values < 0.001.

high dropout rate (8), inclusion of only neurotic patients (9, 11), and inadequate statistical analysis (10). Despite such flaws, these studies still consistently found that the checklist method identified more adverse events than open-ended questioning. Our study—which randomly assigned patients to different methods of collecting information on adverse events; included patients who were all participating in a single-blind, placebo run-in period; and had complete data collection on all patients—validates the findings of earlier work.

One could make rational arguments for using a checklist, an open-ended question, or other techniques to identify adverse events in clinical trials. However, use of different methods for collecting data in different clinical trials limits the ability to compare the side effect profile of drugs. The use of the CONSORT (Consolidated Standards of Reporting Trials) statement has helped to make clinical trials more uniform and comparable (12, 13), and we strongly advocate that this effort be extended to include a uniform method for identifying adverse events in clinical trials.

From University of California and San Francisco Veterans Administration Medical Center, San Francisco, California.

**Acknowledgments:** The authors thank Drs. Suzanne Staccone and Evelyn Badua, Ms. Bertina Lee, and Ms. Arleen Sakamoto for their work and for establishing the patient rapport that made this study possible.

**Grant Support:** By the National Center for Complementary and Alternative Medicine (grant 1 K08 ATO1338-01) and by the National Institute of Diabetes and Digestive and Kidney Diseases (grant 1 RO1 DK56199-01).

**Potential Financial Conflicts of Interest:** None disclosed.

**Requests for Single Reprints:** Stephen Bent, MD, San Francisco Veterans Administration Medical Center, 111-A1, 4150 Clement Street,

San Francisco, CA 94121; e-mail, bent@itsa.ucsf.edu.

Current author addresses and author contributions are available at [www.annals.org](http://www.annals.org).

## References

1. U.S. Food and Drug Administration. Clinical Safety Data Management: Definitions and Standards for Expedited Reporting 1995. Accessed at [www.fda.gov/cder/guidance/iche2a.pdf](http://www.fda.gov/cder/guidance/iche2a.pdf) on 29 September 2005.
2. Gotzsche PC. Non-steroidal anti-inflammatory drugs. *BMJ*. 2000;320:1058-61. [PMID: 10764369]
3. National Center for Complementary and Alternative Medicine. Clinical Trial: Saw Palmetto Extract in Benign Prostatic Hyperplasia. National Institutes of Health; 2005. Accessed at <http://clinicaltrials.gov/show/NCT00037154> on 29 September 2005.
4. Safer DJ. Design and reporting modifications in industry-sponsored comparative psychopharmacology trials. *J Nerv Ment Dis*. 2002;190:583-92. [PMID: 12357091]
5. Ethgen M, Boutron I, Baron G, Giraudeau B, Sibilia J, Ravaud P. Reporting of harm in randomized, controlled trials of nonpharmacologic treatment for rheumatic disease. *Ann Intern Med*. 2005;143:20-5. [PMID: 15998751]
6. Reidenberg MM, Lowenthal DT. Adverse nondrug reactions. *N Engl J Med*. 1968;279:678-9. [PMID: 5670907]
7. Wallin J, Sjoval J. Detection of adverse drug reactions in a clinical trial using two types of questioning. *Clin Ther*. 1981;3:450-2. [PMID: 7020939]
8. Wallander MA, Dimenas E, Svardsudd K, Wiklund I. Evaluation of three methods of symptom reporting in a clinical trial of felodipine. *Eur J Clin Pharmacol*. 1991;41:187-96. [PMID: 1748134]
9. Lapierre YD. Évaluation des effets secondaires chez les névrotiques—un essai avec la mesoridazine et le placebo. *Can Psychiatr Assoc J*. 1975;20:61-6. [PMID: 1091347]
10. Greenblatt M. Controls in clinical research. *Clin Pharmacol Ther*. 1964;5:864.
11. Downing RW, Rickels K, Meyers F. Side reactions in neurotics. I. A comparison of two methods of assessment. *J Clin Pharmacol J New Drugs*. 1970;10:289-97. [PMID: 5272671]
12. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134:663-94. [PMID: 11304107]
13. Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004;141:781-8. [PMID: 15545678]

**Current Author Addresses:** Drs. Bent and Avins and Ms. Padula: San Francisco Veterans Administration Medical Center, 111-A1, 4150 Clement Street, San Francisco, CA 94121.

**Author Contributions:** Conception and design: S. Bent, A.L. Avins. Analysis and interpretation of the data: S. Bent, A. Padula, A.L. Avins. Drafting of the article: S. Bent.

Critical revision of the article for important intellectual content: S. Bent, A. Padula, A.L. Avins.

Final approval of the article: S. Bent, A.L. Avins.

Statistical expertise: S. Bent.

Administrative, technical, or logistic support: A. Padula.

Collection and assembly of data: A. Padula.