The validity of a patient-reported adverse drug event questionnaire using different recall periods

Sieta T. de Vries · Flora M. Haaijer-Ruskamp · Dick de Zeeuw · Petra Denig

Abstract

Purpose To assess the validity of a patient-reported adverse drug events (ADEs) questionnaire with a 3-month or 4-week recall period.

Methods Patients receiving at least one oral glucose-lowering drug were asked to report potential ADEs they experienced related to any drug in a daily diary for a 3-month period. Thereafter, they completed the ADE questionnaire with either a 3-month or 4-week recall period. The validity was assessed by comparing ADEs reported in each version with those reported in the diary at class level and at specific ADE level. At class level, a comparison was made using (1) primary system organ classes (SOCs) of the medical dictionary for regulatory activities and (2) other related SOCs. Sensitivity and positive predictive value (PPV) were calculated.

Results Each version of the questionnaire was completed by 39 patients. In the 3-month group, 21 patients reported 70 ADEs in the diary. In the 4-week group, six patients reported seven ADEs in the last 4 weeks of the diary. Sensitivity to assess ADEs at primary SOC was low for both recall groups (33 %). PPV was 51 and 10 % for, respectively, the 3-month and 4-week group. Taking other related SOCs into account slightly increased the sensitivity for the 3-month group (38%). Sensitivity of reporting the same ADE was 41 and 43 % for, respectively, the 3-month and 4-week group.

Conclusions Regardless of the recall period and level of comparison, the validity for assessing ADEs was low with the patient-reported ADE questionnaire. Further refinement is needed to improve the validity.

Keywords Validity · Recall period · Patient-reported outcome · Adverse drug events · Longitudinal studies · Diary · Sensitivity · Positive predictive value

S. T. de Vries · F. M. Haaijer-Ruskamp · D. de Zeeuw · P. Denig (✉)
Department of Clinical Pharmacy and Pharmacology (FB20), University of Groningen, University Medical Center Groningen, PO Box 196, 9700 AD Groningen, The Netherlands
e-mail: p.denig01@umcg.nl

Introduction

In clinical trials, a daily diary is often used to record adverse events experienced by patients [1, 2]. The diary method brings the reporting closer to the occurrence of adverse events than retrospective questionnaires [3] and can be seen as a gold standard in the assessment of symptoms due to the quality and richness of the collected information [4, 5]. Daily diaries have also been used in observational studies assessing adverse events [6]. However, keeping a daily diary is burdensome. The use of a retrospective questionnaire with a longer recall period might be an alternative. The recall period in a questionnaire is the time period for which the patient has to consider the answer of the question [7].

A patient's recall is influenced by factors such as forgetting an event or its correct date, its accessibility in mind (influenced by for instance its recency, frequency, and salience), and one's mood [5, 8–10]. In addition, a patient's evaluation of a specific event or health state may change over time due to a response shift, which is defined as the use of a different reference category [11, 12]. Therefore,

the recall period in a questionnaire is seen as a limitation of patient reporting [13]. An inappropriate recall period may introduce measurement error [14]. The food and drug administration recommends to pay attention to this issue when constructing patient-reported outcome (PRO) instruments [15].

An optimal recall period depends on various issues, and debate is ongoing about what recall period is suitable for which questionnaire [14, 16]. Some studies have assessed the impact of different recall periods on reporting symptoms in questionnaires (e.g., [17–19]). The general conclusion is that there is an inverse association between the length of recall period and accuracy of recall used in questionnaires [14].

Currently, information about the optimal recall period to assess adverse drug events (ADEs) is lacking. Commonly used recall periods in PRO instruments are between 1 day up to 4 weeks [14]. However, from our study about the content validation of a patient-reported ADE questionnaire, it became clear that several patients found a recall period of 4 weeks relatively short [20]. A longer recall period can be preferred since it is not always immediately clear for patients whether or not a symptom is an ADE, and ADEs that occur irregularly or after some time may not be captured in a 4-week period.

In the current study, we examined the validity of a retrospective questionnaire using a 4-week or a 3-month recall period for the assessment of ADEs. The primary objective was to assess the validity of reported ADEs at aggregated class level and at individual ADE level. The secondary objective was to explore whether the validity of the questionnaire might be dependent on either the class of ADE, or characteristics at patient level.

