Susan M Rogers
7 January 2011

Integrity Check for Chronic Prostatitis Collaborative Research Network 2 (CPCRN2) - Alfuzosin Study

As a partial check of the integrity of the CPCRN2-Alfuzosin dataset archived in the NIDDK data repository, a set of tabulations was performed to verify that published results can be reproduced using the archived dataset. Analyses were performed to duplicate selected results for the data published by Nickel et al [1] in the *New England Journal of Medicine* in December 2008. The results of this dataset integrity check (DSIC) are described below. The full text of the *New England Journal of Medicine* article can be found in Attachment 1 and the STATA code for our tabulations is included in Attachment 2.

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is ***not*** to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected on a first (or second) exercise in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC), however this process is labor-intensive for both DCC and Repository staff. It is thus not our policy to resolve every discrepancy that is observed in an integrity check. Specifically, we do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, *unless NIDDK Repository staff suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff*. We do, however, document in footnotes to the integrity check those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

Background. The Chronic Prostatitis Collaborative Research Network 2 Alfuzosin study was a multicenter randomized double-blind placebo-controlled clinical trial to evaluate the efficacy of alfuzosin in reducing symptoms associated with chronic prostatitis-chronic pelvic pain in men. Men were randomly assigned to treatment for 12 weeks with 10 mg of either alfuzosin, an alpha-adrenergic receptor blocker, or placebo. The primary outcome was a reduction of at least 4 points on the NIH Chronic Prostatitis Symptom Index (NIH-CPSI, range 0-43 with higher scores indicating more severe symptoms) from baseline to 12 weeks. An additional outcome was measured using a 7-point global response assessment (GRA) of self-reported change in chronic prostatitis-related symptoms. Men who reported moderate or marked improvement in the GRA at the end of the study were considered treatment responders.

Other measures included assessment of general pain and urinary urgency on a Likert scale, the McGill Pain Questionnaire, the Medical Outcomes Study Short Form Health Survey 12 (physical and mental component summaries), the Hospital Anxiety and Depression Scale, the International Index of Erectile Function, and the Male Sexual Health Questionnaire.

There were four clinic visits during which data for the outcomes were measured: visit 1 involved screening, visit 2 involved baseline data collection and randomization, visit 3 was the 6-week evaluation and visit 4 was the 12-week and final evaluation. A total of 272 participants were randomized. The paper by Nickel et al. provides the main study results for the efficacy trial.

In this DSIC, we compared our results to the published results in Table 1 (Baseline Characteristics of the Study Subjects), Table 2 (Response Rates for NIH-CPSI and Global Response Assessment According to Study Group) and Table 3 (Changes at 12 Weeks in Scores for Measures of Secondary Outcomes According to Study Group).

The tables in this DSIC reproduce closely the results presented in Tables 1 through 3.

Baseline Characteristics of the Study Subjects. Table A lists the data variables and algorithms for derived analysis variables that we used in our replication for Table 1 of Nickel et al. The dataset "cp01_long.sas7bdat" is an analysis file provided by the DCC along with the form-based files. We did not use the original form-based variables – rather we used the variables as contained in cp01_long, the analysis file.

TABLE A Variables and analysis variable algorithms for Nickel et al (2008), Table 1

	Analysis dataset variable	Analysis algorithm
Treatment group	arm	egen treatment=group(arm)
Age, mean/range	age base	
Race — no. (%)†	race	recode (5=1) (3=2) (2=3) (else=4)
NIH-CPSI		
Total score	cpsibase	
Pain score	painbase	
Urinary score	uribase	
Quality-of-life score	qolbase	
Likert Pain and Urinary Urgency Scale		
Average pain score	cppainv1	
Urgency score	urgv1	
McGill Pain Questionnaire, No.		
Total score	mcgill_total_v2	
Sensory score	mcgill_sensory_v2	
Affective score	mcgill_affective_v2	
SF-12, No.		
Physical-component summary	pcs12_v2	
Mental-component summary	mcs12_v2	
Hospital Anxiety and Depression Scale, No.		
Total score	had15_v2	
International Index of Erectile Function, No.		
Total score	iief_total_v2	
Male Sexual Health Questionnaire, No.		
Total score	mshq_total_v2	

The published results and the DSIC results for Table 1 of the manuscript are shown in Table B. The base Ns and mean values for the baseline characteristics and primary and secondary measures calculated by the DSIC match very closely the published values. In the published article, P values were calculated controlling for clustering by clinical center; these values were not recreated in this DSIC.

