Equivalence testing using open symmetric intervals – an evaluation

Michael Meyners

11th Sensometrics, July 11-13, 2012

P&G

The equivalence test problem and Two one-sided tests (TOST)

The equivalence test problem

3

$$H_0: \quad \mu < \delta_1 \quad \vee \quad \mu > \delta_2$$

vs.

$$H_1: \quad \delta_1 \le \mu \le \delta_2$$

Often: $\delta_1 = -\delta_2$

Two one-sided tests (TOST)

4

$$H_0: \quad \mu < \delta_1 \quad \vee \quad \mu > \delta_2$$

vs.

$$H_1: \quad \delta_1 \le \mu \le \delta_2$$

- use two tests, yielding p_1 and p_2
  - H_{01}: $\mu < \delta_1$ vs. H_{11}: $\mu \ge \delta_1$
  - H_{02}: $\mu > \delta_2$ vs. H_{12}: $\mu \le \delta_2$
- reject H_0 if $\max(p_1, p_2) \le \alpha$

Open symmetric intervals (Ennis & Ennis 2009, 2010)
Binomially distributed data

6

Binomially-distributed data ($\pi = ½$)

H_0: $\mu \leq ½ - \theta$ v $\mu \geq ½ + \theta$

vs.

H_1: $½ - \theta < \mu < ½ + \theta$

$$p = \sum_{k=0}^{n-m} \binom{n}{k} (½ - \theta)^k (½ + \theta)^{n-k}$$

$$- \sum_{k=0}^{m-1} \binom{n}{k} (½ - \theta)^k (½ + \theta)^{n-k}$$

m = min(x, n-x)

Binomially-distributed data (π = ½)

$$p = \sum_{k=m}^{n-m} \binom{n}{k} (½-\theta)^k (½+\theta)^{n-k}$$

$$p_{TOST} = \sum_{k=m}^{n} \binom{n}{k} (½-\theta)^k (½+\theta)^{n-k}$$

➢ higher power than TOST

How much to gain?

TOST vs. Ennis & Ennis: θ = 0.2

- $n \in$ {10 (1) 1000}
- m: E&E significant, TOST not

		p value [%]		**power [%]**	
n	***m***	***EE***	***TOST***	***EE***	***TOST***
16	8	4.9	7.4	19.6	0.0
18	9	3.9	6.0	18.5	0.0
23	11	4.7	5.5	32.2	0.0

TOST vs. Ennis & Ennis: θ = 0.1

		p value [%]		power [%]	
n	*m*	*EE*	*TOST*	*EE*	*TOST*
44	22	4.9	11.6	12.0	0.0
46	23	4.6	10.9	11.7	0.0
48	24	4.3	10.3	11.5	0.0
50	25	4.0	9.8	11.2	0.0
52	26	3.8	9.3	11.0	0.0
54	27	3.6	8.8	10.8	0.0
56	28	3.4	8.3	10.6	0.0
58	29	3.2	7.9	10.4	0.0
60	30	3.0	7.5	10.3	0.0
62	31	2.8	7.1	10.1	0.0
64	32	2.7	6.7	9.9	0.0
66	33	2.5	6.4	9.8	0.0
68	34	2.4	6.0	9.6	0.0
69	34	4.7	7.5	19.0	0.0
70	35	2.3	5.7	9.5	0.0
71	35	4.5	7.1	18.7	0.0
72	36	2.2	5.5	9.4	0.0
73	36	4.3	6.7	18.5	0.0
74	37	2.0	5.2	9.2	0.0
75	37	4.0	6.4	18.2	0.0

		p value[%]		power [%]	
n	*m*	*EE*	*TOST*	*EE*	*TOST*
77	38	3.8	6.1	18.0	0.0
79	39	3.6	5.8	17.8	0.0
81	40	3.4	5.5	17.6	0.0
83	41	3.3	5.2	17.4	0.0
84	41	4.9	6.3	25.6	8.7
86	42	4.6	6.0	25.3	8.6
88	43	4.4	5.7	25.1	8.5
90	44	4.2	5.4	24.8	8.4
92	45	4.0	5.2	24.5	8.3
97	47	4.8	5.6	31.5	16.1
99	48	4.6	5.3	31.2	15.9
101	49	4.3	5.1	30.9	15.8
106	51	5.0	5.5	37.3	22.9
108	52	4.8	5.2	36.9	22.7
117	56	4.8	5.1	42.1	28.8
133	63	4.9	5.1	51.2	39.7
156	73	5.0	5.0	62.1	52.9

