

2020 United States COVID-19 Vaccination Preference (CVP) Study

Appendix 2. Internal Validity

16 November 2020

Benjamin M. Craig, PhD, benjamin.craig@iahpr.org

This appendix examines the association between the randomized components of the survey instrument and respondent behaviors. In this study, each respondent is randomly assigned an attribute order and a series of eight sets (i.e., block), which are shown in a random sequence and with random object position (i.e., A, B, C). These random components may have influenced respondent behaviors:

1. **length of survey:** time in seconds from first page load to clicking submit
2. **straight lining:** identical responses for all tasks (always same column, e.g., always opt-out)
3. **consistency:** choosing opt-out in the singleton and choice tasks consistently
4. **dominance:** selection of a dominated object during the warm-up task
5. **response time:** time in seconds from page load to clicking next in a choice task
6. **changed response:** one or more responses before the final response in a choice task
7. **response location:** the location of the response option selected in a choice task

Understanding the effects of random components aids in the interpretation of the preference evidence. This study includes four choice tasks: singleton, non-adaptive, adaptive, and qualitative; however, this report focuses on behaviors related to the non-adaptive tasks.

Except for blocks, all randomized components were assigned independently using respondent-level random number seeds. Blocks were assigned serially to distribute them more evenly within quotas.

To assess whether the randomized components were uniformly distributed, chi-square goodness-of-fit tests were conducted to produce p-values. Rejecting a uniform distribution is noteworthy but does not imply a loss in internal validity if the component is randomized. Failing to reject a uniform distribution is also a noteworthy indicator of entropy. In addition, chi-square tests were conducted to assess the association between individual-level random components and quotas.

To assess the association between the randomized components and the time-related behaviors, Wald tests for quantile regressions were conducted to produce p-values.

To assess the association between the randomized components and discrete behaviors, chi-square tests were conducted to produce p-values.

An association between response behaviors and randomized components implies that aspects of the experiment beyond preferences caused respondents to alter their behaviors. The identification of a significant association implies a loss in internal validity and may motivate further investigation of its source as well as its control within a sensitivity analysis.

Table 1. Randomized components

	Proportions	H ⁰ : uniform p-value	Quotas p-values
Designs for each respondent (3 designs)*	31.83 to 35.82	0.197	0.923
Blocks for each respondent (21 blocks)	3.64 to 5.98	0.477	1.000
Attributes order for each respondent (120 orders [5!])	0.09 to 7.03	<0.001	0.290
Sequence positions of each set (8 positions)	9.80 to 14.48	0.714	
Object positions in each set (6 orders [3!])	6.24 to 34.17	<0.001	

* For example, the respondents were randomly assigned to one of three possible designs: 31.83% random design, 32.35% generator-developed design, and 35.82% efficient design. The chi-square goodness of fit test fails to reject the null hypothesis of a uniform distribution (33%, 33%, 33%) and the second chi-square test fails to find association between design and quota.

Table 2. Length of survey

	Median	IQR	p-value
All	1130	790-1584	
Designs (3)			
Random	1122	803-1618	0.932
Generator-developed	1122	760-1540	
Efficient design	1139	800-1590	
Blocks (21)			
Minimum	898	764-1515	0.013
Maximum	1465	967-1909	

Table 5. Median length of survey conditional on an attribute placed at a specific row

	Proof*	Setting	Effectiveness	Duration	Risk**	p-value
First row	1121.34	1139.29	1036.04	1131.94	1177.45	0.325
Second row	1149.26	1092.26	1141.74	1133.98	1130.4	0.913
Third row	1157.63	1132.87	1101.95	1118.76	1132.8	0.946
Fourth row	1136.08	1133.98	1197.26	1087.79	1051.75	0.198
Fifth row	1055.32	1155.81	1131.34	1156.93	1141.58	0.455
p-value	0.512	0.927	0.158	0.869	0.359	

* For example, when proof of vaccination is in the first row, the median length of survey is 1121.34 seconds. The attribute in the first row not affects the median response time (p-value = 0.325).

Table 3. Straight lining, consistency, and dominance*

	All	Straight lining	Always opt-out	Never opt-out	Sometimes opt-out	p-value
All	100.00 (1153)	2.78 (32)	7.55 (87)	76.76 (885)	12.92 (149)	
Vaccination question						
Yes	61.06 (704)	2.98 (21)	0.85 (6)	88.07 (620)	8.10 (57)	<0.001
Don't know/not sure	22.38 (258)	1.55 (4)	9.69 (25)	68.60 (177)	20.16 (52)	
No	16.57 (191)	3.66 (7)	29.32 (56)	46.07 (88)	20.94 (40)	
Warmup task						
Best vaccination	64.18 (740)	2.03 (15)	1.89 (14)	83.24 (616)	12.84 (95)	<0.001
Dominated once	14.83 (171)	5.85 (10)	0.00 (0)	83.04 (142)	11.11 (19)	
Dominated twice	12.32 (142)	4.93 (7)	1.41 (2)	82.39 (117)	11.27 (16)	
Opt-out	8.67 (100)	0.00 (0)	71.00 (71)	10.00 (10)	19.00 (19)	
Designs (3)**						
Random	35.82 (413)	2.66 (11)	8.72 (36)	74.09 (306)	14.53 (60)	0.329
Generator-developed	32.35 (373)	3.75 (14)	7.24 (27)	78.55 (293)	10.46 (39)	
Efficient design	31.83 (367)	1.91 (7)	6.54 (24)	77.93 (286)	13.62 (50)	

