

Apéndice:

Evaluación de sesgos de selección del marco muestral y de la muestra

Teniendo en cuenta que el estudio no contó con la información de contacto de todos los y las candidatas (25%), así como de candidatos y candidatas que no terminaron la encuesta en su totalidad (35%), realizamos una evaluación de las diferencias principales de la muestra analizada, con la población de la cual teníamos contacto, y de la población que respondió pero no terminó la encuesta. Esto con el objetivo de identificar principales sesgos de selección que la muestra principal tiene con respecto a estos dos grupos.

Para ello, en la lista población total de los y las candidatas creamos varias variables dicotómicas. Primero, asignamos el valor '1' aquellos candidatos y candidatas de los que teníamos sus correos electrónicos, y '0' a los que no lo hicieron (*contacto*). Segundo, asignamos el valor a '1' a los candidatos y candidatas que comenzaron nuestra encuesta '1', y '0' a los candidatos que no lo hicieron (*participantes*). Tercero, asignamos el valor '1' a los candidatos y candidatas que terminaron la encuesta en un 90% con el valor '1', y '0' a los que no lo hicieron (*muestra final*).

Para observar si hay diferencias sustantivas entre los grupos contactados, participantes y que finalizaron la encuesta, calculamos los porcentajes de cada grupo con respecto a múltiples características de interés. La Tabla A.1. incluye variables de característica de las cuales teníamos acceso:

- (1) El género,
- (2) Categoría de edad,
- (3) Electos,
- (4) Experiencia previa como candidatos en las últimas tres elecciones locales,
- (5) Cargo de elección popular,
- (6) Macro regiones
- (7) Partidos Políticos
- (8) Renuncia oficial al a candidatura
- (9) Candidatos y candidatas en zonas de riesgo electoral identificada por la MOE Departamentos

La Tabla A.1.1 también presenta la diferencia porcentual entre la población total de candidaturas y los porcentajes de población con contacto, población participante y la muestra final. Estas diferencias se obtienen restando el porcentaje de cada uno de estos grupos con respecto a la población total. Para facilitar la interpretación, hemos asignado colores que van del rojo al verde: los tonos más intensos de rojo indican subrepresentación en el grupo, mientras que los tonos más intensos de verde reflejan sobrerepresentación

En general, no se observan sesgos muy grandes en ninguna los grupos al evaluar las diferencias entre los grupos. Primero, al revisar las diferencias entre los contactados y la muestra final, no

encontramos muchas diferencias significativas. Solo hay una discrepancia del 2% en el caso del Partido MAIS y las candidaturas de coaliciones. La subrepresentación del Partido MAIS fue la más consistente en los tres grupos evaluados. En cuanto a las candidaturas de coaliciones, la diferencia se corrigió en la muestra final, que muestra unos porcentajes similares a la población.

Segundo, las diferencias entre los participantes que iniciaron la encuesta y las personas que completaron la muestra no dan cuenta de diferencias significativas. La única variación relevante se observa en el género, donde, aunque las mujeres comenzaron la encuesta en proporciones similares a la población general, fueron más propensas a abandonarla que los hombres. Esta diferencia es de aproximadamente un 3%, lo que genera una ligera subrepresentación de las mujeres candidatas en la muestra final de análisis.

Tercero, la muestra final si presenta algunas subrepresentaciones y sobre-representaciones específicas ya discutidas en el texto principal. Observamos que las candidaturas entre los 51 y 75 años están más subrepresentadas, al igual que las candidaturas para el Concejo. Por otro lado, las candidaturas entre los 35 y 50 años están más sobre-representadas, junto con las de alcaldías y, en menor medida, asambleas. Sin embargo, a pesar de estas diferencias, las distribuciones en las categorías de edad y tipo de cargo se mantienen en línea con las de la población general, por lo que las diferencias porcentuales no afectan significativamente la similitud de la muestra analizada con la población.

Tabla A.1. Proporción de la población de candidatos, de los contactados, los participantes, y la muestra.

