

Informe Panel de Expertos Marco Muestral Casen 2009¹

Junio del 2010

Antecedentes

1. El Panel de Expertos se origina en una solicitud del Ministro de Mideplan al Programa de Naciones Unidas para el Desarrollo (PNUD) y se constituye con expertos de entidades académicas y de un coordinador del PNUD. Su objetivo es emitir una opinión técnica e independiente respecto del marco muestral usado en la encuesta Casen 2009, habida cuenta que se disponía de una alternativa en la materia que era recomendada por el Instituto Nacional de Estadísticas. Específicamente, al Panel de Expertos se le ha pedido responder a las siguientes tres preguntas:
 - a. Descripción comparativa del marco muestral usado por la Casen 2009 y el recomendado por el INE, haciendo referencia, al menos, a las principales diferencias existentes (en lo cuantitativo y cualitativo).
 - b. Pronunciamiento sobre las implicancias en términos de sesgos relevantes que se podrían anticipar producto del uso de un marco muestral respecto del otro.
 - c. Pronunciamiento sobre la conveniencia de usar el marco muestral utilizado en Casen 2009 respecto de la alternativa propuesta por el INE y las implicancias en los resultados de la encuesta Casen 2009.
2. La mayoría de las encuestas sociales de importancia que se levantaron en el país en los años siguientes al censo de población del 2002 usaron muestras basadas en el marco muestral maestro de secciones (MMS). Este marco fue construido por el INE a partir de los datos censales y debe ser actualizado para dar cuenta de las viviendas que se construyen entre la fecha del censo y la encuesta correspondiente². Las secciones corresponden a áreas geográficas que contienen un determinado número de viviendas definidas al interior de los distritos censales. Las encuestas que usaron este marco muestral corresponden a aquellas desarrolladas por el mismo INE o por instituciones que adquirieron la muestra en el INE.
3. La encuesta Casen 2009 usó el citado marco muestral de secciones, actualizado para tales efectos por el INE por petición (y financiamiento) de Mideplan. La actualización se realizó en terreno a través de una revisión de las áreas incluidas en las secciones

¹ Los miembros del Panel de Expertos son Andrés Hernando, Bernardita Vial, Jeanette Fuentes, Pablo Marshall y Osvaldo Larrañaga (ccordinador). La secretaría técnica estuvo a cargo de Mariana Huepe

² Entre los censos 1992 y 2002 la tasa anual de crecimiento de nuevas viviendas fue 2,6% en promedio.

pertenecientes a la muestra, incluyendo las nuevas viviendas encontradas en dicha área. De acuerdo a información entregada por el INE, consideraciones de costo económico determinaron que solo se actualizaran las áreas geográficas de primera etapa seleccionadas en las muestras.

4. La directora del INE ha planteado al actual Ministro de Mideplan que la decisión de utilizar el marco muestral de secciones en la Casen 2009 fue errónea, puesto que para esa fecha se disponía de un marco muestral más completo, conocido como “Marco Muestral de Manzanas” (MMM). Las manzanas son unidades geográficas con límites definidos para las cuales se dispone de información cartográfica y del número de viviendas en cada una de ellas. Si bien este marco muestral también se basa en los datos censales del 2002, no estaría afecto a los problemas del marco muestral de secciones.
5. El INE plantea que hacia el año 2009 el MMS había acumulado dos problemas importantes. Primero, se trataba de un marco muestral “agotado” en el sentido que había sido usado profusamente en encuestas durante los años anteriores, de modo que las viviendas ya habían sido visitadas para efectos de una u otra encuesta. La probabilidad de repetición de la vivienda depende directamente del número y tamaño de encuestas basadas en tal marco muestral; siendo necesario considerar al respecto que solo la encuesta nacional de empleo del INE entrevistaba cada mes un total cercano a 34.500 viviendas en la modalidad de panel rotatorio. El segundo problema es que el proceso de actualización usado en el marco muestral de secciones tenía insuficiencias debido a que las secciones no eran áreas físicas fácilmente reconocibles, dificultando la posterior actualización, que este proceso no se realizaba con un procedimiento sistemático, y que estaba restringido a mantener un número fijo de viviendas en cada sección.
6. El resultado, de acuerdo al mismo INE, es que el marco muestral de secciones estaba solo parcialmente actualizado y no incorporaba bien el cambio demográfico y espacial entre el 2002 y el 2009, por lo que no era totalmente representativo de la población nacional en esa última fecha. Además, la sobre explotación a que había estado afecto ese marco muestral indujo a que se efectuaran reemplazos por secciones que estaban fuera de la muestra maestra, introduciéndose sesgos adicionales en la muestra.
7. El INE argumenta que el nuevo marco muestral de manzanas se desarrolló como respuesta a los problemas que acumulaba el marco de secciones, y que para el año 2008 todas las encuestas sociales de esa institución estaban ya migrando al nuevo marco muestral, así como aquellas desarrolladas por otras entidades que usaban estas muestras (el INE dejó de actualizar el marco muestral de secciones en el año 2007). Bajo tal contexto el INE propuso a Mideplan que la encuesta Casen 2009 utilizara el nuevo marco muestral y considera una decisión equivocada de Mideplan no haberlo hecho así.

