Appendix H

Reliability and Measurement Bias Discussion and Findings

Reliability and Measurement Bias Discussion and Findings

The Effect of Elapsed Time on Reliability

We examined whether the gap in time between the facility and gold rater assessments of nursing home residents had any effect on the reliability of measurement. The average gap between the facility rater assessment and the gold rater assessment was 25 days (SD=27). We examined whether facility and gold raters in agreement on each quality measure differed from those that disagreed in terms of the length of time elapsed between their assessments. We found no significant differences for any of the 21 quality measures. In cases where there were modest trends in terms of differences in elapsed time, the direction of difference was counter-intuitive (i.e., those in agreement tended to have a larger gap between the two assessments). However, the absolute number of disagreements for those quality measures was low, suggesting the disagreement was concentrated in a few facilities in which gold rater assessments were uniformly done nearer in time to the facility assessments. Thus, all assessments of both the facility and the research nurse assessors were included in all reliability and directional bias analyses.

Inter- and Intra-State Variation in Reliability

In Table 1 we present the distribution within each state of the 21 quality measures. There is considerable variability of reliability coefficients (kappas) within and across states. Figures 1a, 1b, and 1c illustrate the variability in graphical form for several of the quality measures.

Table 1
Inter- and Intra-State Variation in Kappas

	Illinois					Tennessee					Ohio				
	N	**Mean**	**Med**	**25th**	**75th**	**N**	**Mean**	**Med**	**25th**	**75th**	**N**	**Mean**	**Med**	**25th**	**75th**
Behavior high & low risk (chsra; beh01)	37	**0.70**	0.78	0.47	1.00	31	**0.62**	0.68	0.46	0.89	15	**0.49**	0.51	0.14	1.00
Behavior high risk (chsra; beh02)	37	**0.75**	0.84	0.44	1.00	31	**0.65**	0.72	0.44	0.88	15	**0.59**	0.71	0.29	1.00
Behavior low risk (chsra; beh03)	37	**0.73**	1.00	0.00	1.00	31	**0.66**	1.00	0.00	1.00	15	**0.79**	1.00	1.00	1.00
Little or no activities (chsra; soc02)	37	**0.19**	0.12	0.00	0.31	31	**0.33**	0.27	0.08	0.60	15	**0.18**	0.11	0.00	0.30
Catheter (chsra; cat02)	37	**0.71**	0.87	0.51	1.00	31	**0.78**	0.84	0.63	1.00	15	**0.47**	0.41	0.15	0.71
Incontinence hi & lo risk (chsra; cnt01)	37	**0.84**	0.81	0.75	1.00	31	**0.84**	0.88	0.78	0.92	15	**0.74**	0.73	0.55	1.00
UTI (chsra; cnt04)	37	**0.37**	0.42	0.00	0.60	31	**0.45**	0.42	0.13	0.79	15	**0.34**	0.29	0.26	0.44
Incontinence high risk (chsra; cnt05)	37	**0.65**	1.00	0.00	1.00	31	**0.69**	1.00	0.00	1.00	15	**0.53**	1.00	0.00	1.00
Incontinence low risk (chsra; cnt06)	37	**0.81**	0.80	0.66	1.00	31	**0.85**	0.89	0.77	1.00	15	**0.65**	0.71	0.37	1.00
Tube feeding (ramsey; nut01)	37	**0.84**	1.00	0.78	1.00	31	**0.92**	1.00	1.00	1.00	15	**0.60**	0.66	0.00	1.00
Low BMI (megaqi; bmi0x)	37	**0.82**	0.87	0.75	1.00	31	**0.87**	0.91	0.77	1.00	15	**0.67**	0.84	0.00	1.00
Infection flare-up (megaqi; inf0x)	37	**0.31**	0.35	0.18	0.46	31	**0.39**	0.43	0.24	0.59	15	**0.36**	0.37	0.28	0.48
Pain poorly managed (megaqi; pai0x)	37	**0.52**	0.54	0.35	0.78	31	**0.64**	0.66	0.53	0.77	15	**0.34**	0.36	0.05	0.55
PU high & low risk (chsra; pru01)	37	**0.54**	0.52	0.35	0.78	31	**0.60**	0.63	0.44	0.79	15	**0.40**	0.45	0.20	0.55
PU high risk (chsra; pru02)	37	**0.58**	0.69	0.35	1.00	31	**0.59**	0.60	0.41	0.78	15	**0.48**	0.53	0.00	1.00
PU low risk (chsra; pru03)	37	**0.77**	1.00	0.53	1.00	31	**0.85**	1.00	0.72	1.00	15	**0.84**	1.00	0.77	1.00
Burns abrasions bruises (megaqi; bur0x)	37	**0.72**	1.00	0.45	1.00	31	**0.50**	0.51	-0.05	1.00	15	**0.64**	1.00	0.03	1.00
Restraints (chsra; res01)	37	**0.51**	0.59	0.00	1.00	31	**0.57**	0.65	0.35	1.00	15	**0.63**	0.78	0.14	1.00
Antipsych hi & low risk (chsra; drg01)	37	**0.79**	0.87	0.72	1.00	31	**0.81**	0.81	0.71	0.92	15	**0.80**	0.78	0.65	0.90
Antipsychotic high risk (chsra; drg02)	37	**0.58**	1.00	0.00	1.00	31	**0.61**	1.00	0.00	1.00	15	**0.61**	1.00	0.00	1.00
Antipsychotic low risk (chsra; drg03)	37	**0.82**	1.00	0.70	1.00	31	**0.79**	0.83	0.63	1.00	15	**0.78**	0.87	0.64	1.00

Table 1
Inter- and Intra-State Variation in Kappas (continued)