Method

The study had a longitudinal design, where patients first completed a daily diary for a period of 3 months followed by a previously developed retrospective questionnaire [20]. The reporting of ADEs in the questionnaire was compared with a daily diary, which was used as the gold standard. Although the patient-reported questionnaire can be used to assess adverse drug reactions (ADRs), we use the term ADE instead of ADR for two reasons. First, patients may be uncertain about a causal relation between a symptom and a drug [20]. Second, patients may perceive unintended responses due to medication errors or overdoses. This implies that our questionnaire is not restricted to assess ADRs as defined by the World Health Organization as "a noxious and unintended response to a medicine that occurs at normal therapeutic doses used in humans for prophylaxis, diagnosis, or therapy of disease, or for the modification of physiologic function"

[21]. The study was carried out in accordance with the Code of Ethics of the World Medication Association (Declaration of Helsinki) for experiments involving humans. The Medical Ethics Committee of the University Medical Center Groningen in The Netherlands determined that ethical approval was not needed for this study.

Participants

Inclusion criteria for the patients were age 18 years or older, being dispensed an oral glucose-lowering drug, availability of an e-mail address, access to the internet, and ability to read and write Dutch language. These patients were recruited in 2012 and 2013 via pharmacies in the northern part of The Netherlands. In around 30 pharmacies, a randomly selected sample of 15 patients aged 18 years or older, being dispensed an oral glucose-lowering drug, were contacted by telephone and sent an information letter when they fulfilled the inclusion criteria and were interested in participation. In another 4 pharmacies, an information letter was sent to all patients aged ≥18 years and being dispensed an oral glucose-lowering drug (three pharmacies) or glucose-lowering drug (one pharmacy). The patients who returned a completed consent form and who fulfilled the inclusion criteria were included in the study. After completing the study, patients were compensated with a voucher of €10 for participation.

Reported ADEs

As primary outcome, we compared reported ADEs at the primary System Organ Class level of the Medical Dictionary for Regulatory Activities (MedDRA®) terminology version 13.0. Each symptom included in the patient-reported questionnaire was assigned to a Lowest Level Term of the MedDRA® [20], which is linked to at least one System Organ Class. This is the level at which all adverse reactions should be tabulated according to the European guideline of summary of product characteristics [22]. Although each symptom has at least a primary System Organ Class, the symptom may also be linked to a secondary and even tertiary System Organ Class. Therefore, we additionally assessed the validity by including secondary and tertiary System Organ Class levels of the MedDRA®, if applicable, which takes into account possible misclassifications. As a third step, the validity was assessed for reporting the same ADE at the lowest level.

Material and procedure

A paper-based diary was sent by mail to the participants, to be filled in daily for a period of 3 months. The diary was developed for this study and consisted of an open-ended

question asking for symptoms experienced. An additional closed-ended question asked whether or not the patient attributed the symptom(s) to any drug they used, not restricted to their oral glucose-lowering drug. Telephone reminders were given to patients who did not return their diary within a month after it should have been completed.

After returning the completed diary, the patient received an e-mail message with the URL (uniform resource locator) to open the web-based version of the patient-reported ADE questionnaire, which was constructed using the Unipark Enterprise Feedback Suite 8.0 version 1.1 (http://www.unipark.de). The e-mail message included a personal login code to prevent multiple completions of the questionnaire by a patient [23]. The patient-reported ADE questionnaire is a generic questionnaire which includes general questions about patient characteristics and drug use, and a list with symptoms in lay terms which can be checked by the patient as a symptom unrelated to any drug or as a potential ADE [20]. Additional questions about the nature of the ADE and the drugs a patient relates to the ADE are asked for each potential ADE. Two versions of the patient-reported ADE questionnaire were used, one with a recall period of 3 months (e.g., "Which symptoms involving your 'eyes and/or eyelids' did you experience during the past 3 months") and one with a recall period of 4 weeks (e.g., "Which symptoms involving your 'eyes and/or eyelids' did you experience during the past 4 weeks"). Patients were randomized using blocked randomization [24] to one of the two groups that differed in the recall period of the questionnaire. We aimed to include 100 patients (50 per group), which has been suggested as a reasonable number for reliability studies [25]. Although the current study does not assess the reliability, we used this number as a reference since no data about different recall periods in assessing ADEs were available for calculating the required sample size.

Analyses

Differences in patient characteristics between those who completed the questionnaire with a recall period of 4 weeks and those who completed the questionnaire with a recall period of 3 months were compared using the Pearson χ^2-test, Fisher–Freeman–Halton test, and t-test, depending on the type of variable.