Categoría	Vaiable	Población	Contacto	Participantes	Muestra final	Diff contacto	Diff Participants	Diff muestra final
Género	Femenino	39,04	38,70	39,02	36,17	-0,35	-0,03	-2,87
	Masculino	60,96	61,30	60,98	63,83	0,35	0,03	2,87
Categoría de edad	18-35	31,23	31,48	32,43	32,83	0,25	1,21	1,60
	36-50	38,96	39,34	42,84	43,89	0,38	3,89	4,93
	51-75	29,21	28,68	24,44	23,06	-0,53	-4,76	-6,15
	75+	0,58	0,50	0,28	0,23	-0,08	-0,30	-0,35
Electos	Electos	13,43	14,10	15,43	14,86	0,67	2,00	1,43
Experiencia previa como candidatos	Experiencia	30,65	31,50	29,52	29,01	0,85	-1,13	-1,64
Cargo	Alcaldía	4,76	5,38	8,44	9,38	0,62	3,67	4,61
	Asamblea	2,91	2,91	4,54	5,03	0,00	1,63	2,11
	Concejo	80,57	81,27	75,15	73,82	0,70	-5,42	-6,75
	Gobernación	0,19	0,24	0,33	0,33	0,05	0,14	0,14
	JAL	11,56	10,19	11,54	11,44	-1,37	-0,02	-0,12
Regiones	Amazonía	3,57	3,66	4,19	4,27	0,09	0,62	0,70
	Caribe	21,51	21,93	23,23	23,35	0,42	1,71	1,84
	Central	37,62	37,75	34,15	33,34	0,14	-3,47	-4,28
	Eje Cafetero-Antioquia	16,56	15,40	15,33	15,68	-1,16	-1,22	-0,88
	Orinoquía	7,11	7,96	8,11	7,90	0,85	1,01	0,79
	Pacífica	13,63	13,30	14,98	15,46	-0,33	1,35	1,83
Partidos	Partido Liberal	9,09	9,03	8,82	8,95	-0,06	-0,27	-0,15
	Partido Conservador	8,07	8,65	9,01	8,76	0,58	0,94	0,69
	Partido de la U	7,34	8,83	7,62	7,24	1,49	0,27	-0,10
	Partido Cambio Radical	7,20	8,46	7,86	7,71	1,26	0,66	0,52
	Partido Alianza Verde	6,52	7,21	7,31	7,42	0,68	0,79	0,90
	Partido ASI	5,36	6,33	6,45	6,16	0,97	1,09	0,80
	Partido Centro Democrático	5,28	6,85	6,18	6,42	1,56	0,90	1,13
	Partido MAIS	3,79	1,16	1,04	1,05	-2,63	-2,75	-2,74
	Partido Colombia Humana	0,81	0,74	1,16	1,32	-0,07	0,35	0,51
	Partido Nuevo Liberalismo	3,53	3,39	3,25	3,08	-0,13	-0,28	-0,44
	Partido AICO	2,74	2,79	2,89	2,78	0,05	0,15	0,04
	Renuncio a la Candidatura	Renunció	3,07	3,00	2,24	2,04	-0,07	-0,83
Zona de Riesgo Electoral	Zonas	19,59	18,97	21,07	21,55	-0,62	1,48	1,96
	Coaliciones	9,64	7,63	9,43	10,13	-2,01	-0,21	0,49

Tipo de candidatura	GSC	1,08	1,25	1,58	1,71	0,17	0,51	0,64
	Partido Político	89,27	91,11	88,97	88,16	1,84	-0,29	-1,11
Departamento	Amazonas	0,30	0,32	0,44	0,50	0,02	0,14	0,20
	Antioquia	12,12	11,35	11,54	11,86	-0,77	-0,58	-0,25
	Arauca	0,71	0,81	1,01	1,06	0,09	0,30	0,34
	Atlántico	2,82	3,39	4,07	4,16	0,56	1,24	1,34
	Bogotá	1,76	1,88	2,30	2,56	0,11	0,53	0,80
	Bolívar	4,17	4,55	4,62	4,69	0,38	0,45	0,52
	Boyacá	6,57	7,01	5,42	5,09	0,44	-1,15	-1,48
	Caldas	2,78	2,58	2,32	2,35	-0,19	-0,45	-0,42
	Caquetá	1,20	1,29	1,32	1,28	0,09	0,12	0,07
	Casanare	1,64	1,94	2,01	2,02	0,31	0,37	0,38
	Cauca	3,04	2,67	3,05	3,35	-0,37	0,01	0,32
	Cesar	2,81	2,44	2,46	2,71	-0,37	-0,35	-0,10
	Chocó	1,52	1,33	1,38	1,25	-0,19	-0,14	-0,27
	Córdoba	3,38	3,37	3,43	3,24	-0,01	0,05	-0,14
	Cundinamarca	10,58	10,42	9,55	9,07	-0,16	-1,03	-1,51
	Guanía	0,22	0,25	0,32	0,37	0,02	0,10	0,14
	Guaviare	0,39	0,46	0,50	0,52	0,07	0,11	0,13
	Huila	3,24	3,17	2,76	2,79	-0,08	-0,48	-0,45
	La Guajira	2,07	1,92	2,13	2,08	-0,15	0,07	0,01
	Magdalena	3,73	3,98	4,09	3,99	0,26	0,36	0,27
	Meta	2,69	3,03	2,97	2,81	0,34	0,28	0,12
	Nariño	3,57	3,47	3,65	3,80	-0,09	0,08	0,24
	Norte de Santander	3,76	3,86	4,01	4,16	0,10	0,25	0,40
	Putumayo	1,29	1,20	1,50	1,51	-0,09	0,21	0,22
	Quindío	1,67	1,47	1,48	1,46	-0,19	-0,19	-0,20
	Risaralda	1,66	1,80	1,81	1,69	0,15	0,15	0,04
	San Andrés	0,13	0,16	0,17	0,14	0,03	0,04	0,01
	Santander	7,13	7,26	6,42	6,09	0,14	-0,71	-1,03
	Sucre	2,40	2,12	2,25	2,34	-0,28	-0,15	-0,06
	Tolima	4,57	4,15	3,69	3,57	-0,42	-0,88	-1,00
	Valle del Cauca	5,51	5,82	6,90	7,04	0,32	1,39	1,54
	Vaupés	0,16	0,15	0,11	0,10	-0,01	-0,05	-0,05
Vichada	0,41	0,37	0,32	0,32	-0,04	-0,10	-0,09	
Total	Observaciones	128129	95699	14901	9865			