	Missouri					Pennsylvania					California				
	N	**Mean**	**Med**	**25th**	**75th**	**N**	**Mean**	**Med**	**25th**	**75th**	**N**	**Mean**	**Med**	**25th**	**75th**
Behavior high & low risk (chsra; beh01)	19	**0.50**	0.53	0.42	0.60	30	**0.60**	0.62	0.52	0.74	18	**0.58**	0.62	0.44	0.78
Behavior high risk (chsra; beh02)	19	**0.53**	0.58	0.38	0.72	30	**0.57**	0.58	0.45	0.75	18	**0.63**	0.65	0.46	1.00
Behavior low risk (chsra; beh03)	19	**0.63**	1.00	0.00	1.00	30	**0.86**	1.00	1.00	1.00	18	**0.83**	1.00	1.00	1.00
Little or no activities (chsra; soc02)	19	**0.08**	0.08	0.00	0.10	30	**0.23**	0.15	0.00	0.36	18	**0.15**	0.02	0.00	0.36
Catheter (chsra; cat02)	19	**0.84**	1.00	0.68	1.00	30	**0.56**	0.54	0.33	0.91	18	**0.63**	0.67	0.44	1.00
Incontinence hi & lo risk (chsra; cnt01)	19	**0.74**	0.75	0.64	0.90	30	**0.73**	0.74	0.58	0.88	18	**0.84**	0.90	0.74	1.00
UTI (chsra; cnt04)	19	**0.44**	0.46	0.21	0.76	30	**0.46**	0.51	0.06	0.71	18	**0.47**	0.51	0.34	0.65
Incontinence high risk (chsra; cnt05)	19	**0.82**	1.00	1.00	1.00	30	**0.85**	1.00	1.00	1.00	18	**0.94**	1.00	1.00	1.00
Incontinence low risk (chsra; cnt06)	19	**0.70**	0.71	0.54	0.85	30	**0.68**	0.69	0.55	0.89	18	**0.78**	0.86	0.61	1.00
Tube feeding (ramsey; nut01)	19	**0.91**	1.00	0.84	1.00	30	**0.81**	1.00	0.80	1.00	18	**0.74**	0.95	0.65	1.00
Low BMI (megaqi; bmi0x)	19	**0.85**	1.00	0.66	1.00	30	**0.88**	1.00	0.88	1.00	18	**0.83**	1.00	0.74	1.00
Infection flare-up (megaqi; inf0x)	19	**0.35**	0.38	0.22	0.50	30	**0.45**	0.45	0.35	0.59	18	**0.37**	0.40	0.11	0.61
Pain poorly managed (megaqi; pai0x)	19	**0.28**	0.29	0.00	0.51	30	**0.48**	0.53	0.26	0.71	18	**0.62**	0.64	0.41	0.95
PU high & low risk (chsra; pru01)	19	**0.42**	0.38	0.11	0.65	30	**0.53**	0.51	0.36	0.75	18	**0.67**	0.65	0.56	0.84
PU high risk (chsra; pru02)	19	**0.50**	0.40	0.15	1.00	30	**0.57**	0.52	0.41	0.85	18	**0.74**	0.79	0.57	1.00
PU low risk (chsra; pru03)	19	**0.76**	1.00	0.31	1.00	30	**0.85**	1.00	0.68	1.00	18	**0.78**	1.00	0.47	1.00
Burns abrasions bruises (megaqi; bur0x)	19	**0.43**	0.35	0.09	0.65	30	**0.54**	0.68	-0.04	1.00	18	**0.49**	0.41	0.16	1.00
Restraints (chsra; res01)	19	**0.47**	0.47	0.15	0.66	30	**0.30**	0.25	0.00	0.46	18	**0.68**	0.81	0.55	0.93
Antipsych hi & low risk (chsra; drg01)	19	**0.71**	0.70	0.62	0.91	30	**0.82**	0.85	0.71	1.00	18	**0.71**	0.74	0.66	0.88
Antipsychotic high risk (chsra; drg02)	19	**0.77**	1.00	0.50	1.00	30	**0.75**	1.00	0.48	1.00	18	**0.73**	1.00	0.38	1.00
Antipsychotic low risk (chsra; drg03)	19	**0.67**	0.76	0.47	1.00	30	**0.77**	0.85	0.65	1.00	18	**0.70**	0.76	0.46	1.00