TOST vs. Ennis & Ennis: θ = 0.05

		p value [%]		power [%]	
n	*m*	*EE*	*TOST*	*EE*	*TOST*
98	49	4.9	18.6	8.0	0.0
100	50	4.8	18.3	8.0	0.0
102	51	4.7	18.0	7.9	0.0
104	52	4.6	17.7	7.8	0.0
106	53	4.5	17.4	7.7	0.0
108	54	4.5	17.2	7.7	0.0
110	55	4.4	16.9	7.6	0.0
112	56	4.3	16.6	7.5	0.0
114	57	4.2	16.4	7.5	0.0
116	58	4.1	16.1	7.4	0.0
118	59	4.1	15.9	7.3	0.0
120	60	4.0	15.6	7.3	0.0
122	61	3.9	15.4	7.2	0.0
124	62	3.8	15.2	7.2	0.0
126	63	3.8	15.0	7.1	0.0
128	64	3.7	14.7	7.0	0.0
130	65	3.6	14.5	7.0	0.0
132	66	3.6	14.3	6.9	0.0
134	67	3.5	14.1	6.9	0.0
136	68	3.4	13.9	6.8	0.0

…

		p value[%]		power [%]	
n	*m*	*EE*	*TOST*	*EE*	*TOST*
507	247	4.9	5.1	46.6	40.6
509	248	4.8	5.0	46.5	40.5
522	254	4.9	5.1	48.8	43.1
524	255	4.9	5.0	48.8	43.0
537	261	5.0	5.1	51.0	45.4
539	262	4.9	5.1	50.9	45.3
541	263	4.9	5.0	50.8	45.3
554	269	5.0	5.1	53.0	47.6
556	270	4.9	5.0	52.9	47.5
569	276	5.0	5.1	54.9	49.8
571	277	4.9	5.0	54.9	49.7
586	284	5.0	5.0	56.7	51.7
601	291	5.0	5.0	58.5	53.7
616	298	5.0	5.0	60.2	55.6
631	305	5.0	5.0	61.9	57.4
646	312	5.0	5.0	63.4	59.1
661	319	5.0	5.0	64.9	60.8

Approach of Ennis & Ennis for binomially-distributed data

- higher power than TOST
- gain in power limited for reasonably powered studies
  (none for power > 70%)
- only for symmetrical margins
- only for π = ½

Adjusted non-central chi-square (ANC) Normally distributed data

Adjusted non-central chi-square (ANC)

H_0: $\mu^2 \geq \theta^2$ vs. H_1: $\mu^2 < \theta^2$

- based on non-central χ^2-distribution
- numerical optimization required for critical values
- approximate solution called ANC