TABLE B. Baseline Characteristics of the Study Subjects (Table 1, Nickel et al, 2008)

Characteristic	From Nickel et al., 2008 Placebo Group (N = 134)	From Nickel et al., 2008 Alfuzosin Group (N = 138)	P Value	From DSIC Placebo Group (N = 134)	From DSIC Alfuzosin Group (N = 138)
Age — yr			0.99		
Mean	40.1±12.3	40.1±11.4		40.1±12.3	40.1±11.4
Range	19–66	19–68		19–66	19–68
Race — no. (%)†					
White	97 (72.4)	88 (63.8)	0.19	97 (72.4)	88 (63.8)
Black	16 (11.9)	20 (14.5)		16 (11.9)	20 (14.5)
Asian	14 (10.4)	13 (9.4)		14 (10.4)	13 (9.4)
Other	7 (5.2)	17 (12.3)		7 (5.2)	17 (12.3)
NIH-CPSI					
Total score	25.1±5.9	23.8 ±6.3	0.06	25.1±5.9	23.8 ±6.3
Pain score	11.5±3.4	11.1±3.3	0.15	11.5±3.4	11.1±3.3
Urinary score	4.9±2.9	4.5±2.8	0.3	4.9±2.9	4.6±2.8
Quality-of-life score	4.7±0.9	4.5±1.1	0.15	4.7±0.9	4.5±1.1
Likert Pain and Urinary Urgency Scale					
Average pain score	5.0±1.9	4.9±2.0	0.52	4.9±2.2	4.9±2.0
Urgency score	4.7±2.6	4.4±2.6	0.4	4.8±2.7	4.3±2.7
McGill Pain Questionnaire					
No. evaluated	133	134		133	134
Total score	11.6±8.7	11.2±8.6	0.6	11.6±8.7	11.2±8.6
Sensory score	8.9±6.2	8.6±5.9	0.55	8.9±6.2	8.5±5.9
Affective score	2.6±3.2	2.6±3.3	0.83	2.6±3.2	2.6±3.3
SF-12					
No. evaluated	130	136	0.84	130	136
Physical-component summary	45.6±8.4	45.5±9.6	0.5	45.6±8.4	45.5±9.6
Mental-component summary	44.9±10.2	44.0±10.9		44.9±10.2	44.0±10.9
Hospital Anxiety and Depression Scale					
No. evaluated	133	137		133	137
Total score	12.8±7.1	12.8±7.6	0.81	12.8±7.1	12.8±7.6
International Index of Erectile Function					
No. evaluated	126	133		126	133
Total score	52.8±17.4	53.5±17.9	0.57	52.8±17.4	53.5±17.9
Male Sexual Health Questionnaire					
No. evaluated	126	130		126	130
Total score	30.4±6.6	30.0±7.4	0.81	30.4±6.6	30.0±7.4

Note: DSIC analysis for Likert average pain score was based on Ns of 131 and 135 and urgency score was based on Ns of 132 and 135 subjects, in the placebo and alfuzosin groups, respectively.

Response Rates for NIH-CPSI and Global Response Assessment According to Study Group. Table C contains the published results and the DSIC results for Table 2 by treatment group.

Primary outcomes. The response rate for the National Institutes of Health Chronic Prostatitis Symptom Index (NIH-CPSI) was based on a decline in the total score of =4 at 12 weeks; a high score indicates more severe symptoms. A primary endpoint was a participant reported global response assessment at 12 weeks relative to overall baseline symptoms. A 7-point centered scale was used and included markedly worse, moderately worse, slightly worse, no change, slightly improved, moderately improved, and markedly improved. Participants who reported either of the last two categories (marked or moderate improvement) were defined as treatment responders.