Figure 1a
Distribution of KAPPAs for PRESSURE ULCER #1

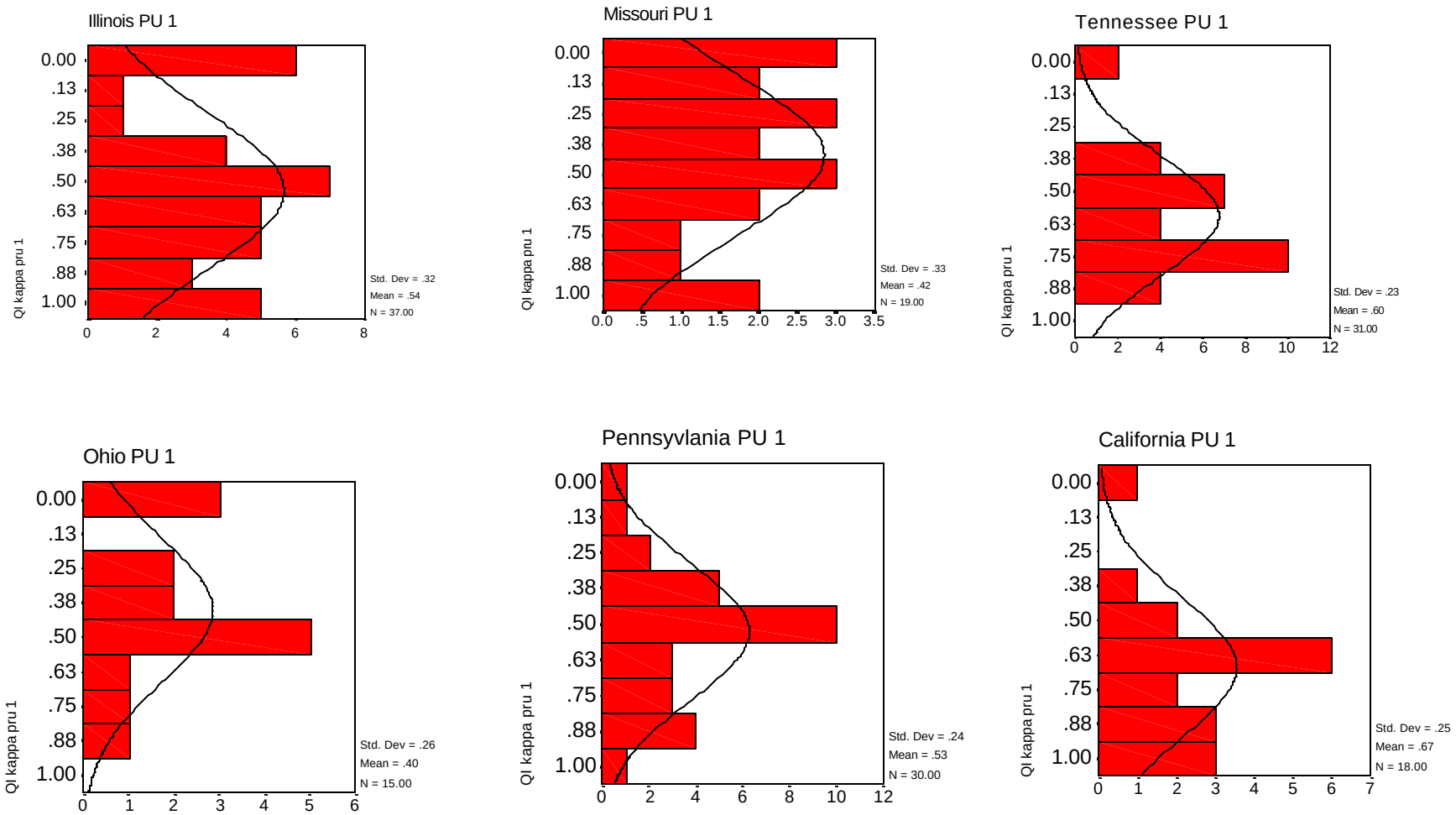


Figure 1b
Distribution of KAPPAs for DRUG #1

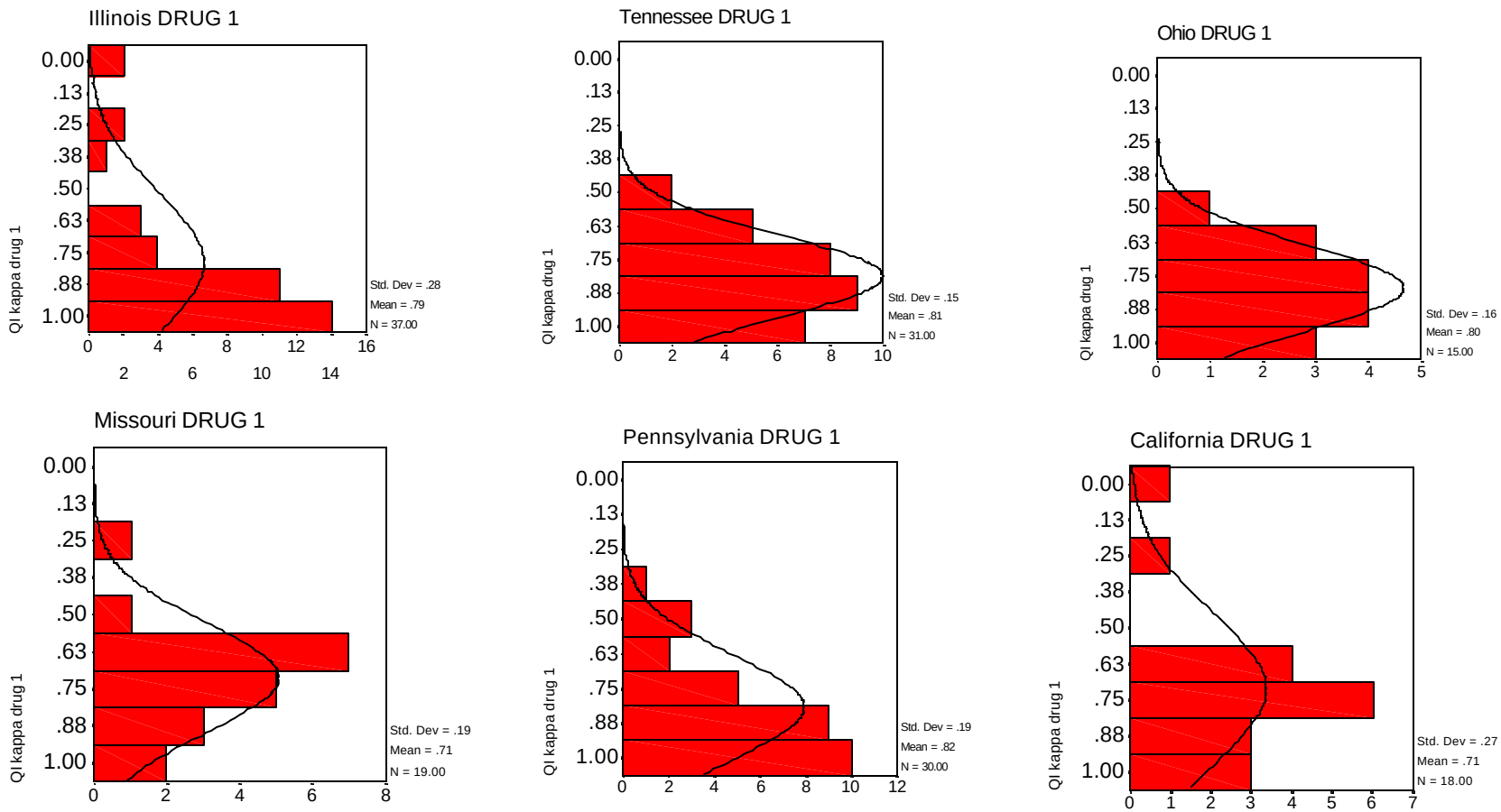
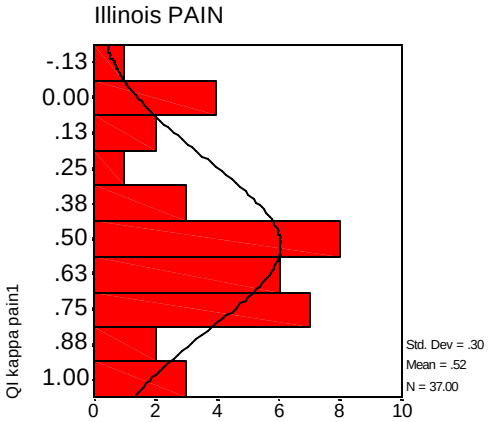
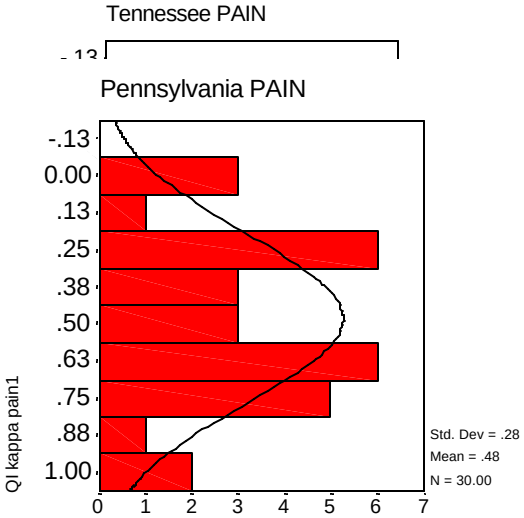
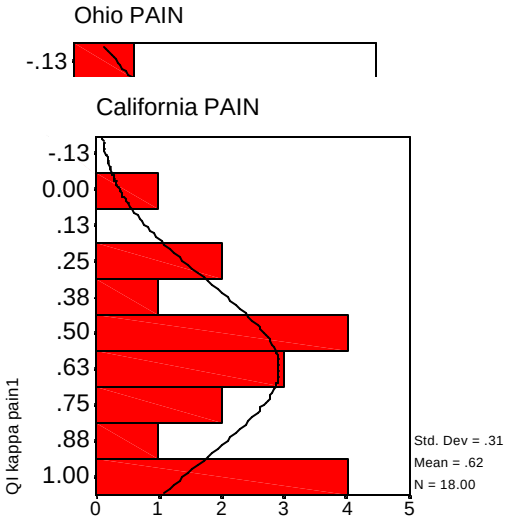
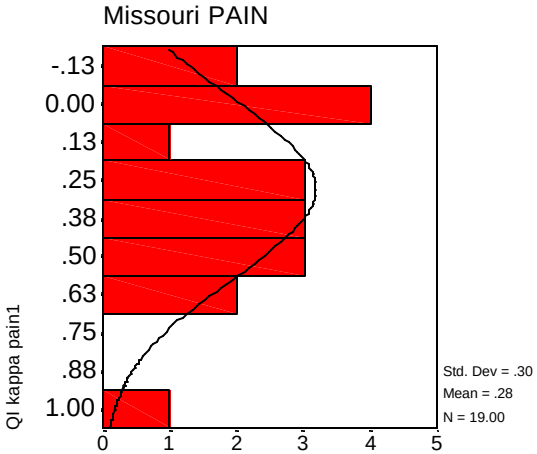