For the comparison of the questionnaire with the diary, the ADEs were used as unit of analysis and the number of true-positive, true-negative, false-positive, and false-negative ADEs are presented where relevant. The validity of reporting ADEs at primary System Organ Class level of the MedDRA® was assessed by calculating the sensitivity [26] and positive predictive value [27]. Specificity and negative predictive value were not calculated since these values are

expected to be high and non-informative due to the high number of true negatives. A positive outcome was defined as the detection of an ADE at this *primary class level*. Exact confidence intervals (CI) based on binomial probabilities were calculated for these validity measures [28]. The questionnaire with a recall period of 3 months was compared with the full 3-month diary, whereas the questionnaire with a recall period of 4 weeks was compared with the last 4 weeks reported in the diary. In addition, the questionnaire with a recall period of 4 weeks was compared with the full 3-month period in the diary to assess the validity in ADE reporting within this wide time frame to allow for incorrect recall of the date of occurrence. Sensitivity analyses were performed to assess whether delayed completion of the diary or the questionnaire affected the results by excluding (1) those patients with >14 days between the last date reported in the diary and receiving the completed diary by the researchers (delayed diary completers), (2) those patients who completed the questionnaire >14 days after the diary was received by the researchers (delayed questionnaire completers), and (3) both the delayed diary and delayed questionnaire completers. Differences between the two recall groups in days of delay were compared using Mann–Whitney U tests.

The sensitivity of both versions of the questionnaire was additionally calculated at (1) MedDRA® *additional class level* for taking not only the primary but also the secondary or tertiary System Organ Classes of the MedDRA® into account if applicable and (2) *specific ADE level* for reporting the same ADEs among the questionnaire and the diary. Two researchers (PD and STdV) independently classified the reported ADEs in the diary to a System Organ Class of the MedDRA® and checked whether or not the specific ADEs reported in the diary were the same ADEs as reported in the retrospective questionnaire. Discrepancies in the judgments of the researchers were resolved by discussion. All participants were included in the analyses comparing the questionnaire with the diary.

To explore whether the validity of the questionnaire was dependent on the class of ADE, or on characteristics at patient level, the reports of both recall groups were combined. The sensitivity per primary System Organ Class of the MedDRA® was assessed for those classes in which at least five ADEs were reported in the diary and/or the questionnaire. The age, gender, and education level of the patients were compared between those patients with no agreement (no corresponding ADEs), partial agreement (some but not all corresponding ADEs), and full agreement (all corresponding ADEs) between the ADEs reported in the diary and the questionnaire.

The analyses were conducted using IBM SPSS Statistics version 20 (Armonk, New York, USA), and P values <0.05 were considered statistically significant. Confidence

Table 1 Patient characteristics per recall group

	4 weeks (N = 39)	3 months (N = 39)	P value
Mean age (SD)	63 (10.0)	67 (7.2)	0.069[•]
Females (%)	20 (51.3)	11 (28.2)	0.037*
Education (%)			0.594[‡]
Lower education[a]	8 (20.5)	8 (20.5)	
Middle education[b]	17 (43.6)	21 (53.8)	
Higher education[c]	12 (30.8)	7 (17.9)	
Other	2 (5.1)	3 (7.7)	

[•] T test; * Pearson χ^2-test; [‡] Fisher–Freeman–Halton test

[a] No education; elementary school; junior secondary vocational education

[b] Junior general secondary education; senior secondary vocational education

[c] Senior general secondary education; higher professional education; university education

intervals were calculated using Stata version 12 (Stata Corp., College Station, TX).

Results

Of the 113 patients who returned an informed consent form, 78 patients (69 %) completed the study. These patients did not significantly differ in age, sex, and education level from the patients who did not complete the study (data not shown). No differences between the completers of the 2 recall groups were found in age and education level, but more males were included in the 3-month recall group than in the 4-week recall group (P < 0.05; Table 1). In total, 27 of the 78 participants reported 77 individual ADEs in the diary. Of these ADEs, 61 were linked to a System Organ Class of the MedDRA® (multiple ADEs reported by one participant within the same System Organ Class were counted as one).