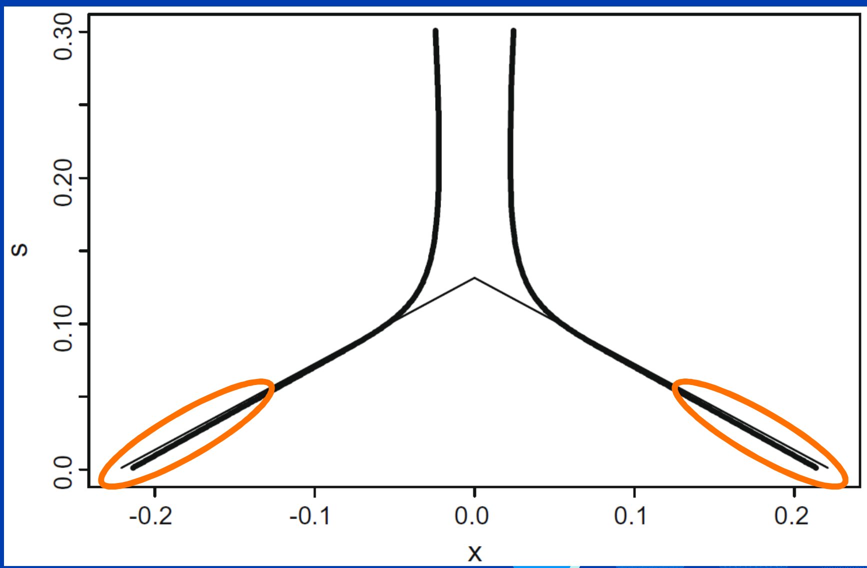
Power comparison

as reported by Berger & Hsu (1996) and Ennis & Ennis (2009)

σ	**0.04**	**0.08**	**0.12**	**0.16**	**0.20**	**0.30**	**∞**
TOST	100	72.0	15.8	0.7	0.0	0.0	0.0
ANC	100	70.6	26.0	13.0	9.1	6.5	5.0
BHM	100	72.1	26.0	13.1	9.3	6.6	5.0
BH	100	72.0	24.7	12.8	9.2	6.6	5.0

15

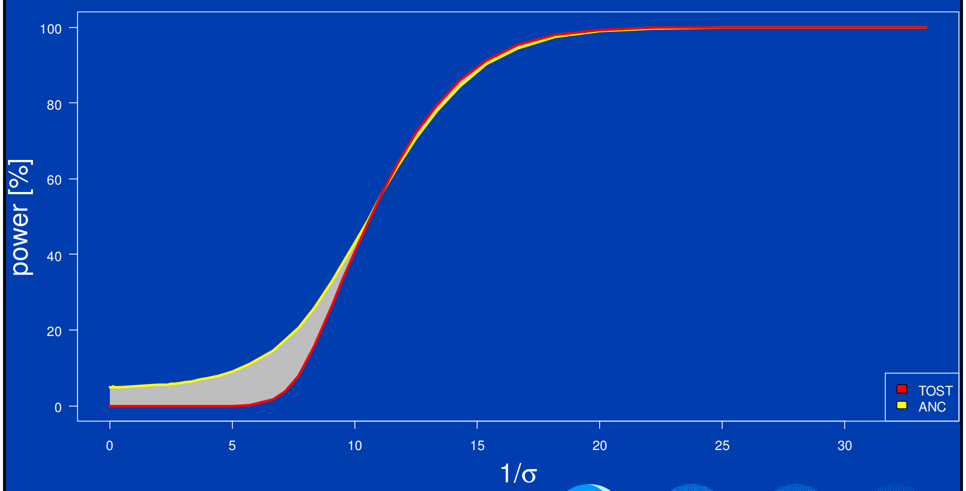
Rejection regions TOST and ANC



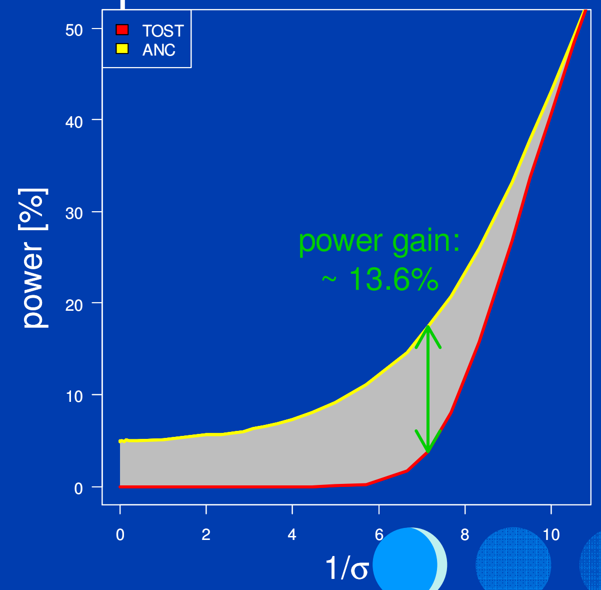
From Ennis & Ennis (2009, 2010)

