

Uses models:	Yes
Uses documents:	No
Emulates:	Responsiveness

	Initial summaries				Update summaries			
	Pyramid		Responsiveness		Pyramid		Responsiveness	
	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)
Pearson	0.883	0.000 (0.804 - 0.932)	0.914	0.000 (0.853 - 0.950)	0.960	0.000 (0.930 - 0.977)	0.955	0.000 (0.923 - 0.974)
Spearman	0.948	0.000	0.951	0.000	0.930	0.000	0.917	0.000
Kendall	0.823	0.000	0.836	0.000	0.780	0.000	0.764	0.000

Table 1: Correlations for CLASSY1, initial and update summaries, all peers.

	Initial summaries				Update summaries			
	Pyramid		Responsiveness		Pyramid		Responsiveness	
	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)
Pearson	0.976	0.000 (0.955 - 0.987)	0.979	0.000 (0.960 - 0.988)	0.964	0.000 (0.934 - 0.980)	0.959	0.000 (0.925 - 0.978)
Spearman	0.936	0.000	0.930	0.000	0.921	0.000	0.909	0.000
Kendall	0.812	0.000	0.804	0.000	0.771	0.000	0.776	0.000

Table 2: Correlations for CLASSY1, initial and update summaries, no models.

		A>B	A=B	A<B
CLASSY1	A>B	194	0	
	A=B	150	0	
	A>B			0

Table 3: Discriminative power for CLASSY1 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY1	A>B	302	0	
	A=B	42	0	
	A>B			0

Table 4: Discriminative power for CLASSY1 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY1	A>B	194	0	
	A=B	150	0	
	A>B			0

Table 5: Discriminative power for CLASSY1 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY1	A>B	302	0	
	A=B	42	0	
	A>B			0

Table 6: Discriminative power for CLASSY1 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY1	A>B	284	39	
	A=B	34	546	
	A>B			0

Table 7: Discriminative power for CLASSY1 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY1	A>B	194	71	
	A=B	33	605	
	A>B			0

Table 8: Discriminative power for CLASSY1 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY1	A>B	269	54	
	A=B	25	555	
	A>B			0

Table 9: Discriminative power for CLASSY1 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY1	A>B	186	79	
	A=B	22	616	
	A>B			0

Table 10: Discriminative power for CLASSY1 compared with Responsiveness. Update summaries, non-model summaries.

Uses models:	Yes
Uses documents:	No
Emulates:	Pyramid and Responsiveness

	Initial summaries				Update summaries			
	Pyramid		Responsiveness		Pyramid		Responsiveness	
	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)
Pearson	0.901	0.000 (0.832 - 0.943)	0.918	0.000 (0.860 - 0.952)	0.946	0.000 (0.906 - 0.969)	0.941	0.000 (0.899 - 0.966)
Spearman	0.935	0.000	0.941	0.000	0.881	0.000	0.864	0.000
Kendall	0.796	0.000	0.820	0.000	0.702	0.000	0.680	0.000

Table 1: Correlations for CLASSY2, initial and update summaries, all peers.

	Initial summaries				Update summaries			
	Pyramid		Responsiveness		Pyramid		Responsiveness	
	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)
Pearson	0.972	0.000 (0.948 - 0.985)	0.975	0.000 (0.955 - 0.987)	0.958	0.000 (0.923 - 0.977)	0.956	0.000 (0.920 - 0.976)
Spearman	0.924	0.000	0.906	0.000	0.911	0.000	0.898	0.000
Kendall	0.792	0.000	0.765	0.000	0.766	0.000	0.774	0.000

Table 2: Correlations for CLASSY2, initial and update summaries, no models.

		A>B	A=B	A<B
CLASSY2	A>B	231	0	
	A=B	113	0	
	A>B			0

		A>B	A=B	A<B
CLASSY2	A>B	304	0	
	A=B	40	0	
	A>B			0

Table 3: Discriminative power for CLASSY2 compared with Pyramid. Initial summaries, models vs. non-models.

Table 4: Discriminative power for CLASSY2 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY2	A>B	231	0	
	A=B	113	0	
	A>B			0

		A>B	A=B	A<B
CLASSY2	A>B	304	0	
	A=B	40	0	
	A>B			0

Table 5: Discriminative power for CLASSY2 compared with Responsiveness. Initial summaries, models vs. non-models.

Table 6: Discriminative power for CLASSY2 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY2	A>B	295	83	
	A=B	23	502	
	A>B			0

		A>B	A=B	A<B
CLASSY2	A>B	194	71	
	A=B	33	605	
	A>B			0

Table 7: Discriminative power for CLASSY2 compared with Pyramid. Initial summaries, non-model summaries.

Table 8: Discriminative power for CLASSY2 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY2	A>B	278	100	
	A=B	16	509	
	A>B			0

		A>B	A=B	A<B
CLASSY2	A>B	184	81	
	A=B	24	614	
	A>B			0

Table 9: Discriminative power for CLASSY2 compared with Responsiveness. Initial summaries, non-model summaries.

