

Original Article

Comparison of 7-day and repeated 24-hour recall of symptoms of cystic fibrosis[☆]

Antonia V. Bennett^{a,*}, Donald L. Patrick^b, James F. Lymp^c,
Todd C. Edwards^b, Christopher H. Goss^d

^a Health Outcomes Research Group, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 307 East 63rd St, 2nd Floor, New York, NY 10065, United States

^b Department of Health Services, University of Washington, Box 359445, Seattle, WA 98195, United States

^c Cystic Fibrosis Foundation Therapeutics Inc. Therapeutics Development Network Coordinating Center, Seattle Children's Research Institute, 1100 Olive Way, Suite 500, Seattle, WA 98101, United States

^d Department of Medicine and Department of Pediatrics, University of Washington Medical Center, Box 356552, Seattle, WA 98195, United States

Received 3 February 2010; received in revised form 25 June 2010; accepted 8 August 2010

Abstract

Background: Patient reporting of symptoms in a questionnaire with a 7-day recall period is expected to differ from reporting in daily symptom diaries.

Methods: 38 patients with cystic fibrosis (CF) completed 77 week-long symptom diaries. Each diary day comprised 13 symptom items with 5-point response scales. Days 1–6 of the diary had a 24-hour recall period. Day 7 had a 7-day recall period. Concordance of 7-day recall with summary descriptors of daily reports (e.g. mean, maximum) was examined and ability of 7-day recall and mean of daily reports to discriminate between well and ill periods of health compared.

Results: The average difference in scores was less than 0.25 response scale points. 7-day recall was most concordant with the mean of daily reports. Discriminant ability was comparable.

Conclusions: In this study sample, a questionnaire with 7-day recall provided information similar to a daily diary about the week-long experience of CF symptoms.

© 2010 European Cystic Fibrosis Society. Published by Elsevier B.V. All rights reserved.

Keywords: Questionnaires [MESH]; Mental recall [MESH]; Validation studies [MESH]; Signs and symptoms [MESH]; Cystic fibrosis [MESH]

1. Introduction

The respiratory symptoms patients with cystic fibrosis (CF) experience include difficulty breathing and/or shortness of breath, cough, sputum production, chest tightness, and

wheezing. These symptoms are sequelae of the underlying chronic airways infection and airway obstruction seen patients with CF [1]. In addition, patients report fever and feeling tired, emotional impacts including worry, irritability, frustration, sadness and/or depression, difficulty sleeping, and activity impacts including time spent sitting or lying down, reduction of usual activities, and missing school or work [2]. In many chronic diseases, including CF, patient reported outcome (PRO) measures are increasingly used to detect change in symptoms for the purposes of evaluating treatments and monitoring patient health [3,4]. Understanding the recall period for CF respiratory symptoms is key to clarifying at what interval symptoms should be assessed by questionnaires.

[☆] Preliminary results of this recall analysis were presented at a research meeting: Bennett AV, Patrick DL, Edwards TC, Lymp JF, Goss CH. Observed differences between daily diary and weekly report of cystic fibrosis symptoms. Poster. International Society for Pharmacoeconomics and Outcomes Research, Toronto, Canada, May 3–7, 2008.

* Corresponding author. Tel.: +1 646 735 8205; fax: +1 646 735 0011.

E-mail address: bennetta@mskcc.org (A.V. Bennett).

The Food and Drug Administration (FDA) has written detailed guidelines for the development of PRO measures to be used in clinical trials. One criterion for an acceptable measure is justification of the recall period [5]. Theory developed from the field of psychology about cognitive processes of memory support that retrospective recall may not be accurate [e.g. 6]. However the accuracy of short term retrospective recall i.e. recall over a period of one to four weeks, in measuring daily experience has been tested in only a few areas, specifically urinary incontinence [7–9], physical activity [10,11], and alcohol consumption [12,13]. In these studies, the correlation between daily diary and retrospective reporting, measured by Pearson’s correlation coefficient, varied from 0.33 to 0.89. A significant shortcoming of these studies is that findings are based on Pearson’s correlation coefficient, which does not incorporate differences in mean scores or the slope of the relationship between each type of measurement (e.g. if the two types of measurement were the same, the means would be equal and the slope would be 45°).