Validity of reporting ADEs at primary class level

The sensitivity and positive predictive value were low for both recall periods (Table 2). Sensitivity analyses by excluding delayed diary and/or questionnaire completers revealed similar validity levels (Online Resource 1). The comparison of the 4-week recall questionnaire with the full 3-month diary revealed a similar sensitivity (32 %; 95 % CI 14–55 %) and a slightly increased positive predictive value (from 10 %; 95 % CI 1–30 %, to 33 %; 95 % CI 15–57 %).

Validity of reporting ADEs at additional class level

The sensitivity of the 4-week recall group remained the same when taking also secondary and tertiary System

Table 2 Validity of the retrospective questionnaire with a recall period of 4 weeks or 3 months compared with the daily diary in reporting adverse drug events at MedDRA® primary class level (N = 702)

	TP	FP	TN	FN	Se (95 % CI)	PPV (95 % CI)
4-week recall; last 4 weeks of diary	2	19	677	4	33 % (4–78)	10 % (1–30)
3-month recall; full 3-month diary	18	17	630	37	33 % (21–47)	51 % (34–69)

The N at MedDRA® level is the number of patients per recall group (39) times 18 different MedDRA® System Organ Classes which are covered by the ADEs in the questionnaire

TP true positive, FP false positive, TN true negative, FN false negative; Se sensitivity; PPV positive predictive value, CI confidence interval

Organ Classes of the MedDRA® into account. For the 3-month recall group, a slightly increased sensitivity was shown (from 33 %; 95 % CI 21–47 %, to 38 %; 95 % CI 25–52 %).

Specific ADE level

In the 3-month recall group, 21 patients (54 %) reported in total 70 ADEs in the diary. The sensitivity of the questionnaire in reporting the same ADE was 41 % (95 % CI 30–54 %; number of true positives 29; number of false negatives 41). In the 4-week recall group, 6 patients (15 %) reported in total 7 ADEs in the last 4 weeks of the diary, and the sensitivity of the questionnaire was 43 % (95 % CI 10–82 %; number of true positives 3; number of false negatives 4).

Differences per class of ADE and in characteristics at patient level

Sensitivity levels ranged from 0 to 50 % per System Organ Class of the MedDRA®, but confidence intervals were overlapping (Table 3). Of the 27 patients who reported one or more ADEs in the diary, 6 (22 %) patients had full agreement by reporting all of these ADE also in the questionnaire, 11 (41 %) had partial agreement, and 10 (37 %) had no agreement. Patients with no agreement were somewhat younger than patients with full or partial agreement [mean age in years 64 (sd: 6) vs. 66 (sd: 10) and 67 (sd: 8)], and more often female (60 vs. 33 and 36 %). The education level of the patients appeared to be similar among the three groups.

Discussion

Regardless of the recall period, the patient-reported ADE questionnaire had a low sensitivity to identify patients who

Table 3 Validity of the questionnaire* per MedDRA® system organ class level for those classes in which ≥5 adverse drug events are reported in the diary or the questionnaire

MedDRA® System Organ Class level	TP	FP	TN	FN	Se (95 % CI**)
Gastrointestinal disorders	7	2	62	7	50 % (23–77)
General disorders and administration site conditions	2	3	69	4	33 % (4–78)
Metabolism and nutrition disorders	0	4	72	2	0 % (0–84)
Musculoskeletal and connective tissue disorders	3	3	66	6	33 % (7–70)
Nervous system disorders	4	6	62	6	40 % (12–74)
Psychiatric disorders	0	2	73	3	0 % (0–71)
Respiratory, thoracic, and mediastinal disorders	0	5	71	2	0 % (0–84)
Skin and subcutaneous tissue disorders	3	3	66	6	33 % (7–70)

TP true positive, *FP* false positive, *TN* true negative; *FN* false negative; *Se* sensitivity, *PPV* positive predictive value, *CI* confidence interval

* Results of the questionnaire with a 4-week and 3-month recall period are combined

** One-sided, 97.5 % confidence interval when TP equals 0

experienced an ADE at organ class level and at specific ADE level. In addition, the questionnaire had low positive predictive value. There may be differences among classes of ADEs but additional studies are needed to confirm this finding. In addition, further studies are needed to assess whether characteristics at patient level, such as age and gender, influence the validity of patient-reported ADE questionnaires.

The positive predictive value of the questionnaire was especially low for the 4-week recall period. Patients in this recall group more often reported an ADE at MedDRA® level in the questionnaire than in the diary. This higher reporting could be due to reporting additional ADEs in the questionnaire, or to forward telescoping, that is, ADEs were reported as being more recent than they actually occurred [5, 29, 30]. Forward telescoping probably occurred at least five times in the 4-week recall group, since five additional true positives were found when the full 3-month diary was taken into account. However, it should be noted that patients may have been primed to the 3-month period when answering the questionnaire because of our study design, since they completed the diary for this time period.