Figure 1c
Distribution of KAPPAs for PAIN









Case Summaries

As is apparent, the level of detailed information that is relevant in understanding how and why some QIs end up with a poor reliability coefficient and others have an acceptable measure is considerable. We've seen that there is inter-state variation and within state variation on selected QIs. Since the Kappa statistic is very sensitive to the prevalence of the condition under study (if the event is rare, any disagreement is weighted heavily because the expected is for the condition to be absent), and that may be relevant to both the facility as well as the number of dual assessments done in the facility, we created a number of "case summaries" to provide the reader with the complete detailed information on all facilities for a selected QI. This entailed examining the kappa of the underlying items used to calculate the QI as well as the resulting QI itself.

The table below lists a detailed summary of the pain-related items analyzed as part of the reliability analysis. Each row represents an individual facility. The average agreement between the gold and facility-raters are presented for each item of the pain score (pain frequency and pain intensity). The kappas for each item are then presented, with (.) indicating insufficient information present to compute a kappa for the item. The actual prevalence of pain is then presented (as measured by the gold and facility-raters), followed by the average facility-level percent agreement on the pain QI. The kappa for the pain QI is shown in the final column.

Table 2
Detailed Case Summaries

Facility Number	Facility level agree J2a Pain freq.	Facility level agree J2b Pain Intensity	Kappa J2a pain freq.	Kappa J2b pain intensity	Facility-scored Pain QI	Gold-standard Pain QI	Average facility-level % agree on Pain QI	Kappa Pain QI
1	0.88	0.85	0.69	0.63	0.08	0.08	1.00	1.00
2	0.90	0.90	.	.	0.03	0.03	0.93	-0.04
3	0.89	0.89	.	0.64	0.04	0.04	1.00	1.00
4	0.97	0.87	0.95	.	0.50	0.47	0.97	0.93
5	0.75	0.75	0.55	.	0.21	0.08	0.88	0.51
6	0.61	0.54	0.41	0.33	0.48	0.14	0.67	0.32
7	0.74	0.85	0.35	.	0.07	0.19	0.81	0.20
8	0.79	0.83	0.37	.	0.13	0.08	0.96	0.78
9	0.79	0.62	0.68	.	0.24	0.28	0.83	0.55
10	0.57	0.50	0.30	0.25	0.21	0.07	0.86	0.44
11	0.85	0.73	0.79	.	0.28	0.15	0.88	0.66
12	0.68	0.61	0.35	0.20	0.32	0.07	0.75	0.28
13	0.83	0.70	0.73	0.55	0.53	0.37	0.83	0.67
14	0.71	0.67	0.53	0.48	0.13	0.13	1.00	1.00
15	0.70	0.63	0.43	0.30	0.07	0.19	0.89	0.52
16	0.68	0.64	0.34	.	0.16	0.08	0.92	0.63
17	0.89	0.89	.	.	0.11	0.07	0.96	0.78
18	0.95	0.90	0.64	.	0.00	0.00	1.00	1.00
19	0.83	0.97	0.67	0.93	0.23	0.23	0.87	0.63
20	0.80	0.83	0.54	0.63	0.03	0.00	0.97	0.00
21	0.68	0.68	.	0.34	0.00	0.11	0.89	0.00

Table 2
Detailed Case Summaries

Facility Number	Facility level agree J2a Pain freq.	Facility level agree J2b Pain Intensity	Kappa J2a pain freq.	Kappa J2b pain intensity	Facility-scored Pain QI	Gold-standard Pain QI	Average facility-level % agree on Pain QI	Kappa Pain QI
22	0.77	0.73	0.59	0.54	0.23	0.07	0.77	0.13
23	0.90	0.83	0.83	.	0.07	0.13	0.93	0.63
24	0.62	0.59	0.53	.	0.12	0.17	0.92	0.71
25	0.71	0.71	0.51	0.48	0.07	0.11	0.96	0.78
26	0.67	0.70	0.50	0.53	0.60	0.30	0.70	0.44
27	1.00	0.97	1.00	0.93	0.13	0.10	0.97	0.84
28	0.70	0.73	0.40	.	0.03	0.10	0.93	0.47
29	0.90	0.83	0.77	.	0.10	0.07	0.97	0.78
30	0.86	0.86	0.64	0.73	0.61	0.50	0.89	0.79
31	0.80	0.70	0.55	.	0.10	0.13	0.90	0.52
32	0.73	0.67	0.60	0.55	0.47	0.37	0.83	0.66
33	0.73	0.73	0.55	0.53	0.20	0.27	0.80	0.44
34	0.79	0.86	0.53	0.68	0.03	0.10	0.93	0.47
35	0.83	0.83	0.51	0.52	0.10	0.07	0.97	0.78
36	0.77	0.73	0.43	.	0.14	0.13	0.86	0.42
37	0.90	0.90	-0.04	-0.04	0.04	0.00	0.96	0.00
38	0.73	0.73	0.46	0.45	0.07	0.13	0.93	0.63
39	0.73	0.47	0.60	0.42	0.30	0.33	0.70	0.31
40	0.62	0.59	0.43	0.42	0.24	0.31	0.86	0.66
41	0.46	0.25	0.24	0.01	0.19	0.29	0.65	0.09
42	1.00	0.96	1.00	0.94	0.07	0.07	1.00	1.00
43	0.89	0.82	0.60	.	0.00	0.00	1.00	1.00
44	0.59	0.36	0.46	.	0.33	0.26	0.81	0.54
45	0.83	0.80	.	0.53	0.00	0.03	0.97	0.00
46	0.79	0.69	0.60	.	0.24	0.17	0.93	0.79
47	0.52	0.48	0.33	0.33	0.11	0.17	0.93	0.63
48	0.83	0.76	0.64	0.51	0.07	0.07	1.00	1.00
49	0.83	0.87	0.72	0.79	0.33	0.33	0.93	0.85
50	0.67	0.70	0.29	0.36	0.07	0.07	0.87	-0.07
51	0.60	0.60	0.37	0.42	0.43	0.40	0.77	0.52
52	0.45	0.45	0.13	0.21	0.50	0.62	0.68	0.36
53	0.57	0.63	0.36	0.48	0.30	0.40	0.70	0.35
54	0.90	0.87	0.85	0.77	0.50	0.40	0.90	0.80
55	0.55	0.59	0.30	0.34	0.28	0.21	0.79	0.44
56	0.62	0.62	.	.	0.04	0.08	0.92	0.00
57	0.75	0.64	0.34	.	0.04	0.04	1.00	1.00
58	0.80	0.80	0.51	.	0.04	0.04	0.92	-0.04
59	0.61	0.61	0.40	.	0.11	0.23	0.88	0.61
60	0.43	0.43	.	.	0.04	0.14	0.82	-0.06
61	0.67	0.53	0.43	0.14	0.11	0.07	0.82	-0.09
62	0.73	0.64	0.42	.	0.10	0.07	0.89	0.34
63	0.66	0.59	0.41	.	0.11	0.14	0.89	0.51