Table 10: Discriminative power for CLASSY2 compared with Responsiveness. Update summaries, non-model summaries.

Uses models:	Yes
Uses documents:	No
Emulates:	Responsiveness

	Initial summaries				Update summaries			
	Pyramid		Responsiveness		Pyramid		Responsiveness	
	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)
Pearson	0.898	0.000 (0.828 - 0.941)	0.888	0.000 (0.811 - 0.935)	0.798	0.000 (0.670 - 0.880)	0.790	0.000 (0.657 - 0.875)
Spearman	0.854	0.000	0.850	0.000	0.721	0.000	0.692	0.000
Kendall	0.666	0.000	0.685	0.000	0.596	0.000	0.562	0.000

Table 1: Correlations for CLASSY3, initial and update summaries, all peers.

	Initial summaries				Update summaries			
	Pyramid		Responsiveness		Pyramid		Responsiveness	
	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)
Pearson	0.961	0.000 (0.929 - 0.979)	0.961	0.000 (0.928 - 0.979)	0.889	0.000 (0.803 - 0.939)	0.899	0.000 (0.820 - 0.944)
Spearman	0.875	0.000	0.881	0.000	0.829	0.000	0.823	0.000
Kendall	0.716	0.000	0.738	0.000	0.653	0.000	0.665	0.000

Table 2: Correlations for CLASSY3, initial and update summaries, no models.

		A>B	A=B	A<B
CLASSY3	A>B	247	0	
	A=B	97	0	
	A>B			0

Table 3: Discriminative power for CLASSY3 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY3	A>B	303	0	
	A=B	41	0	
	A>B			0

Table 4: Discriminative power for CLASSY3 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY3	A>B	247	0	
	A=B	97	0	
	A>B			0

Table 5: Discriminative power for CLASSY3 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY3	A>B	303	0	
	A=B	41	0	
	A>B			0

Table 6: Discriminative power for CLASSY3 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY3	A>B	290	50	
	A=B	28	535	
	A>B			0

Table 7: Discriminative power for CLASSY3 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY3	A>B	199	171	
	A=B	28	505	
	A>B			0

Table 8: Discriminative power for CLASSY3 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY3	A>B	276	64	
	A=B	18	545	
	A>B			0

Table 9: Discriminative power for CLASSY3 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY3	A>B	185	185	
	A=B	23	510	
	A>B			0

Table 10: Discriminative power for CLASSY3 compared with Responsiveness. Update summaries, non-model summaries.

Uses models:	Yes
Uses documents:	No
Emulates:	Pyramid

	Initial summaries				Update summaries			
	Pyramid		Responsiveness		Pyramid		Responsiveness	
	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)
Pearson	0.890	0.000 (0.813 - 0.936)	0.916	0.000 (0.857 - 0.951)	0.952	0.000 (0.916 - 0.972)	0.945	0.000 (0.904 - 0.968)
Spearman	0.946	0.000	0.949	0.000	0.888	0.000	0.856	0.000
Kendall	0.828	0.000	0.833	0.000	0.729	0.000	0.679	0.000

Table 1: Correlations for CLASSY4, initial and update summaries, all peers.

	Initial summaries				Update summaries			
	Pyramid		Responsiveness		Pyramid		Responsiveness	
	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)	value	p (95% CI)
Pearson	0.961	0.000 (0.929 - 0.979)	0.954	0.000 (0.915 - 0.975)	0.959	0.000 (0.925 - 0.978)	0.960	0.000 (0.927 - 0.978)
Spearman	0.874	0.000	0.875	0.000	0.905	0.000	0.910	0.000
Kendall	0.720	0.000	0.720	0.000	0.744	0.000	0.787	0.000

Table 2: Correlations for CLASSY4, initial and update summaries, no models.

		A>B	A=B	A<B
CLASSY4	A>B	204	0	
	A=B	140	0	
	A>B			0

Table 3: Discriminative power for CLASSY4 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY4	A>B	312	0	
	A=B	32	0	
	A>B			0

Table 4: Discriminative power for CLASSY4 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY4	A>B	204	0	
	A=B	140	0	
	A>B			0

Table 5: Discriminative power for CLASSY4 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY4	A>B	312	0	
	A=B	32	0	
	A>B			0

Table 6: Discriminative power for CLASSY4 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY4	A>B	246	33	
	A=B	72	552	
	A>B			0

Table 7: Discriminative power for CLASSY4 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY4	A>B	205	103	
	A=B	22	573	
	A>B			0

Table 8: Discriminative power for CLASSY4 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY4	A>B	234	45	
	A=B	60	564	
	A>B			0

Table 9: Discriminative power for CLASSY4 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY4	A>B	192	116	
	A=B	16	579	
	A>B			0

Table 10: Discriminative power for CLASSY4 compared with Responsiveness. Update summaries, non-model summaries.