There are several factors that influence the validity of the questionnaire when comparing it with a diary. First of all, patients complete a diary with the knowledge they have at that moment, whereas the questionnaire is completed with the knowledge they gained over time about their symptoms. This additional knowledge may change their opinion about, for instance, a symptom being an ADE. Second, an open-ended question was used in the diary, whereas a symptom checklist was used in the questionnaire. Previously, it was shown that more patients report an ADE and that the number of reported ADEs is higher on a checklist than on an open-ended question [31]. Using the open-ended question in the diary as the gold standard, the use of a closed instead of open-ended question in the questionnaire may lead to higher false-positive rates. Furthermore, some patients appeared to have delayed the completion of the diary or the questionnaire, which can be expected to result in lower validity. The sensitivity analyses, however, showed also low validity levels when such patients were excluded.

Previously, we found low test–retest reliability of assessing ADEs at specific level using the patient-reported ADE questionnaire, which may be due to problems in the questionnaire as well as a patient's uncertainty about a symptom being an ADE [20]. This uncertainty was also demonstrated in the current study, in which patients related a specific drug to the ADE in less than half of the cases (data not shown). In addition, reported symptoms were sometimes indicated as an ADE and sometimes as 'I do not know' in the diaries. The uncertainty may particularly occur in patients with multiple comorbidity and comedication, which is common in the patient population included in this study [32]. Therefore, the low validity observed in our study may in part be due to the complexity of acknowledging ADEs in this specific patient population. The performance of the questionnaire might be better in patients who only have one disease or use one drug. Qualitative studies are needed to assess to what extent patients in general, and patients with type 2 diabetes more specifically, are able to report all (possible) ADEs, and to gain more knowledge about discrepancies in reported ADEs between a diary and a questionnaire.

We observed slightly lower sensitivities at organ class levels as compared to specific ADE level. This suggests that the direct linkage of symptoms in the checklist to MedDRA® terms may be inadequate. We observed that the System Organ Class of the checked symptom may differ from the System Organ Class that would be linked to the additional information given by the patients about the symptom (e.g., the System Organ Class of a checked symptom "tingling or prickling sensation" differs from the System Organ Class of the additional description provided by the patient being "muscle pain").

Only small differences were found in the validity of reporting specific ADEs between a questionnaire with a recall period of 4 weeks and 3 months. This finding of similar validity among the recall periods differs from a recent study in which higher accuracy of reporting

headache frequency was found when a recall period of 30 days was used compared to 90 days [17]. This inconsistency indicates that conclusions about recall periods cannot easily be transferred from one questionnaire to another, as has been stated before [14, 16].

Strengths and limitations

To our knowledge, this is the first study assessing the validity of different recall periods for assessing ADEs in a patient-reported questionnaire. Some limitations need to be acknowledged. The first and major limitation is the small sample size included in this study in combination with the low number of patients reporting an ADE, especially in the 4-week recall group. This limitation resulted in wide confidence intervals. A post hoc analyses showed that given the current data, a sample size of 43 for the 3-month recall group and 378 for the 4-week recall group would be necessary to achieve an accuracy of 5 % for the observed sensitivity of 33 % at primary System Organ Class level of the MedDRA® [33]. This finding indicates that for a more precise indication of the validity of a recall period of 4 weeks in an ADE questionnaire, a (preselected) sample of patients with a higher expected ADE rate would be preferable. Secondly, we included a selective sample of patients that responded to the letters sent via pharmacists. These were patients consenting to keep a diary for 3 months, using an oral glucose-lowering drug, and with internet access. A previous study with a web-based version of the patient-reported ADE questionnaire showed that the responders were younger than the non-responders [20]. Thirdly, more males were included in the 3-month recall group than in the 4-week recall group, indicating that the randomization was not completely successful. Fourthly, there are some limitations with the use of a daily diary as a gold standard in the reporting of ADEs. It has been noted that daily diaries also require recall and may be influenced by the same factors that apply to retrospective questionnaires [3]. In addition, we are not sure whether patients completed the diary each day with the risk of loss of validity [17]. On the other hand, keeping a daily diary before completing a questionnaire may positively affect the recall in the questionnaire. We expect these factors to be similar for both recall groups. Furthermore, patients may become tired of keeping a diary, which can lead to less validity in the last period and therefore lower validity in the 4-week recall group. Although the number of patients reporting an ADE was relatively stable over time (data not shown), we cannot exclude this possibility.