Table C. Response Rates for NIH-CPSI and Global Response Assessment According to Study Group

	From Nickel et al., 2008			From DSIC *	
Measure	Placebo Group (N = 134)	Alfuzosin Group (N = 138)	Absolute Difference in Rates	Placebo Group (N = 134)	Alfuzosi Group (N = 138
	no. (%)		% (95% CI)	no. (%)	
NIH-CPSI score, decline of =4 points (primary efficacy end point)	66 (49.3)	68 (49.3)	0.1 (.11.2 to 11.0)	66 (49.3)	68 (49.3)
Global response assessment					
Marked or moderate improvement	45 (33.6)	48 (34.8)	1.8 (-9.0 to 2.5)	45 (36.6)	48 (40.3)
Marked improvement	16 (11.9)	23 (16.7)		16 (13.0)	23 (19.3)
Moderate improvement	29 (21.6)	25 (18.1)		29 (23.6)	25 (21.0)
Slight improvement	32 (23.9)	26 (18.8)		32 (26.0)	26 (21.8)
No change	36 (26.9)	36 (26.1)		36 (29.3)	36 (30.2)
Slight worsening	5 (3.7)	6 (4.3)		5 (4.1)	6 (5.0)
Moderate worsening	4 (3.0)	2 (1.4)		4 (3.2)	2 (1.7)
Marked worsening	1 (0.7)	1 (0.7)		1 (0.8)	1 (0.8)

Notes:

* The NIH-CPSI score, decline of =4 points was derived from the variable, cpsibin4pt, in the analysis dataset. The global response assessment values were derived from the variable, gralast. While the Ns for each category of the 7-point global response assessment match the published Ns, they actually total 123 for the placebo group and 119 for the alfuzosin group (not 134 and 138 as indicated in the table heading). Percentages in the published data reflect the Ns of 134 and 138 for the placebo and alfuzosin groups, respectively, whereas the percentages from the DSIC reflect total Ns of 123 and 119, respectively.

The absolute difference in rates (and 95% CI) was not calculated in this DSIC exercise.

Changes at 12 Weeks in Scores for Measures of Secondary Outcomes According to Study. Table D lists the data variables used in our replication of Table 3 of Nickel et al. and Table E presents the published and DSIC results. Our DSIC tabulations provided estimates of changes at 12 weeks in secondary outcome scores identical to the published results; 95% CIs and *P* values were similar, but not identical due to the use of different statistical commands for calculating estimate variance (see Note at end of Table E).

TABLE D. Changes at 12 Weeks in Scores for Measures of Secondary Outcomes According to Study Group

	Analysis dataset variable	Analysis algorithm
NIH-CPSI		
Total score (0–43)	cpsidiff_12wk	gen cpsichange=cpsiv4-cpsibase
Pain score (0–21)	paindiff_12wk	gen painchange=painv4-painbase
Urinary score (0–10)	uridiff_12wk	gen urichange=uriv4-uribase
Quality-of-life score (0–12)	qoldiff_12wk	gen qolchange=qolv4-qolbase
McGill Pain Questionnaire		
Total score (0–45)	mcgill_total_12wkcng	
Sensory score (0–33)	mcgill_sensory_12wkchn	
Affective score (0–12)	mcgill_affective_12wkchng	
SF-12		
Physical component summary (0–100)	pcs12_12wkcng	
Mental-component summary (0–100)	mcs12_12wkcng	
Hospital Anxiety and Depression Scale		
Total score	had15_12wkcng	
International Index of Erectile Function		
Total score (0–43)	iief_total_12wkcng	
Male Sexual Health Questionnaire		
Total score (0–40)	mshq_total_12wkcng	