Según el INE, el nuevo marco estaba disponible; había sido ya probado en otras encuestas; representaba para Mideplan un menor costo monetario respecto de la actualización requerida en el marco muestral de secciones; y ofrecía mayor facilidad en el trabajo de terreno considerando que las viviendas no habían sido entrevistadas previamente. En cambio, plantea el INE, el marco de secciones que utilizó en definitiva Mideplan produce resultados con sesgos, originados en los problemas señalados.

8. El responsable directo de la encuesta Casen 2009 por parte de Mideplan es el director de la división social, quien tiene una visión distinta del proceso que originó la decisión de usar el marco de secciones en la Casen 2009. El director de la división social plantea lo siguiente: (i) que en la encuesta Casen 2006 se adoptó el marco muestral de secciones por iniciativa del INE, por lo cual no se entiende que esa institución argumente en su contra tres años más tarde; (ii) que el marco muestral de manzanas no estaba suficientemente validado en el 2009, y que durante todo ese año el INE realizó en forma paralela la encuesta nacional de empleo vigente y la nueva encuesta nacional de empleo, antes que esta última saliera a luz pública a inicios del 2010 (la nueva encuesta supone una modificación en el marco muestral y en el cuestionario); (iii) que las conversaciones con el INE durante el 2009 ocurrieron demasiado tarde en el año como para que Mideplan pudiera haber tomado una decisión responsable a efectos de modificar el marco muestral de la encuesta Casen; (iv) que el INE no entregó mayores antecedentes metodológicos, ni de resultados esperados, ni del trabajo de campo del nuevo marco, dificultando el citado proceso de decisión.
9. Cabe agregar que la discusión de las bondades relativas entre los marcos muestrales es atingente sólo a las áreas urbanas, puesto que el marco muestral de manzanas sólo cubre la zona urbana.

Opinión del Panel

10. No corresponde a este panel realizar una investigación inquisitoria para dilucidar los aspectos contradictorios de las versiones que entregaron las partes en disputa sobre los acontecimientos del 2009 que derivaron en la adopción del marco muestral de secciones en el encuesta Casen de ese año, pero cabe mencionar que se trató de un proceso conflictivo que derivó en el quiebre de las conversaciones entre los responsables del INE y Mideplan en la época. También hay que hacer notar sobre la existencia de una tensión subyacente entre el mandato institucional del INE y de Mideplan, que puede proyectarse hacia el futuro y generalizar a otras entidades del sector público. En efecto, de los antecedentes vistos por este Panel se desprende que el INE cree entender que es la institución rectora del sector público en materia de las estadísticas generadas por encuestas y censos, y que le corresponde fijar estándares entre las agencias del sector.