Table 2
Detailed Case Summaries

Facility Number	Facility level agree J2a Pain freq.	Facility level agree J2b Pain Intensity	Kappa J2a pain freq.	Kappa J2b pain intensity	Facility-scored Pain QI	Gold-standard Pain QI	Average facility-level % agree on Pain QI	Kappa Pain QI
64	0.53	0.43	0.21	.	0.14	0.13	0.89	0.51
65	0.64	0.57	0.42	.	0.22	0.07	0.85	0.44
66	0.62	0.59	0.37	0.31	0.21	0.14	0.79	0.28
67	0.62	0.62	0.45	0.46	0.10	0.25	0.79	0.29
68	0.69	0.59	0.32	.	0.07	0.14	0.93	0.63
69	0.72	0.72	0.18	.	0.00	0.14	0.86	0.00
70	0.63	0.56	.	.	0.15	0.19	0.81	0.33
71	0.63	0.67	0.26	0.32	0.07	0.17	0.76	-0.11
72	0.63	0.53	0.38	.	0.17	0.13	0.77	0.09
73	0.34	0.28	0.03	-0.04	0.33	0.31	0.67	0.23
74	0.63	0.53	0.47	.	0.21	0.21	0.79	0.36
75	0.54	0.50	0.34	.	0.38	0.11	0.73	0.35
76	0.52	0.45	0.33	0.24	0.55	0.17	0.62	0.29
77	0.39	0.39	0.03	.	0.08	0.00	0.92	0.00
78	0.69	0.65	0.57	0.52	0.04	0.08	0.96	0.65
79	0.65	0.61	0.43	0.32	0.52	0.39	0.52	0.05
80	0.62	0.50	.	0.17	0.73	0.52	0.44	-0.14
81	0.90	0.79	0.85	.	0.21	0.24	0.90	0.70
82	0.61	0.54	0.42	0.31	0.14	0.29	0.79	0.38
83	0.36	0.46	-0.03	0.16	0.11	0.25	0.71	0.06
84	0.64	0.61	0.46	.	0.32	0.21	0.82	0.55
85	0.70	0.60	0.52	0.42	0.17	0.27	0.83	0.51
86	0.59	0.59	0.41	.	0.19	0.16	0.71	0.05
87	0.57	0.50	0.31	0.35	0.28	0.41	0.79	0.52
88	0.77	0.70	0.73	0.65	0.32	0.30	0.93	0.84
89	0.67	0.47	0.49	0.28	0.41	0.33	0.86	0.71
90	0.67	0.63	0.17	0.12	0.07	0.10	0.90	0.35
91	0.71	0.68	0.36	.	0.14	0.29	0.86	0.59
92	0.79	0.82	.	0.23	0.00	0.07	0.93	0.00
93	0.74	0.70	.	.	0.00	0.07	0.92	0.00
94	0.80	0.83	0.64	.	0.14	0.13	0.93	0.71
95	0.76	0.72	0.59	0.53	0.18	0.24	0.93	0.79
96	0.83	0.71	.	.	0.00	0.04	0.96	0.00
97	0.64	0.64	0.36	.	0.15	0.07	0.93	0.63
98	0.70	0.77	0.42	0.56	0.07	0.17	0.90	0.53
99	0.63	0.67	0.40	0.46	0.13	0.10	0.83	0.19
100	0.63	0.63	0.37	.	0.07	0.17	0.90	0.53
101	0.72	0.62	0.50	0.31	0.11	0.07	0.96	0.78
102	0.62	0.52	0.36	0.14	0.28	0.07	0.72	0.10
103	0.77	0.67	0.61	0.43	0.13	0.23	0.77	0.23
104	0.77	0.65	0.54	0.34	0.08	0.15	0.85	0.26
105	0.62	0.54	0.43	.	0.04	0.23	0.83	0.28