Retrospective recall has also been compared to the results of momentary reporting. Three studies of recall bias in pain measurement, comparing retrospective recall to ecological momentary report, have found 1) recalled pain is most similar to the maximum and last reports, 2) patients with greater variability in real-time reports of pain will recall a higher level of pain than the average of their real-time reports, and 3) while between-person correspondence between real-time reports and recall is moderate, the within-person correspondence is low [14–16].

This small body of research on short term retrospective recall supports that there is difference between retrospective report and daily experience, and suggests the amount of difference depends on the how the event is measured (e.g. occurrence or severity) as well as the patterns of response. Only two of the areas previously researched are symptoms (incontinence and pain) [7–9,14–16]. It is not known if the difference between daily and retrospective measurement identified in these symptoms and activities will be found consistently in assessment of respiratory symptoms.

This study had two aims. The first aim was further exploration of the relationship between weekly and daily diary reporting. It was hypothesized based on the results of published studies that the weekly response would be most similar to the maximum daily report and the most recent daily report, in this case Day 6. The second aim was to compare the ability of each type of measurement to detect differences between periods of being well and periods of pulmonary exacerbation. A hypothesis was not put forth for the second aim as there was no existing evidence to support it.

2. Methods

2.1. Diary

The CF Respiratory Symptom Diary (CFRSD) is a 16 item measure (13 of which are analyzed here) of symptoms and impacts of CF. The diary was developed via patient interviews and clinician input [2]. As part of a preliminary validation study [17], the CFRSD was completed for 7 days. On days 1–6

patients report the occurrence, frequency and severity of symptoms and impacts in the past 24 hours. On day 7 patients report the occurrence, frequency and severity of symptoms and impacts during the past 7 days. The analyses reported in this paper were conducted on the patient reports of symptom frequency. The questionnaire items included in the analysis are shown in Fig. 1.

Symptom frequency is measured in the daily diary and weekly questionnaire by items that are identical except for the recall period. Daily diary items ask, for example, “Overall during the past 24 hours, how often did you have difficulty breathing?” The corresponding weekly questionnaire item asks: “Overall during the last 7 days, how often did you have difficulty breathing?” The response options for both items are: no (did not have difficulty breathing), a little bit of the time, some of the time, much of the time, and all of the time, which are coded here as 0, 1, 2, 3, 4. For each symptom, the weekly report is the answer to the item from the weekly questionnaire, and the mean daily report is the mean of the six daily responses from the daily diary.

2.2. Data collection

Data was collected through an observational study. Patients were recruited through two CF clinics in Seattle, WA, the University of Washington Adult CF clinic and Seattle

Stem for each item in Daily Diary –		Stem for each item in 7-Day Questionnaire –	
“Because of cystic fibrosis, during the last 24 hours:		“Because of cystic fibrosis, during the last 7 days:	
Group	Item text		
Respiratory	Did you have difficulty breathing?		
	Did you cough?		
	Did you cough up mucus?		
	Did you have tightness in the chest?		
	Did you wheeze?		
Tired	Did you feel tired?		
Temperature	Did you feel feverish (have a temperature)?		
	Did you have chills or sweats?		
Mood	Did you have difficulty sleeping?		
	Did you feel worried?		
	Did you feel cranky?		
	Did you feel sad or depressed?		
	Did you feel frustrated?		
Response Options for Each Item			
Yes			
No (If no, go to next question)			
If Yes, patient answers follow- up question:			
Overall during the [recall period], how often [item text] ?			
A little bit of the time			
Some of the time			
Much of the time			
All of the time			

Note: This figure describes only the 13 CFRSD items included in this analysis

Fig. 1. CF Respiratory Symptoms Diary (CFRSD) items.