Practice implications

The patient-reported ADE questionnaire is a generic questionnaire which is intended to measure all ADEs experienced by patients. However, the questionnaire is not sufficiently sensitive to detect all experienced ADEs. In addition, the questionnaire has low positive predictive value. Therefore, adaptations to the patient-reported ADE questionnaire are needed before it can be generally used. The direct linkage of checked symptoms to MedDRA® terms may introduce misclassifications. In addition, patients check multiple symptoms describing one ADE, as has been shown previously [20]. Therefore, starting with an open-ended question in which patients give a description of their ADEs which is then linked to a MedDRA® term may be preferred. In additional research, the validity of such an adaptation should be tested. Further research is also needed to gain more insight into whether there are differences in accuracy among classes of ADEs. For observational studies assessing ADEs in patients using chronic medication, a recall period of 3 months may be preferable compared to a 4-week recall period. A 3-month recall period has the advantage of covering a longer time period facilitating the identification of more ADEs. Our study suggests that the validity of reporting specific ADEs is hardly affected using a longer recall period, but further validation may be needed when the questionnaire is adapted. Shorter recall periods, however, may be needed for clinical trials and studies that try to assess ADEs experienced at different stages of treatment [14, 17].

Conclusion

This study showed that a retrospective patient-reported ADE questionnaire is insufficiently valid for assessing ADEs, regardless of the recall period and the level of comparison. The use of a 3-month recall period may be preferred over a 4-week recall period since it covers a longer time period. However, further refinement of the questionnaire is needed to improve its validity.

References

1. Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials, 24*(2), 182–199.
2. Shiffman, S. (2004). Electronic diaries: Impact on drug development. http://www.google.nl/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=2&ved=0CDoQFjAB&url=http%3A%2F%2Fwww.fda.gov%2Fohrms%2Fdockets%2Fac%2F04%2Fslides%

2F2004-4078OPH1_01_Shiffman.ppt&ei=doRfUqz2CsbZswam54HoDg&usg=AFQjCNG2FAPyq7Q1qPEzbt-4gyBCsO2CEA. Accessed March 28, 2014.

3. Gendreau, M., Hufford, M. R., & Stone, A. A. (2003). Measuring clinical pain in chronic widespread pain: Selected methodological issues. *Best Practice & Research. Clinical Rheumatology, 17*(4), 575–592.

4. Monk, T. H., Buysse, D. J., Kennedy, K. S., Pods, J. M., De-Grazia, J. M., & Miewald, J. M. (2003). Measuring sleep habits without using a diary: The sleep timing questionnaire. *Sleep, 26*(2), 208–212.

5. Verbrugge, L. M. (1980). Health diaries. *Medical Care, 18*(1), 73–95.

6. Klocke, K. R., Stauch, K., & Landen, H. (2003). Effect of add-on acarbose to insulin therapy in routine clinical practice. *Clinical Drug Investigation, 23*(10), 621–627.

7. DeWalt, D. A., Rothrock, N., Yount, S., Stone, A. A., & PROMIS Cooperative Group. (2002). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care, 45*(5 Suppl 1), S12–S21.

8. Hufford, M. R., & Shiffman, S. (2002). Methodological issues affecting the value of patient-reported outcomes data. *Expert Review of Pharmacoeconomics & Outcomes Research, 2*(2), 119–128.

9. Crosby, R. D., Kolotkin, R. L., & Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology, 56*(5), 395–407.

10. Bohner, G., & Wänke, M. (2002). *Attitudes and attitudes change*. East Sussex: Psychology Press Ltd.

11. Breetvelt, I. S., & Van Dam, F. S. (1991). Underreporting by cancer patients: The case of response-shift. *Social Science and Medicine, 32*(9), 981–987.

12. Litwin, M. S., & McGuigan, K. A. (1999). Accuracy of recall in health-related quality-of-life assessment among men treated for prostate cancer. *Journal of Clinical Oncology, 17*(9), 2882–2888.

13. King, A., Daniels, J., Lim, J., Cochrane, D. D., Taylor, A., & Ansermino, J. M. (2010). Time to listen: A review of methods to solicit patient reports of adverse events. *Quality and Safety in Health Care, 19*(2), 148–157.