Por su parte, profesionales de Mideplan expresan que el INE hace uso y abuso de poder monopólico, tanto en la fijación de precios para los productos que ofrece a otras instituciones como en la (falta de) información que hace disponible respecto de sus procesos de producción. Se argumenta que Mideplan es responsable de generar y difundir la cifra oficial de pobreza en el país, por lo cual debe velar por la calidad de la información en que se basan tales estimadores (incluyendo el marco muestral del INE, entre otros insumos). Es muy posible que la discusión del marco muestral de la Casen 2009 haya estado cruzada por este tipo de disputa sobre los mandatos institucionales.

11. El nuevo marco muestral en el cual se basa el INE para la selección de muestras a hogares en zonas urbanas está basado en unidades de selección de primera etapa conformado por manzanas geográficas en lugar de secciones. Algunas características de este cambio son: (i) la unidad primaria de muestreo constituida por la manzana es más fácil de identificar en el trabajo de campo, al estar anclada a una variable geográfica bien definida, por lo cual evita cometer errores no muestrales en la selección de las viviendas; además hace más fácil el monitoreo del trabajo de campo y posibilita que el proceso de actualización de viviendas sea fácil, rápido y preciso; (ii) el trabajo de campo debería ser más barato de realizar en el marco muestral de manzanas, puesto que la búsqueda de direcciones se concentra en una zona geográfica acotada; (iii) las manzanas geográficas son unidades más homogéneas que las secciones, a las cuales se les puede asociar a variables de interés de modo de optimizar los tamaños muestrales; (iv) el nuevo marco muestral está constituido por el censo completo, razón por la cual el problema de saturación de viviendas se minimiza.
12. Por su parte, las secciones son áreas no estrictamente acotadas en términos geográficos, y en cuya definición importan factores como el número de viviendas a incluir. En general, las secciones incluyen un mayor número de viviendas que la manzana; por lo cual el marco muestral de secciones tiende a tener un menor número de unidades primarias y un mayor número de unidades secundarias de muestreo, respecto del marco muestral de manzanas. Al respecto, no hay estudios que permitan establecer fehacientemente ventajas de un tipo de marco sobre otro en materia del nivel de precisión de las estimaciones.
13. El marco muestral de secciones usado por la Casen 2009 podría estar afectos a dos tipos de sesgo. Primero, puede haber una pérdida de representatividad de la muestra originada en una actualización parcial de las nuevas viviendas, sea por que las secciones son unidades difíciles de seguir en el tiempo o sea porque la actualización del 2009 sólo se realizó a nivel de las secciones identificadas en el marco muestral existente. Segundo, el agotamiento del marco muestral puede provocar una mayor tasa de rechazo de las entrevistas en algunos de los hogares que habían sido entrevistados en encuestas anteriores. Si estos hogares son distintos a los hogares que responden la entrevista se

genera un sesgo de selección, que puede ser solo parcialmente corregido a través de los factores de expansión. El agotamiento de la muestra también puede producir otros sesgos que son muy difíciles de cuantificar, como es el error de respuesta durante la entrevista.

14. La pregunta relevante es cuán importante es el sesgo asociado a los problemas descritos, para lo cual se requiere examinar evidencia empírica en la materia. A tal respecto hubiese sido importante contar con los datos de la Casen 2009, pero ello no fue posible atendiendo a que se encuentran en etapa de validación y análisis. Los antecedentes que se entregan a continuación permiten una aproximación razonable, pero no exacta, al problema.
15. Para aproximar cuantitativamente el sesgo que se produce por falta de actualización de la muestra se pueden examinar las características de los hogares según período de adquisición de la vivienda, considerando que el problema de la muestra está asociada a la sub-representación de las viviendas nuevas. Este análisis se realiza en base a los datos de la encuesta Casen 2006. A tal respecto hay que precisar que la información disponible se refiere a la fecha de adquisición de la vivienda, la cual es una aproximación no exacta de la fecha de construcción de la vivienda, puesto que una parte de las viviendas recién adquiridas son viviendas antiguas. En el caso de las viviendas que se adquieren con subsidios, que corresponden aproximadamente a la mitad del total, la correspondencia entre fecha de adquisición y fecha de construcción es mucho mayor, porque prácticamente todos los programas sociales de adquisición de vivienda financiaban solo la compra de viviendas nuevas (hasta muy recientemente).
16. El Cuadro 1 muestra el resultado del ejercicio descrito, considerando como viviendas nuevas a las adquiridas en los tres últimos años.³ El análisis se realiza a nivel de áreas urbanas, puesto que este es el segmento de interés para fines del Informe (los marcos muestrales de secciones y manzanas bajo análisis difieren solo en el componente urbano). Como podría esperarse, los jefes de hogar en viviendas nuevas son considerablemente más jóvenes que en las viviendas antiguas (43,9 años vs 55,3 años), a la vez que hay un mayor número de menores de 14 años en el hogar (1,10 vs 0,79). Asimismo, estos jefes de hogar tienen mayor escolaridad, pero ello podría explicarse por su menor edad dado que las cohortes más jóvenes son en promedio más educadas.