Children's Hospital CF clinic. The clinics and patients were chosen based on convenience and availability. Although children from 2 years and older were recruited, only subjects 12 years and older are included in this analysis. Subjects who had undergone prior solid organ transplantation were excluded. Following enrollment into the study, patients completed the diary for 7 days at a time, twice while they were well and once when they were ill, if patients were ill during the study period. Patients were deemed ill when they were experiencing an exacerbation that required clinical intervention beyond their usual clinical care. Enrolled patients were asked to contact the study coordinator when they were beginning a period of being ill so that this could be recorded and study forms could be provided.

Demographic information, including age, gender, ethnicity and education level, was collected during the patient screening process. Patients completed the diary each day during the 7-day well or ill period through a web-based questionnaire. Patients were prompted by phone call from the study coordinator if they did not complete the diary on any given day, and the data taken over the phone if necessary. If patients knew they would be away from internet access during the diary period, they were provided with paper versions of the diary questionnaire to complete on those days.

Institutional Review Board approval was granted for the study and all enrolled patients provided informed consent. Data collection and analysis were conducted in accordance with ethical standards described in the 1964 Declaration of Helsinki.

2.3. Statistical analyses

Analyses were conducted on observations (daily responses and weekly item response for one item) which were not missing the weekly item response. Two-sided *p*-values were employed throughout. Statistical significance was defined when $p < 0.05$. Standard errors were estimated using the clustered sandwich estimator [18] to account for correlation from multiple items in each item group and multiple diaries per patient.

2.4. Aim 1: Exploration of the relationship between weekly and daily diary reports

The weekly item response was compared with summary descriptors of the daily diary item responses: mean, median, mode, maximum (most severe), minimum (least severe), day 1 of 6, and day 6 of 6. Since it is possible the two measures could have the same mean without being correlated, or be highly correlated without having the same mean, the two measures were compared using the Concordance Correlation Coefficient (CCC). This statistic is composed of Pearson's correlation coefficient and a bias correction factor calculated from the differences in the means and variance of the two variables. The bias correction factor indicates the amount of deviation of the Pearson's best-fit line from the 45° origin line (the concordance line) [19,20]. The CCC ranges from -1 to $+1$, where absolute values of the CCC closer to 1 indicate greater concordance (negative CCC values indicate reverse concordance). The bias

correction factor ranges from 0 to 1, and values closer to 1 indicate less difference between the Pearson's best-fit line and the 45° origin line. A similar statistic of agreement, the intra-class correlation coefficient (ICC) has the disadvantage that it cannot be composed into the correlation coefficient and the bias factor.

The sample mean and 95% confidence intervals were calculated for the weekly report, for each summary descriptor of the daily responses, and for the difference between the weekly report and each summary descriptor of the daily responses.

2.5. Aim 2: Compare the ability of weekly and mean of daily reports to detect differences between health states

To compare the ability of weekly and mean of daily reports to detect differences between health states, item scores of well-period diaries were compared with the item scores of ill-period diaries. For each type of measurement, a *t*-statistic was calculated for the difference in mean scores between well and ill period diaries, where the null hypothesis is that the difference is zero. The effect size of the difference in mean scores between well and ill period diaries, calculated as Cohen's *d*—the difference in means divided by the pooled standard deviation, is also reported [21,22].

3. Results

3.1. Patient characteristics

Fifty-two patients CF patients 12 years and older were screened, and all were eligible. Twelve patients could not be scheduled for the initial study visit (and consented) due to family time conflicts. Of the 40 patients attending the initial study visit, 38 agreed to participate in study. Study participants completed a total of 96 diaries. Nineteen diaries were excluded from analysis because they were missing the weekly report data. The remaining 77 diaries represent 38 patients. These diaries had very little missing data; of the 462 diary days (77 diaries \times 6 days) only 12 diary days (2.6%) were not completed by the patients.

The 38 patients ranged in age from 12 to 38 years old. Thirteen (34%) were between the ages of 12 and 18, and 25 (66%) were between the ages of 19 and 38. The mean (SD) age was 21.8 (7.0) years. Twenty-four (63%) were female and 32 (84%) were white. Fifteen (58%) of patients age 18 or older reported completing at least some college (see Table 1).