Table 2
Detailed Case Summaries

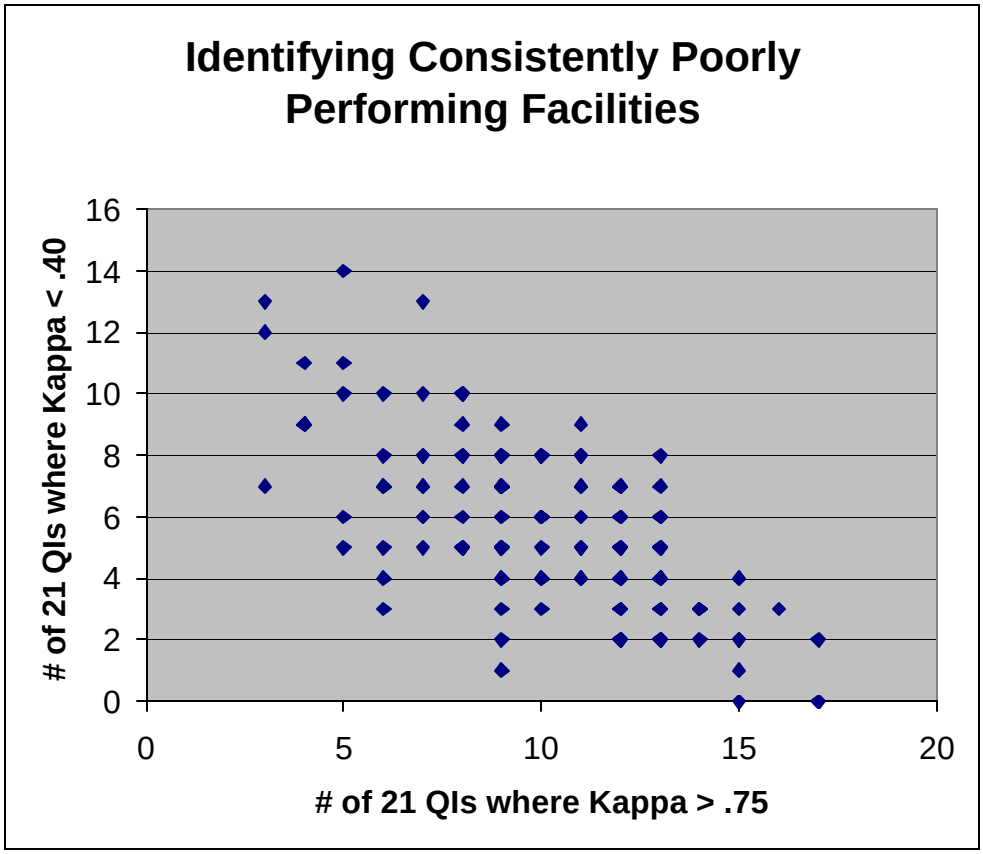
Facility Number	Facility level agree J2a Pain freq.	Facility level agree J2b Pain Intensity	Kappa J2a pain freq.	Kappa J2b pain intensity	Facility-scored Pain QI	Gold-standard Pain QI	Average facility-level % agree on Pain QI	Kappa Pain QI
106	0.52	0.59	0.23	0.23	0.07	0.30	0.78	0.32
107	0.83	0.83	0.59	.	0.10	0.10	0.93	0.63
108	0.70	0.70	0.58	0.55	0.24	0.47	0.76	0.51
109	0.67	0.67	0.45	.	0.13	0.27	0.80	0.39
110	0.79	0.69	0.69	.	0.11	0.17	0.93	0.71
111	0.90	0.87	0.81	0.80	0.07	0.03	0.97	0.65
112	0.80	0.80	0.61	.	0.27	0.20	0.93	0.81
113	0.57	0.73	0.25	.	0.77	0.43	0.60	0.25
114	0.73	0.73	0.43	0.43	0.07	0.13	0.93	0.63
115	0.62	0.59	0.27	.	0.07	0.14	0.86	0.27
116	0.79	0.79	0.19	.	0.03	0.03	1.00	1.00
117	0.90	0.80	0.70	.	0.03	0.03	1.00	1.00
118	0.87	0.80	0.77	0.68	0.13	0.17	0.90	0.61
119	0.73	0.73	0.59	0.60	0.27	0.30	0.90	0.75
120	0.82	0.79	0.72	0.70	0.32	0.36	0.89	0.76
121	0.67	0.77	0.48	0.60	0.27	0.30	0.77	0.43
122	0.70	0.70	0.54	0.54	0.30	0.27	0.83	0.59
123	0.61	0.52	0.37	.	0.20	0.20	0.80	0.38
124	0.80	0.73	0.31	.	0.07	0.07	0.93	0.46
125	0.67	0.67	.	.	0.10	0.20	0.90	0.62
126	0.60	0.67	0.31	0.39	0.23	0.10	0.73	0.07
127	0.73	0.67	0.57	.	0.20	0.13	0.93	0.76
128	0.57	0.53	0.20	.	0.07	0.17	0.90	0.53
129	0.67	0.53	0.38	.	0.13	0.23	0.83	0.45
130	0.69	0.76	0.54	.	0.54	0.38	0.86	0.72
131	0.66	0.62	0.36	0.27	0.10	0.14	0.90	0.51
132	0.86	0.86	0.72	0.05	0.69	0.59	0.83	0.63
133	0.90	0.87	0.85	0.79	0.40	0.40	0.93	0.86
134	0.97	0.90	0.94	0.83	0.03	0.07	0.97	0.65
135	0.80	0.80	0.69	0.69	0.33	0.43	0.90	0.79
136	0.83	0.90	0.72	0.83	0.23	0.17	0.87	0.59
137	0.67	0.73	0.37	.	0.10	0.23	0.80	0.30
138	0.68	0.57	0.50	0.39	0.36	0.50	0.79	0.57
139	0.80	0.73	0.68	0.57	0.13	0.13	0.93	0.71
140	0.71	0.75	0.61	0.67	0.19	0.21	0.89	0.66
141	0.67	0.67	0.49	0.51	0.30	0.23	0.87	0.66
142	0.83	0.87	0.74	.	0.33	0.30	0.90	0.77
143	0.75	0.86	0.64	0.78	0.39	0.25	0.86	0.68
144	0.77	0.77	0.60	.	0.20	0.20	1.00	1.00
145	0.83	0.79	0.74	0.70	0.21	0.28	0.93	0.81
146	0.79	0.76	0.52	.	0.14	0.10	0.97	0.84
147	0.71	0.71	0.48	0.48	0.07	0.11	0.96	0.78

Table 2
Detailed Case Summaries

Facility Number	Facility level agree J2a Pain freq.	Facility level agree J2b Pain Intensity	Kappa J2a pain freq.	Kappa J2b pain intensity	Facility-scored Pain QI	Gold-standard Pain QI	Average facility-level % agree on Pain QI	Kappa Pain QI
148	0.73	0.80	0.57	0.74	0.21	0.20	0.86	0.58
149	0.83	0.83	0.74	0.74	0.21	0.21	0.93	0.79
150	0.77	0.60	0.63	0.39	0.17	0.27	0.77	0.32

Careful scrutiny of the data suggests that by and large the percent agreement for most facilities on the actual pain QI is quite high, although the kappa statistic varies dramatically, largely due to the prevalence of the pain QI. When the prevalence is reasonably high, the kappa tends to be better. Furthermore, there is a fairly close correspondence between the weighted kappa statistic for the MDS item on the frequency of pain and the kappa for the overall QI.

Figure 2
Scatterplot of Facility Counts of Kappa Performance



Facilities were classified in terms of the absolute count of the 21 QIs for which they had a kappa exceeding .75 versus the absolute number of QIs for which the kappa fell below .40. This basically contrasts the likelihood that a facility has QIs with unacceptable kappas and the likelihood that they have exceptionally good reliability on some QIs. These two values were then plotted against each other to visually identify facilities that had relatively few HIGH kappa values while having an exceptionally large number of LOW kappas. As can be seen, there are 5 facilities with greater than 10 QIs with kappas under .4 and 5 or fewer QIs on which they had excellent kappa levels, exceeding .75. There was an additional facility with 13 QIs with poor kappa and only 7 QIs with excellent reliability. Overall, however, the correlation between these two measures was good – the more excellent QI kappas the fewer the poor kappas with the average facility having nearly 10 kappas in the excellent range and around 6 in the "poor" range.