Secondary outcomes. For the NIH-CPSI, higher scores indicate more severe symptoms (for the quality-of-life score, higher scores indicate a more negative effect). Score ranges are as follows: total score, 0 to 43; pain score, 0 to 21; urinary score, 0 to 10, quality-of-life score, 0 to 12; and average pain and urgency scores, 0 to 10. For the McGill Pain Questionnaire, higher scores indicate greater pain. Score ranges are as follows: total score, 0 to 45; sensory score, 0 to 33; affective score, 0 to 12. For the Medical Outcomes Study Short Form Health Survey 12 (SF-12), higher scores indicate better quality of life. Score range for both the physical and mental component summaries is 0 to 100. For the Hospital Anxiety and Depression Scale, higher scores indicate greater anxiety and depression; range, 0 to 42. For the International Index of Erectile Function, higher scores indicate better sexual function; range, 0 to 75. For the Male Sexual Health Questionnaire, higher scores

indicate better function with respect to erection and ejaculation and greater satisfaction with sexual life; range, 0 to 40.

TABLE E. Changes at 12 Weeks in Scores for Measures of Secondary Outcomes According to Study Treatment

	From Nickel et al., 2008				From DSIC			
Measure	Placebo Group (N = 134)	Alfuzosin Group (N = 138)	Absolute Difference Between Groups (95% CI)	P Value*	Placebo Group (N = 134)	Alfuzosin Group (N = 138)	Absolute Difference Between Groups (95% CI)	P Value*
NIH-CPSI								
No. evaluated	117	116			117	116		
Total score (0–43)	-6.5±8.5	-7.1±9.0	-0.6 (-2.7,1.5)	0.7	-6.5±8.5	-7.1±9.0	-0.6 (-2.7,1.5)	0.58
Pain score (0–21)	-3.0±4.4	-3.3±4.5	-0.3 (-1.4,1.0)	0.64	-3.0±4.4	-3.3±4.5	-0.3 (-1.4, 1.0)	0.61
Urinary score (0–10)	-1.0±2.6	-1.2±2.6	-0.2 (-0.8,0.4)	0.62	-1.0±2.6	-1.2±2.6	-0.2 (-0.8, 0.4)	0.60
Quality-of-life score (0–12)	-1.2±1.5	-1.2±1.5	0 (-0.4,0.4)	0.99	-1.2±1.5	-1.2±1.5	0 (-0.4,0.4)	0.87
McGill Pain Questionnaire								
No. evaluated	116	112			116	112		
Total score (0–45)	-3.1±6.5	-3.4±6.4	-0.3 (-1.8,1.2)	0.45	-3.1±6.5	-3.4±6.4	-0.3 (-2.0,1.4)	0.73
Sensory score (0–33)	-2.3±4.9	-2.5±5.0	-0.2 (-1.4,1.0)	0.47	-2.3±4.9	-2.5±5.0	-0.2 (-1.5,1.1)	0.72
Affective score (0–12)	-0.9±2.3	-1.0±2.1	-0.1 (-0.6,0.4)	0.89	-0.9±2.3	-1.0±2.1	-0.1 (-0.7,0.5)	0.75
SF-12								
No. evaluated	113	115			113	115		
Physical summary (0–100)	3.5±8.1	3.0±7.4	-0.5 (-2.3,1.3)	0.6	3.5±8.1	3.0±7.4	-0.5 (-2.5,1.5)	0.61
Mental summary (0–100)	1.9±10.6	4.0±10.5	2.1 (-0.4,4.6)	0.16	1.9±10.6	4.0±10.5	2.1 (-0.6,4.9)	0.13
Hospital Anxiety and Depression Scale								
No. evaluated	117	115			117	115		
Total score	-1.5±5.5	- 2.6±5.7	-1.1 (-2.4,0.2)	0.08	-1.5±5.5	-2.6±5.7	-1.1 (-2.6,0.3)	0.13
International Index of Erectile Function								
No. evaluated	109	110			109	110		
Total score (0–43)	-0.2±14.7	-0.5±12.7	-0.6 (-2.6,4.0)	0.94	-0.2±14.7	-0.5±12.7	-0.7 (-3.0,4.3)	0.72
Male Sexual Health Ques.								
No. evaluated	111	107			111	107		
Total score (0–40)	0.6±6.8	1.7±4.5	1.1 (-0.3,2.5)	0.06	0.6±6.8	1.7±4.5	1.1 (-0.4,2.7)	0.15