3.2. Diary characteristics

Seventy-seven diaries were analyzed, of which 30 were ill-period diaries and 47 were well-period diaries. Eight patients (21%) completed 1 diary, 21 patients (55%) completed 2 diaries, and 9 patients (24%) completed 3 diaries. Twenty-eight patients (74%) completed 1 ill-period diary and at least 1 well-period diary.

Table 1
Patient characteristics.

N=38	N	%
Age group		
Age 12–18	13	34
Age 19–24	15	39
Age 25+	10	26
Age, mean (SD) and range	21.8 (7.0)	12–38
Gender		
Female	24	63
Male	14	37
Education level of patients age 18+		
Some high school	3	12
High school graduate/GED	3	12
Some college	7	27
College graduate	8	31
Not reported	5	19
Ethnicity		
White	32	84
White and African-American	1	3
White and Hispanic	1	3
White and Native American and Asian	1	3
Not reported	3	8

The ill-period and well-period diaries reflected the difference in health status of the patients during each period. The ill-period diaries compared to the well-period diaries had significantly higher mean (95% CI) daily frequency of symptoms [0.64 (0.44, 0.82) vs 0.33 (0.23, 0.43); $p < 0.001$].

The within-subject variation in daily item responses within each diary was low. The average standard deviation of responses to an item within one diary ranged from 0.11 to 0.48 response scale points across the 4 item groups; it was largest for respiratory items (0.35) and the tired item (0.48).

3.3. Aim 1: Exploration of the relationship between weekly and daily diary reports

The comparison of the weekly response to each descriptor of the daily responses showed consistent trends among the 4 item groups (Table 2). The weekly response had the highest concordance as well as the highest correlation with the mean of daily responses compared to other descriptors. The single exception to this trend was that in the temperature items group the concordance of the weekly report with the mode was slightly higher than the concordance of the weekly report with the mean (0.88 vs 0.86). The concordance statistic for the comparison of the weekly and mean of daily responses ranged from 0.72 to 0.86 across the 4 item groups. In each item group the concordance statistic was lower than the correlation statistic (range 0.78 to 0.92), because the bias correction factor was less than 1.00 (range 0.92 to 0.99).

The weekly report was higher than the mean of daily responses by less than one-quarter a response scale interval (Table 3). The average difference between the weekly report and mean of daily responses for 3 item groups (respiratory, tired, and mood items) ranged from 0.10 to 0.16 ($p < 0.05$), and for the temperature items group it was not different from zero

Table 2
Concordance of the weekly report with statistical descriptors of the daily diary item responses.

Item Group Descriptors	CCC (r-b)	Pearson's (r)	Bias Factor (b)
Respiratory items (5 items, 385 observations)			
Mean	0.85	0.85	0.99
Median	0.83	0.85	0.98
Mode	0.82	0.83	0.99
Maximum	0.71	0.77	0.93
Minimum	0.65	0.75	0.86
Day 6	0.80	0.81	0.99
Day 1	0.70	0.70	1.00
Tired items (1 item, 77 observations)			
Mean	0.82	0.85	0.96
Median	0.82	0.85	0.96
Mode	0.79	0.81	0.98
Maximum	0.68	0.75	0.91
Minimum	0.45	0.60	0.74
Day 6	0.73	0.77	0.96
Day 1	0.61	0.62	0.99
Temperature items (2 items, 154 observations)			
Mean	0.86	0.92	0.94
Median	0.83	0.91	0.91
Mode	0.88	0.91	0.97
Maximum	0.70	0.75	0.93
Minimum	0.51	0.67	0.76
Day 6	0.62	0.65	0.95
Day 1	0.66	0.67	0.99
Mood items (5 items, 385 observations)			
Mean	0.72	0.78	0.92
Median	0.64	0.72	0.89
Mode	0.64	0.68	0.94
Maximum	0.66	0.71	0.93
Minimum	0.36	0.54	0.67
Day 6	0.65	0.67	0.98
Day 1	0.66	0.66	1.00

Note: Mean=average of daily reports, Median=median of daily reports, Mode=most common daily report (larger value used in case of tie), Maximum = most severe response, Minimum=least severe response, Day 6=response of the day before the weekly report was completed, Day 1=first daily report. CCC=Concordance Correlation Coefficient.