Assessing Ascertainment or the Direction of Measurement Bias

A gamma statistic was computed to characterize the direction of agreement between the gold and facility-raters in each facility. High positive values are indicative of the facility-raters in the nursing home reporting more instances of a given problem, whereas large negative values are indicative of facility-raters underreporting problems relative to the gold-raters.

Five of the six facilities in Figure 2 with at least 11 or more QIs with kappas below .40 had at least five extreme gammas (>.6 or <-.6), and in nearly every case negative bias was more prevalent.

Figure 3 is a histogram of the gamma statistic for all facilities across all the QIs. Gamma was classified into one of five categories, ranging from large negative bias to large positive bias. For the majority of cases, gamma was estimated between –0.2 and 0.2 -- indicating little to no measurement bias. However, there were a substantial number of instances with moderate to large measurement bias; there were more cases of negative bias than positive bias, meaning facilities were more likely to under-report an indicator than over-report it.

Figure 3

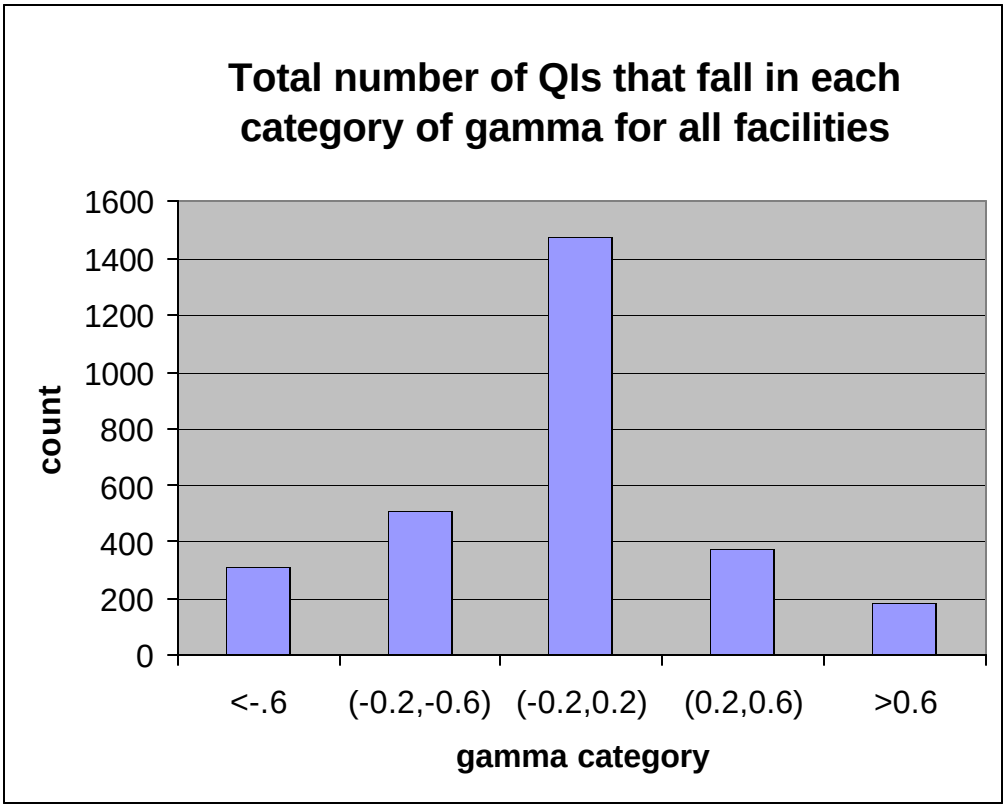


Table 3 displays the distribution of extreme values of gamma by facility. The rows are counts of the number of QIs in a facility that had a gamma greater than 0.6 (large positive bias). The columns are counts of the number of QIs in each facility that had gamma less than –0.6 (large

negative bias). From the margins it is clear that facilities were more likely to under-report indicators. For example, 52 facilities did not have a gamma greater than 0.6 for any of the indicators, whereas only 28 did not have any gamma less than –0.6. On the other hand, 38 of 150 facilities had none or only one QI that had an extreme Gamma score and an additional 36 facilities had two or fewer extreme indicators, reflecting a common perspective between the two assessors on the vast majority of QIs in the plurality of facilities.

Table 3
Extreme Values of Gamma

		Number of 21 QIs for which facility has gamma less than -0.6								
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Total**
Number of 21 QIs for which facility has gamma greater than 0.6	**0**	7	12	9	11	4	7	2	0	52
	1	7	12	11	10	6	3	0	0	49
	2	6	8	2	5	1	0	1	1	24
	3	5	5	1	2	3	1	0	0	17
	4	2	0	0	1	0	0	0	0	3
	5	0	2	0	0	0	0	0	0	2
	6	1	1	0	1	0	0	0	0	3
Total		28	40	23	30	14	11	3	1	150

Table 4 reports the median and inter-quartile range (IQR) for each indicator, stratified by state. The majority of the medians are zero, indicating that the 'typical' facility in each state does not tend to systematically under- or over- report the QIs. There are some exceptions, though. For example, facilities in every state severely under-report social activities; facilities in CA tend to over-report pain whereas facilities in PA tend to under-report pain.