16

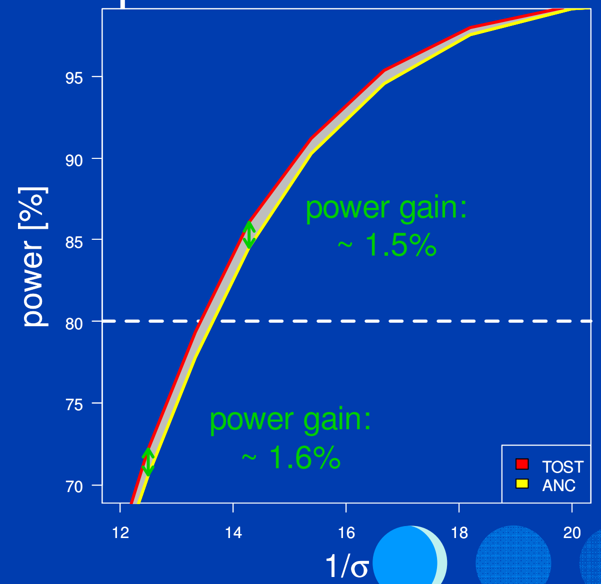
Power comparison TOST vs. ANC



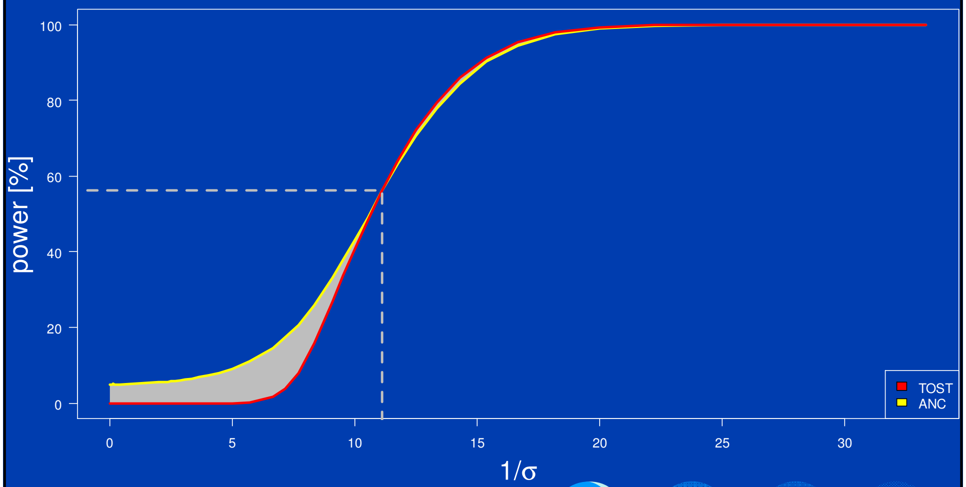
Power comparison TOST vs. ANC



Power comparison TOST vs. ANC



Power comparison TOST vs. ANC



Adjusted non-central chi-square (ANC)

- higher power than TOST for low-powered studies
- lower power for reasonably powered studies (break-even ~ 55% power)
- only for symmetrical margins

Adjusted non-central chi-square (ANC)

- possible gain if σ can be bounded
- Ennis & Ennis (2009) report 72.3% power vs. 72.0% (TOST) and 70.6% (unbounded ANC)
  (note: this assumes we specify the upper bound for the variability exactly at the true value)
- a Bayesian version of the TOST could be expected to benefit similarly from using this prior knowledge

Conclusions

- Binomial data: E&E more powerful than TOST (but often no difference)
- Normally distributed data:
  ANC much higher power for underpowered studies (5 – 40% power)
- TOST outperforms ANC for studies with a reasonable power (> 60% power)

23

References

- Ennis DM & Ennis JM (2009). Hypothesis Testing for Equivalence Defined on Symmetric Open Intervals. *Communications in Statistics – Theory and Methods, 38,* 1792-1803.
- Ennis DM & Ennis JM (2010). Equivalence Hypothesis Testing. *Food Quality and Preference, 21,* 253-256.
- Meyners M (2012). Equivalence testing – a review. *Food Quality and Preference 26, 231-245.*