One diary was not completed on Day 6, therefore the number of observations for each item group was Respiratory=380, Tired=76, Temperature=152, and Mood=380.

(0.02; $p = 0.175$). The weekly report was lower than the maximum daily response in each item group. The average difference between the weekly report and maximum daily response ranged from -0.15 to -0.52 ($p < 0.01$).

3.4. Aim 2: Compare the ability of weekly and mean of daily reports to detect differences between health states

For each item group, the difference between ill-period and well-period diaries was estimated for the weekly report and the mean of daily responses (Table 4). For 3 of the 4 item groups (respiratory, tired, and temperature items) the absolute value of the t-statistic was larger for the mean of daily responses than for the weekly report. However, in each of these 3 item groups, the

Table 3

Differences between the weekly report and the mean and maximum of daily reports.

Item Group	Respiratory Obs=385		Tired Obs=77		Temperature Obs=154		Mood Obs=385	
	Mean	CI	Mean	CI	Mean	CI	Mean	CI
Weekly Report	0.86	[0.66 1.06]	0.95	[0.61 1.28]	0.13	[0.02 0.23]	0.29	[0.15 0.43]
Mean of Daily	0.76	[0.57 0.95]	0.78	[0.52 1.05]	0.11	[0.02 0.19]	0.20	[0.09 0.31]
Difference ^a	0.10	[0.03 0.17]	0.16	[0.02 0.31]	0.02	[−0.01 0.06]	0.09	[0.04 0.14]
t-statistic, p-value	2.76	0.009	2.33	0.026	1.38	0.175	3.70	0.001
Maximum Daily	1.24	[1.01 1.47]	1.47	[1.09 1.84]	0.28	[0.09 0.47]	0.50	[0.28 0.73]
Difference ^b	−0.38	[−0.49 −0.27]	−0.52	[−0.73 −0.31]	−0.15	[−0.25 −0.04]	−0.21	[−0.31 −0.11]
t-statistic, p-value	6.97	<0.001	4.93	<0.001	2.90	0.006	4.36	<0.001

Note: The t-statistics and p-values shown were calculated for the difference between the weekly report and the mean of daily reports and for the difference between the weekly report and maximum daily report. The number of observations reflects multiple items per group and multiple diaries per patient (N=38). Obs=observations. CI=95% confidence interval.

^a Difference between the Weekly Report and Mean of Daily Reports.

^b Difference between the Weekly Report and the Maximum of Daily Reports.

significance of the t-statistic for the weekly report and the daily diary was the same. In the fourth item group (mood items), the absolute value of the t-statistic was larger for the weekly response and the significance of the t-statistics differed; the p-value for the weekly report was 0.012 and was 0.054 for the mean of daily responses. For each item group, the effect size of the weekly report and the mean of daily reports for the difference between ill and well period were similar. The effect size of the weekly report was larger than that of the mean of daily reports for respiratory items (0.50 vs 0.49) and for mood items (0.37 vs 0.25). The effect size of the weekly report was smaller than the mean of daily reports for the tired item (0.61 vs 0.81) and temperature items (0.05 vs 0.10).

4. Discussion

In this study sample the weekly report was consistently higher than the mean of daily reports. The difference between the weekly report and the mean of daily reports was less than one quarter of a response scale interval. When the weekly report

was compared to a variety of summary descriptors of the daily reports (e.g. mean, median, maximum, etc) the weekly report was highly concordant with the mean of daily reports. These results support that although the weekly report was slightly higher than the mean of the daily reports, the weekly reports of the study subjects were most similar to an average of the their daily symptom experience during the recall period. The hypothesis that the weekly response would be most similar to the maximum daily report and the most recent daily report, in this case Day 6, was not confirmed in this analysis. The primary focus of this study was addressing weekly vs daily recall which represents an early stage in instrument validation. The clinically relevant differences or response criteria of the symptom scores are not yet known.

The weekly report and mean of daily responses were comparable in their ability to detect differences in item scores between ill and well periods of health. The amount of difference detected between ill and well period diaries, as measured by the significance of the t-statistics and by the effect size, was consistent between the two types of measurement.