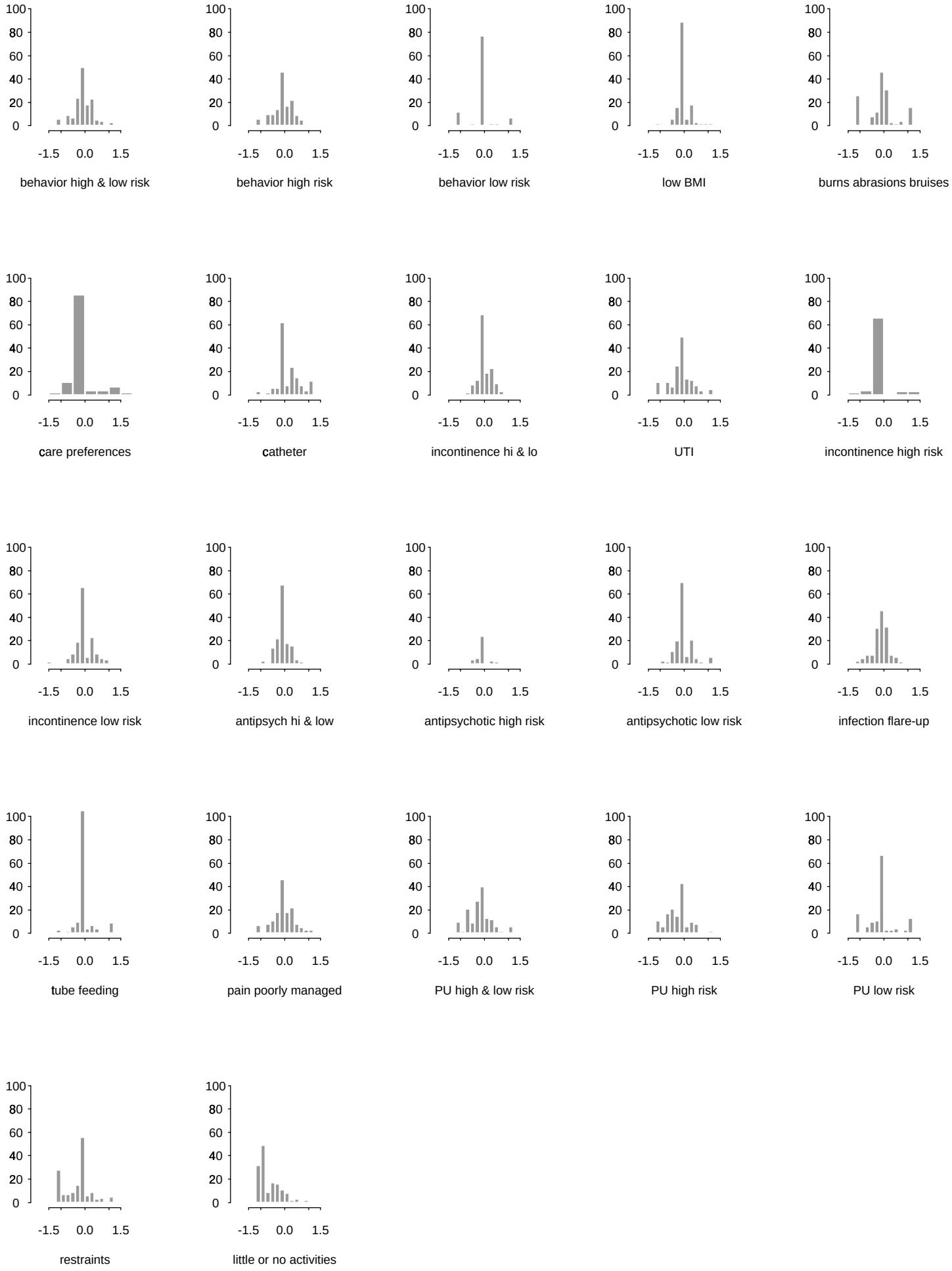
Table 4
Median and Inter-quartile Range (IQR) for Each Indicator

	Gamma statistics by state **Median (IQR)**					
	CA	IL	MO	OH	PA	TN
behavior hi & lo	0.37 (0.59)	0.00 (0.39)	-0.11 (0.32)	0.00 (0.42)	-0.08 (0.56)	0.00 (0.41)
behavior high	0.37 (0.53)	0.00 (0.38)	-0.04 (0.38)	0.00 (0.53)	0.00 (0.59)	-0.05 (0.51)
behavior low	0.00 (0.00)	0.00 (0.00)	0.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
low BMI	0.00 (0.00)	0.00 (0.23)	0.00 (0.00)	0.00 (0.37)	0.00 (0.00)	0.00 (0.19)
burns abrasions	-0.02 (0.27)	0.00 (0.69)	-0.24 (0.56)	-0.16 (1.00)	0.00 (0.28)	0.00 (1.13)
catheter	0.00 (0.63)	0.00 (0.24)	0.00 (0.00)	0.32 (0.78)	0.22 (0.70)	0.00 (0.36)
incont high & low	0.00 (0.32)	0.00 (0.18)	-0.13 (0.27)	0.00 (0.34)	0.14 (0.39)	0.00 (0.23)
UTI	0.15 (0.45)	0.00 (0.49)	-0.16 (0.30)	-0.23 (0.33)	-0.06 (0.63)	-0.17 (0.44)
incontinence hi	0.00 (0.00)	0.00 (0.00)	0.00 (0.000	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
incontinence low	0.00 (0.31)	0.00 (0.17)	-0.22 (0.43)	0.00 (0.33)	0.00 (0.69)	0.00 (0.20)
antipsych hi & lo	0.00 (0.52)	0.00 (0.20)	0.00 (0.51)	0.00 (0.23)	0.00 (0.18)	-0.04 (0.27)
antipsych high	-0.10 (0.83)	0.00 (0.00)	0.00 (0.00)	0.02 (0.79)	0.00 (0.50)	0.00 (0.00)
antipsych low	0.00 (0.29)	0.00 (0.00)	0.00 (0.50)	0.00 (0.55)	0.00 (0.26)	0.00 (0.22)

Table 4
Median and Inter-quartile Range (IQR) for Each Indicator

	Gamma statistics by state **Median (IQR)**					
	CA	IL	MO	OH	PA	TN
infection flare-up	0.00 (0.22)	-0.02 (0.31)	-0.18 (0.17)	-0.10 (0.36)	-0.03 (0.42)	-0.18 (0.44)
tube feeding	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.69)	0.00 (0.26)	0.00 (0.00)
pain poorly man.	0.24 (0.54)	0.00 (0.45)	0.00 (0.50)	0.00 (0.65)	-0.20 (0.71)	0.00 (0.45)
PU high & low	0.00 (0.60)	-0.08 (0.40)	-0.24 (0.65)	-0.62 (0.67)	-0.08 (0.44)	-0.33 (0.42)
PU high risk	0.00 (0.42)	0.00 (0.83)	-0.48 (0.73)	-0.59 (1.00)	-0.02 (0.30)	-0.41 (0.68)
PU low risk	0.00 (0.87)	0.00 (0.36)	0.00 (1.00)	0.00 (0.67)	0.00 (1.36)	0.00 (0.55)
restraints	0.00 (0.32)	0.00 (1.00)	0.00 (0.69)	0.00 (0.52)	-0.35 (0.86)	-0.17 (0.46)
little, no activities	-0.92 (0.74)	-0.83 (0.55)	-0.95 (0.53)	-0.84 (0.67)	-0.47 (0.94)	-0.70 (0.54)

Figure 4 presents histograms of gamma for each indicator. In the majority of cases the distribution is centered around 0 with relatively symmetric tails. There are exceptions. For example, the pressure ulcer and restraint QIs tend to have a number of facilities with a large, negative estimated gamma (large negative bias).


behavior high & low risk
behavior high risk
behavior low risk
low BMI
burns abrasions bruises
care preferences
catheter
incontinence hi & lo
UTI
incontinence high risk
incontinence low risk
antipsych hi & low
antipsychotic high risk
antipsychotic low risk
infection flare-up
tube feeding
pain poorly managed
PU high & low risk
PU high risk
PU low risk
restraints
little or no activities
100
80
60
40
20
0
-1.5
0.0
1.5