Evaluación de los sesgos de atrición

Con respecto a las diferencias en el sesgo de atrición, que se refiere a la diferencia sistemática en el abandono de la encuesta en línea, la Figura 1 muestra, en el eje horizontal, el porcentaje de progreso en la encuesta desde el 0% (izquierda) hasta el 100% (derecha). En el eje vertical, se representa la proporción de respuestas que alcanzaron ese nivel de progreso.

La gráfica indica que, de los 5,036 participantes que abandonaron la encuesta antes del 90%, el 50% dejó el estudio entre el 0% y el 11% de progreso en la encuesta. Estas preguntas eran principalmente socio-demográficas. A partir del 10%, la tasa de abandono se mantiene por debajo del 1% hasta el final de la encuesta. Esto indica que en general no existieron preguntas sensibles en la encuesta que hubieran obligado a los y las candidatas abandonar la encuesta.

Sin embargo, se observa un leve salto en la tasa de abandono entre el 25% y el 30% de progreso en la encuesta, correspondiente al 22% de los que abandonaron; en este nivel, la tasa de abandono de un 2%. En este momento de la encuesta, el equipo investigador incluyó una pregunta experimental que podría haber causado el abandono. Este salto también revela una leve diferencia de género, ya que las participantes femeninas son más propensas a abandonar en este punto del proceso.

No obstante, después de este momento, no se observa otro salto significativo en la tasa de abandono a lo largo de la encuesta. Esto sugiere que no hubo preguntas incómodas que pudieran haber influido de manera sistemática en el abandono, salvo la pregunta experimental mencionada. En consecuencia, el abandono de los participantes está principalmente asociada a una fatiga inicial, pero aquellos que superan el 10% de la encuesta, deciden mantenerse y responder la mayoría de las preguntas de la encuesta.

Figura 1. Tasas de abandono de la encuesta por género (N=14901)

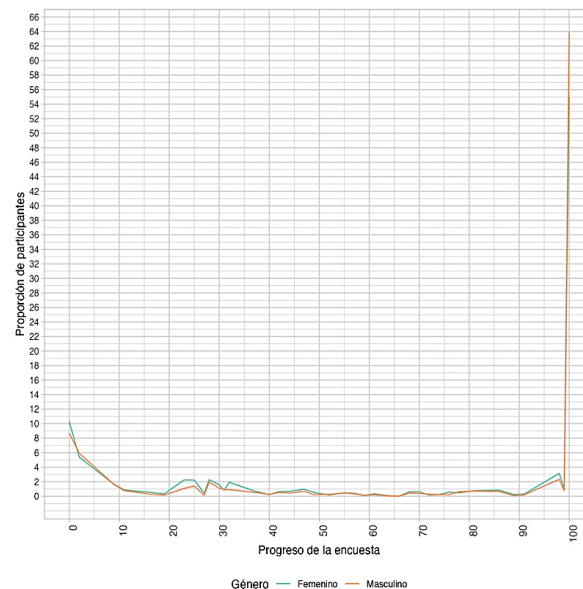


Tabla A2: Regresión logística binomial usando como variable dependiente la violencia física y amenazas

	<u>Violencia física / amenazas</u>
Intercepto	-2.276 *** (0.170)
Masculino	0.048 (0.067)
Afrocolombiano	0.070 (0.093)
Indígena	0.199 (0.128)
Edad joven	0.491 *** (0.070)
Alta posibilidad de ganar	0.152 ** (0.064)
No electo	-0.034 (0.087)
Cargo uninominal	0.630 *** (0.110)
Violencia previa	3.420 *** (0.068)
Trayectoria social	0.312 *** (0.064)
Presencia GAI	0.173 ** (0.083)

Urbano	-0.058
	(0.070)
Trayectoria política (años)	0.005
	(0.003)
Ingreso	0.016
	(0.022)
Educación universitaria	0.004
	(0.074)
<hr/>	
Null deviance	11300.4
Residual deviance	7029.7
AIC	7123.7

Notas: Significancia: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.01$.