Table 4

Sensitivity of the weekly report and mean of daily reports to differences between ill and well period diary item scores.

Item Group	Ill period		Well period		Difference between periods				
	Mean	CI	Mean	CI	Mean	CI	t	p	ES
Respiratory	Obs. = 235		Obs. = 150						
Weekly	1.15	[0.83 1.47]	0.67	[0.52 0.83]	0.48	[0.22 0.75]	3.692	0.001	0.50
Mean Daily	1.03	[0.76 1.30]	0.58	[0.42 0.75]	0.45	[0.24 0.66]	4.332	<0.001	0.49
Tired	Obs. = 47		Obs. = 30						
Weekly	1.33	[0.85 1.82]	0.70	[0.36 1.04]	0.63	[0.17 1.09]	2.803	0.008	0.61
Mean Daily	1.23	[0.84 1.61]	0.50	[0.26 0.75]	0.73	[0.40 1.06]	4.476	<0.001	0.81
Temperature	Obs. = 94		Obs. = 60						
Weekly	0.15	[−0.01 0.31]	0.12	[0.01 0.23]	0.03	[−0.13 0.20]	0.408	0.686	0.05
Mean Daily	0.14	[0.01 0.28]	0.08	[0.00 0.16]	0.06	[−0.05 0.17]	1.054	0.299	0.10
Mood	Obs. = 235		Obs. = 150						
Weekly	0.47	[0.23 0.72]	0.17	[0.04 0.30]	0.30	[0.07 0.53]	2.659	0.012	0.37
Mean Daily	0.30	[0.12 0.48]	0.13	[0.04 0.23]	0.17	[0.00 0.34]	1.993	0.054	0.25

Note: The number of observations reflects multiple items per item group and multiple diaries per patient (N=38). Analysis is based on all ill and well period diaries, not matched pairs. For each item group, the larger t-statistic is shown in bold. ES=Effect size.

During the design of this analysis one consideration was that patient perceptions of their symptoms changed as a result of completing the daily diary; increased awareness of symptoms could affect reports in subsequent days of the diary as well as the weekly report. However, a similar study (not yet published) in patients with type 2 diabetes compared the weekly reports of patients randomized to two groups in which one group completed the daily diary during the recall period of the weekly report and the other group did not. The differences in the weekly report scores were very small and not statistically significant. This finding suggests completing the daily diary during the recall period of the weekly report does not affect the weekly report item responses [23].

Strengths of this study included that the wording and layout of the questionnaire used in the weekly report was identical to the daily diary except that the recall period was changed. In addition, there was very little missing daily diary data because patients were given reminder calls, however 19 diaries did not have the weekly report on Day 7 perhaps due to respondent fatigue and lack of reminder calls for the weekly questionnaire. Although the study sample was small ($n=38$), patients contributed a total of 77 diaries which were completed during both ill and well periods of health. It was possible with this data to compare the sensitivity of each type of measurement to differences in scores between ill and well period diaries, however it is not known which type of measurement would be more sensitive to detecting change over time with treatment with a drug or therapeutic intervention. A limitation to the generalizability of results is that more rigid definitions of an exacerbation are also in use (e.g. Rosenfeld's criteria) but were not used in this study.

This study compared two methods for measuring the 7-day symptom experience of patients with CF, in which the two methods (a single 7-day recall and repeated 24-hour recall) were found to provide similar results for *groups* of patients. In designing a study, investigators will need to choose both the recall period (e.g. 24-hour or 7-day) and the measurement period (e.g. 1-day, 3-day, or 7-day symptom experience). The extra burden on study participants of completing the diary each day compared to a single weekly report, and the additional study resources required for collecting and analyzing diary data are practical considerations for selecting the recall period. The choice of measurement period would depend on the goals of the study. In a clinical practice setting, a diary might be useful to facilitate early detection of a CF exacerbation if clinic staff were promptly notified of high daily scores.

Conflict of interest statement

The authors report no financial or personal relationships which could inappropriately influence this study.