Note: P values for the absolute difference between study groups for each of the measured outcomes differ for the Nickel et al article and the DSIC. The Nickel et al manuscript (p. 4) notes that comparisons of primary study outcomes across study groups (alfuzosin vs placebo) were calculated "using the exact conditional test version of the Mantel–Haenszel test to control for clustering by clinical center. The pooled rate difference, (i.e., the between-group difference in response rates across clinical centers), and the 95% confidence interval for this difference were calculated with the use of the "metan" module in SAS software, version 9.0, to implement a Mantel–Haenszel estimator for the difference" (p 4). The 'metan' module could not be identified in SAS, but appears to be a special module for use in STATA (not SAS). Metan is not included in STATA version 10 (used in this DSIC); the DSIC provides an uncorrected P value based on use of the ttest command in STATA.

Summary. Descriptive results provided by this DSIC for the baseline measurements and the primary and secondary outcomes closely agree with published values in Nickel et al (2008).

ATTACHMENT 1

Full Text of Article

ATTACHMENT 2

```
*****CPCRN2-Alfuzosin
**Reproduce tabulations in Nickel et al (2008) Alfuzosin and symptoms of
*chronic prostatitis-chronic pelvic pain syndrome
*use cp01_long.dta


*** TABLE 1.  Baseline characteristics by treatment group
codebook agebase
tab1 sex arm race
egen treatment=group(arm)
recode treatment (1=2) (2=1)
label define treatment 1"alfuzosin" 2"placebo"
label values treatment treatment
sort treatment
codebook agebase if treatment==1
codebook agebase if treatment==2
tab treatment, summarize(agebase)
gen recrace=race
recode recrace (5=1) (3=2) (2=3) (else=4)
label define recrace 1"white" 2"black" 3"asian" 4"other"
label values recrace recrace
tab recrace treatment, col chi2

ttest cpsibase, by(treatment)
ttest painbase, by(treatment)
ttest uribase, by(treatment)
ttest qolbase, by(treatment)
ttest urgv1, by(treatment)
ttest cppainv1, by(treatment)
ttest mcgill_sensory_v2, by(treatment)
ttest mcgill_affective_v2, by(treatment)
ttest mcgill_total_v2, by(treatment)
ttest pcs12_v2, by(treatment)
ttest mcs12_v2, by(treatment)
ttest had15_v2, by(treatment)
ttest iief_total_v2, by(treatment)
ttest mshq_total_v2, by(treatment)


*****  TABLE 2. Response rates for NIH-CPSI and GRA by treatment group
********
tab cpsidiff_12wk treatment
tab cpsibin4pt treatment, col

tab withdraw treatment

tab gralast treatment, col
tab grav4 treatment, col

********  TABLE 3. Changes in scores at 12 weeks by treatment group  ********
ttest cpsidiff_12wk, by(treatment)
ttest paindiff_12wk, by(treatment)
ttest uridiff_12wk, by(treatment)
```

```
ttest qoldiff_12wk, by(treatment)

gen cpsichange=cpsiv4-cpsibase
gen painchange=painv4-painbase
gen urichange=uriv4-uribase
gen qolchange=qolv4-qolbase
ttest cpsichange, by(treatment)
ttest painchange, by(treatment)
ttest urichange, by(treatment)
ttest qolchange, by(treatment)


ttest mcgill_total_12wkcng, by(treatment)
ttest mcgill_sensory_12wkcng, by(treatment)
ttest mcgill_affective_12wkcng, by(treatment)
ttest mcs12_12wkcng, by(treatment)
ttest pcs12_12wkcng, by(treatment)
ttest had15_12wkcng, by(treatment)
ttest iief_total_12wkcng, by(treatment)
gen iiefchange=iief_total_v4-iief_total_v2
ttest iiefchange, by(treatment)
ttest mshq_total_12wkcng, by(treatment)
```