Este análisis se realizó utilizando un modelo de regresión logística binomial para evaluar la probabilidad de experimentar diferentes formas de violencia en función de múltiples factores. El grupo étnico usado como referencia es “otro grupo étnico” agrupa blancos, mestizos y población ROM. Para la edad se compara mayores a 35 años versus menores a 35 años. El grupo de referencia para presencia GAI es “sin presencia”. Incluye también efectos fijos por departamento.

Tabla A3: Regresión logística binomial usando como variable dependiente si el candidato tiene familia en cargos públicos de elección

	Familia en política
Intercepto	-3.103 ***
	(0.224)
Masculino	-0.150 **
	(0.068)
Afrocolombiano	-0.046
	(0.095)

Indígena	0.032 (0.127)
Edad joven	0.393 *** (0.071)
Alta posibilidad de ganar	0.183 ** (0.065)
Electo	0.294 *** (0.081)
Cercano a un círculo político nacional o regional	0.311 *** (0.064)
Renglón en top 10	0.158 (0.098)
Recibió apoyo económico del partido político u otro político	-0.145 (0.135)
Votos	0.000 (0.000)
Cargo uninominal	0.004 (0.106)
Violencia física / amenazas	0.101 (0.065)
Trayectoria social	0.102 (0.065)
Urbano	0.071 (0.072)
Trayectoria política (años)	0.014 ***

	(0.003)
Ingreso	0.082 ***
	(0.022)
Educación universitaria	0.197 **
	(0.080)
Null deviance	7062.6
Residual deviance	6838.5
AIC	6938.5

Notas: Significancia: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.01$.

Este análisis se realizó utilizando un modelo de regresión logística binomial para evaluar la probabilidad de experimentar diferentes formas de violencia en función de múltiples factores. El grupo étnico usado como referencia es “otro grupo étnico” agrupa blancos, mestizos y población ROM. Para la edad se compara mayores a 35 años versus menores a 35 años. Violencia física /amenazas incluye intentos de asesinato, amenazas, intimidaciones y ataques físico. Renglón en top 10 indica si el candidato en listas plurinominales estuvo en el top 10 de la lista, para uninominales es siempre 1.. Incluye también efectos fijos por departamento.

Tabla A4: Regresión logística binomial usando como variable dependiente si la candidata considera que fue incluida como cuota de género

	Cuota de género
Intercepto	-2.115 ***
	(0.388)
Afrocolombiano	-0.008
	(0.118)
Indígena	0.323 *
	(0.174)
Edad joven	-0.220 **
	(0.089)

Familia en política	-0.173 (0.113)
Alta posibilidad de ganar	-0.324 *** (0.085)
Electo	-0.564 *** (0.134)
Círculo cercano político nacional o regional	-0.045 (0.080)
Renglón top 10	-0.139 (0.116)
Apoyo económico de partido o político	-0.033 (0.171)
Votos	-0.000 (0.000)
Plurinominales	1.995 *** (0.270)
Violencia	0.318 *** (0.083)
Trayectoria Social	-0.010 (0.083)
Urbano	0.167 * (0.089)

Trayectoria política (años)	-0.002
	(0.005)
Ingreso	-0.031
	(0.029)
Con educación universitaria	0.082
	(0.097)
<hr/>	
Null deviance	3880.9
Residual deviance	3671.2
AIC	3771.2
<hr/>	

Notas: Significancia: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.01$.

Este análisis se realizó utilizando un modelo de regresión logística binomial para evaluar la probabilidad de experimentar diferentes formas de violencia en función de múltiples factores. El grupo étnico usado como referencia es "otro grupo étnico" agrupa blancos, mestizos y población ROM. Para la edad se compara mayores a 35 años versus menores a 35 años. Violencia incluye cualquier tipo de violencia experimentada por la candidata.. Renglón en top 10 indica si el candidato en listas plurinominales estuvo en el top 10 de la lista, para uninominales es siempre 1. Incluye también efectos fijos por departamento.