Acknowledgements

Funding for this study was provided in part by the Cystic Fibrosis Foundation, through a research grant (GOSS05A0) and the Leroy Matthew's Physician Scientist Award, and by grants from NIH (RR-00037-39) and NIH/NHLBI (K23 HL72017).

References

- [1] Gibson R, Burns J, Ramsey B. Pathophysiology and management of pulmonary infections in cystic fibrosis. *Am J Respir Crit Care Med* 2003;168:918–51.
- [2] Goss CH, Edwards TC, Ramsey BW, Aitken ML, Patrick DL. Patient-reported respiratory symptoms in cystic fibrosis. *J Cyst Fibros* Jul 2009;8(4):245–52.
- [3] Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res* 2000;9(8):887–900.
- [4] Goss CH, Quittner AL. Patient-reported outcomes in cystic fibrosis. *Proc Am Thorac Soc* 2007 Aug 1;4(4):378–86.
- [5] FDA. Patient Reported Outcome (PRO) Measures: use in medical product development to support labeling claims. Federal Register, 74; 2009. p. 235.
- [6] Hufford MR, Shiffman S. Assessment methods for patient-reported outcomes. *Dis Manage Health Outcomes* 2003;11(2):77–86.
- [7] Homma Y, Ando T, Yoshida M, Kageyama S, Takei M, Kimoto K, et al. Voiding and incontinence frequencies: variability of diary data and required diary length. *Neurourol Urodyn* 2002;21(3):204–9.
- [8] Kenton K, Fitzgerald MP, Brubaker L. What is a clinician to do—believe the patient or her urinary diary? *J Urol* Aug, 2006;176(2):633–5 discussion 5.
- [9] Elser DM, Fantl JA, McClish DK. Comparison of "subjective" and "objective" measures of severity of urinary incontinence in women. Program for Women Research Group. *Neurourol Urodyn* 1995;14(4):311–6.
- [10] Richardson MT, Ainsworth BE, Jacobs DR, Leon AS. Validation of the Stanford 7-day recall to assess habitual physical activity. *Ann Epidemiol* 2001 Feb;11(2):145–53.
- [11] Stel VS, Smit JH, Pluijm SM, Visser M, Deeg DJ, Lips P. Comparison of the LASA Physical Activity Questionnaire with a 7-day diary and pedometer. *J Clin Epidemiol* 2004 Mar;57(3):252–8.
- [12] Gmel G, Daeppen JB. Recall bias for seven-day recall measurement of alcohol consumption among emergency department patients: implications for case-crossover designs. *J Stud Alcohol Drugs* 2007 Mar;68(2):303–10.
- [13] Hilton ME. A comparison of a prospective diary and two summary recall techniques for recording alcohol consumption. *Br J Addict* 1989 Sep;84(9):1085–92.
- [14] Stone AA, Broderick JE, Kaell AT, DelesPaul PA, Porter LE. Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritis? *J Pain* Autumn, 2000;1(3):212–7.
- [15] Stone AA, Broderick JE, Shiffman SS, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain* 2004 Jan;107(1–2):61–9.
- [16] Stone AA, Schwartz JE, Broderick JE, Shiffman SS. Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Pers Soc Psychol Bull* Oct 2005;31(10):1340–6.
- [17] Goss C, Edwards T, Lymp J, Ramsey B, Aitken M, Patrick D. Patient reported respiratory symptoms in cystic fibrosis with the CF Respiratory Symptom Diary. *Pediatr Pulmonol* 2008;43(S31):A691.
- [18] Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986 Mar;42(1):121–30.
- [19] Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989 Mar;45(1):255–68.
- [20] Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials* Aug, 1991;12(4 Suppl):142S–58S.
- [21] Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Ed. ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
- [22] Hartung J, Knapp G, Sinha B. Statistical meta-analysis with applications. Hoboken, NJ: John Wiley and Sons, Inc.; 2008.
- [23] Bennett AV, Patrick DL, Bushnell DM, Chiou CF, Diehr P. Comparison of 7-day and repeated 24-hour recall of type 2 diabetes. Submitted.