³ La información disponible es exacta respecto a las viviendas antiguas, puesto que todas las adquiridas hace más tres años fueron construidas hace más de tres años. Respecto de las viviendas nuevas, la mitad de ellas corresponde a viviendas sociales para las cuales la fecha de adquisición corresponde aproximadamente a la fecha de construcción. Si se supone que un 50% de las viviendas adquiridas sin subsidios son viviendas nuevas, se tiene que un 75% de las viviendas del total adquiridas en los últimos tres años serían efectivamente viviendas recién construidas.

Por otra parte, no hay diferencias de importancia en el número total de personas en el hogar o en el porcentaje de jefes mujeres.

17. En la dimensión socioeconómica se constatan diferencias significativas en la estructura de los hogares que habitan viviendas nuevas y antiguas. Los primeros tienen una mayor tasa de pobreza (12,9% vs 11,3%) y de indigencia (4,4% vs 2,2%). Al mismo tiempo, esos hogares están también más representados en el quintil de ingresos más alto (28,1% vs 22,3%). Por tanto, hay mayor desigualdad en la distribución de ingresos en los hogares en viviendas nuevas, probablemente por la confluencia de diferentes dinámicas que subyacen a la adquisición de viviendas (mecanismo de mercado, programas sociales focalizados en sectores más pobres, ciclo de vida del hogar, etc).

Cuadro 1.

Características de hogares en viviendas nuevas y antiguas, área urbana del país

	Nuevas viviendas	Antiguas viviendas
Promedio personas en hogar	3,76	3,81
Promedio menores de 15	1,10	0,79
Edad promedio jefe hogar	43,9	55,3
Años escolaridad promedio jefe	11,1	9,3
% hogares con jefe mujer	28,6	31,5
% población pobreza	12,9	11,3
% población en indigencia	4,4	2,2
% estrato medio (quintil 3 y 4)	36,2	43,8
% estrato alto (quintil 5)	28,1	22,3

Nota: viviendas nuevas son las adquiridas en últimos tres años

Fuente: Casen 2006

18. De los datos presentados se infiere que hay sesgos en las estimaciones provenientes de encuestas cuyas muestras tengan una sub-representación de viviendas nuevas. La magnitud del sesgo dependerá de cuán significativa sea el peso de las viviendas nuevas en el total, así como de su pérdida de representatividad en la muestra. En la encuesta Casen 2006 las viviendas adquiridas en los últimos tres años representan cerca del 8% del total de viviendas propias, de modo que cualquier variable allí estimada es un promedio ponderado de hogares en viviendas nuevas, con peso de 8%, y hogares en viviendas antiguas, con peso de 92%. Si la muestra disponible sub-representa a los

primeros en un 10%, se tiene que las ponderaciones muestrales efectivamente usadas serían de 7,2% y 92,8%. El sesgo resultante es bastante menor. En el caso de la tasa de pobreza consignada en el Cuadro 1, se tendría un estimador poblacional de 11,42% de pobreza, siendo el verdadero 11,43%. En caso que la pérdida de representatividad fuese un 30%, o sea que 3 de cada 10 nuevas viviendas quedan fuera del marco muestral, resulta un sesgo en la estimación de la tasa de pobreza de de 0,04 puntos porcentuales (11,39% vs 11,43%). En la estimación de otras variables la magnitud del sesgo dependerá de cuánta distancia exista en el valor de la variable entre viviendas nuevas y antiguas, pero en la mayor parte de los casos el efecto final parece ser bastante pequeño.