14. Stull, D. E., Leidy, N. K., Parasuraman, B., & Chassany, O. (2009). Optimal recall periods for patient-reported outcomes: Challenges and potential solutions. *Current Medical Research and Opinion, 25*(4), 929–942.

15. U.S. Department of Health and Human Services, FDA, Center for Drug Evaluation and Research Center for Biologics Evaluation and Research Center for Devices and Radiological Health. Guidance for industry patient-reported outcome measures: Use in Medical Product Development to Support Labeling Claims. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed March 28, 2014.

16. Norquist, J. M., Girman, C., Fehnel, S., DeMuro-Mercon, C., & Santanello, N. (2012). Choice of recall period for patient-reported outcome (PRO) measures: Criteria for consideration. *Quality of Life Research, 21*(6), 1013–1020.

17. Heyer, G.L., Perkins, S.Q., Rose, S.C., Aylward, S.C., & Lee, J.M. (2013). Comparing patient and parent recall of 90-day and 30-day migraine disability using elements of the PedMIDAS and an Internet headache diary. Cephalalgia, Nov 8.

18. Broderick, J. E., Schwartz, J. E., Vikingstad, G., Pribbernow, M., Grossman, S., & Stone, A. A. (2008). The accuracy of pain and fatigue items across different reporting periods. *Pain, 139*(1), 146–157.

19. Broderick, J. E., Schneider, S., Schwartz, J. E., & Stone, A. A. (2010). Interference with activities due to pain and fatigue: Accuracy of ratings across different reporting periods. *Quality of Life Research, 19*(8), 1163–1170.

20. de Vries, S.T., Mol, P.G., de Zeeuw, D., Haaijer-Ruskamp, F.M., & Denig, P. (2013). Development and initial validation of a patient-reported adverse drug event questionnaire. *Drug Safety, 36*(9), 765–777.

21. World Health Organization Drug and therapeutics committee. Session 4. Assessing and managing medicine safety [online]. http://www.who.int/medicines/technical_briefing/tbs/04-Drug-Safety_final-08.ppt. Accessed March 28, 2014.

22. European commission. Enterprise and industry directorate-general. A guideline on summary of product characteristics (SmPC) September 2009. http://ec.europa.eu/health/files/eudralex/vol-2/c/smpc_guideline_rev2_en.pdf. Accessed February 3, 2014.

23. Schleyer, T. K., & Forrest, J. L. (2000). Methods for the design and administration of web-based surveys. *Journal of the American Medical Informatics Association, 7*(4), 416–425.

24. Randomizer recall. Randomly assignment of subjects to groups were performed using the Web site: GraphPad, QuickCalcs, Random number calculators. http://www.graphpad.com/quickcalcs/. Accessed March 28, 2014.

25. De Vet, H., Terwee, C., Mokkink, L., & Knol, D. (2011). *Measurement in medicine: A practical guide*. Cambridge: Cambridge University Press.

26. Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ, 308*(6943), 1552.

27. Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 2: Predictive values. *BMJ, 309*(6947), 102.

28. Deeks, J. J., & Altman, D. G. (1999). Sensitivity and specificity and their confidence intervals cannot exceed 100%. *BMJ, 318*(7177), 193–194.

29. Prohaska, V., Brown, N. R., & Belli, R. F. (1998). Forward telescoping: The question matters. *Memory, 6*(4), 455–465.

30. Sale, H., Hedman, L., & Isberg, A. (2010). Accuracy of patients' recall of temporomandibular joint pain and dysfunction after experiencing whiplash trauma: A prospective study. *Journal of the American Dental Association, 141*(7), 879–886.

31. Bent, S., Padula, A., & Avins, A. L. (2006). Brief communication: Better ways to question patients about adverse medical events: a randomized, controlled trial. *Annals of Internal Medicine, 144*(4), 257–261.

32. Suh, D. C., Choi, I. S., Plauschinat, C., Kwon, J., & Baron, M. (2010). Impact of comorbid conditions and race/ethnicity on glycemic control among the US population with type 2 diabetes, 1988–1994 to 1999–2004. *Journal of Diabetes and Its Complications, 24*(6), 382–391.

33. Jones, S. R., Carley, S., & Harrison, M. (2003). An introduction to power and sample size estimation. *Emergency Medicine Journal, 20*(5), 453–458.