* Straight lining refers to always choose the same vaccination position. Opt-out is “no vaccination for six months.”

** The association between the 21 blocks and these four response categories is 0.981. The association between attribute position and these four categories is 0.994. This test was repeated to examine the attributes in specific positions and the specific positions of each attribute, and their p-values ranged from 0.166 to 0.977.

Table 5. Median response time conditional on an attribute placed at a specific row

	Proof*	Setting	Effectiveness	Duration	Risk**	p-value
First row	10.06	8.35	8.52	10.23	11.09	<0.001
Second row	9.22	9.24	10.42	10.15	10.22	<0.001
Third row	9.96	10.41	9.69	9.25	9.38	0.007
Fourth row	9.07	10.27	10.05	9.38	9.46	0.005
Fifth row	10.01	10.78	10.18	9.74	8.9	<0.001
p-value	0.014	<0.001	<0.001	0.021	<0.001	

* For example, when proof of vaccination is in the first row, the median response time is 10.06 second. The attribute in the first row affects the median response time (p-value <0.001).

Table 5. Changed response conditional on an attribute placed at a specific row

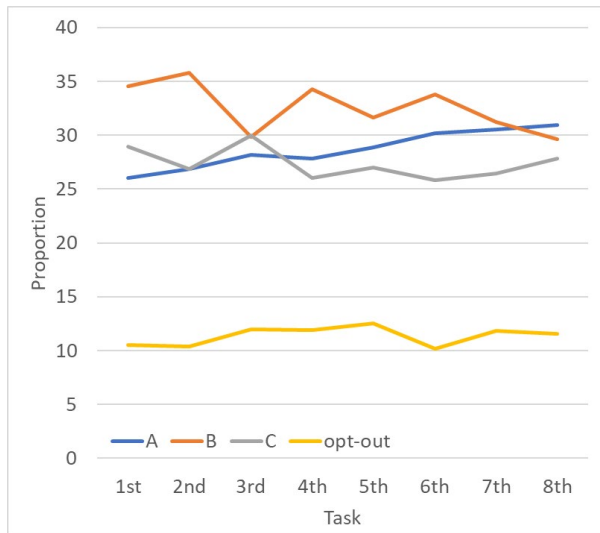
	Proof*	Setting	Effectiveness	Duration	Risk**	p-value
First row	6.12	6.32	5.12	8.67	7.28	<0.001
Second row	7.45	7.20	6.41	5.79	6.26	0.147
Third row	6.86	7.59	6.77	5.5	6.41	0.104
Fourth row	5.84	5.75	7.71	7.56	6.40	0.023
Fifth row	7.13	5.66	7.13	5.88	6.90	0.263
p-value	0.204	0.046	0.026	<0.001	0.681	

* For example, when proof of vaccination is in the first row, the median length of survey is 1121.34 seconds. The attribute in the first row not affect the median response time (p-value = 0.325).

Table 4. Response time and changed response

	Response time			Changed response	
	Median	IQR	p-value	%	p-value
All	9.74	5.01-17.95		6.70	
Designs (3)					
Random	10.25	5.34-18.56	0.001	7.18	0.333
Generator-developed	9.32	4.78-16.93		6.46	
Efficient design	9.55	4.95-18.16		6.39	
Blocks (21)					
Minimum	8.11	4.58-13.46	<0.001	4.48	<0.001
Maximum	12.63	6.76-20.18		11.64	
Set (168 [21 x 8])					
Minimum	6.11	3.93-10.90	0.920	0.00	<0.001
Maximum	12.96	5.75-20.18		19.75	
Task sequence (8)					
Warm-up	18.27	9.42-30.60	<0.001	11.19	<0.001
1 st task	12.79	6.40-23.52		7.55	
2 nd task	10.19	5.46-18.01		6.16	
3 rd task	9.17	4.89-15.98		5.72	
4 th task	8.88	4.58-16.00		6.58	
5 th task	8.63	4.32-15.59		5.72	
6 th task	8.28	4.37-14.85		5.98	
7 th task	7.92	4.34-14.15		5.81	
8 th task	8.17	4.44-14.03		5.46	

Figure 1. Task sequence and response location (p-value = 0.038)*



* The results showed no association between task sequence and the selection of any specific alternative (p-value 0.873)