19. La magnitud del sesgo no se altera mayormente si se consideran algunas modificaciones en los supuestos de la estimación, porque no hay una dirección sistemática en los cambios resultantes. En primer término, si se considera que sólo un 75% de las viviendas adquiridas en los últimos tres años son efectivamente construcciones nuevas (nota pie página 1), la magnitud del sesgo se *reduce* en alrededor de la cuarta parte de la estimación realizada. Segundo, si se supusiera que las viviendas nuevas también están sub-representadas en la encuestas Casen 2006 (porque el marco muestral ya hubiera estaba desactualizado), la magnitud del sesgo *aumenta* en proporción equivalente. Tercero, la magnitud del sesgo no cambia si las características de los hogares que habitan viviendas *nuevas*, incorporadas en el marco muestral, son similares a la de hogares en viviendas *nuevas*, que quedan fuera del marco muestral. Cuarto, el sesgo no se modifica si se considera a las viviendas arrendadas o cedidas, en caso que estas tuvieran igual estructura de antigüedad que las viviendas adquiridas (8% construidas en los últimos tres años).⁴

20. Otra fuente de datos relevante para conocer la eventual magnitud de los problemas de sesgo son los resultados comparados de la encuesta nacional de empleo y de la nueva encuesta nacional de empleo del INE, que durante el año 2009 fueron tomadas en forma conjunta (a nivel de experiencia piloto). Ello en tanto la primera encuesta utiliza el marco muestral de secciones mientras que la segunda usa el marco muestral de manzanas, aún cuando también hay cambios en los cuestionarios respectivos. Los resultados comparados entre ambas encuestas fueron solicitados al INE y de la información recibida se desprende que la tasa de rechazo en la antigua encuesta de empleo es alrededor de diez puntos más elevada, lo cual podría ser reflejo del agotamiento de la muestra y consiguiente negativa a seguir respondiendo encuestas por parte de los hogares ya entrevistados. No obstante, la diferencia en la tasa de rechazo es mayor en el área rural, *donde no ha habido cambio en el marco muestral*, apuntado a otro tipo de causa detrás del diferencial en la tasa de rechazo.

⁴ No hay información de la variables para estas viviendas.

21. Asimismo, este Panel tuvo acceso a alguna información relativa al trabajo de campo en las encuestas Casen 2006 y Casen 2009, no observándose diferencias de importancia en la proporción de unidades no elegibles en la muestra y la tasa de rechazo a la entrevista. En la encuesta Casen 2006, la proporción de la muestra seleccionada que era no elegible era aproximadamente 5% y de los contactos realizados en viviendas, el rechazo era aproximadamente de 5%. Para la encuesta Casen 2009, y según la información disponible para este Panel, cada una de estas proporciones habría subido entre 1 y 2 puntos porcentuales. No hay por tanto evidencia de un deterioro de la muestra entre la Casen 2006 y 2009 por agotamiento del marco muestral utilizado.

22. En conclusión, de los antecedentes disponibles no hay evidencia clara ni sistemática que los resultados de la Casen 2009 tengan un sesgo de importancia originado en la pérdida de representatividad del marco muestral 2009. Los sesgos asociados serían más bien pequeños, y posiblemente inferiores a otros sesgos que están presentes usualmente en encuestas de hogares como la Casen. Este es el caso del subreporte de los ingresos y de la utilización de un informante por hogar cuando no cuenta con buena información de otros miembros del hogar, En el caso específico de la versión 2009 hay que agregar los efectos que producen en las variables de empleo e ingreso el desplazamiento temporal en la fecha de la recolección de los datos.