

## 5. Procesamiento

El procesamiento de la base de datos derivada del censo implicó la ejecución de un proceso de validación de la información con la finalidad de garantizar, tanto su congruencia lógica, como la completitud e integridad de los datos asociados a las preguntas de los módulos del instrumento de captación. Para ello, se aplicó una serie de criterios que se diseñaron tomando fielmente el contenido y la estructura de los cuestionarios.<sup>5</sup>

En términos generales, la validación se llevó a cabo para corregir las inconsistencias de la información derivadas básicamente por la existencia de errores como omisión o falta de respuesta, multirrespuesta, valores inadmisibles o fuera de rango, falta de atención en los pases de preguntas y, también, incongruencias entre respuestas.

Para estos efectos, el proceso se ejecutó en tres grandes fases: la preparación del proceso de validación, la revisión y pruebas del funcionamiento de los vectores de validación y la ejecución de la validación y generación de la base de datos definitiva.

<sup>5</sup> Es importante asentar que en los casos en que los informantes no contaron con datos o elementos para mantener la consistencia en todos los niveles de desagregación que se solicitaron, se estableció una categoría no considerada en el cuestionario denominada "No especificado".

### 5.1 Fase de preparación

Antes de validar la base de datos, se realizó un conjunto de actividades que no fueron ejecutadas necesariamente de manera secuencial, pero que resultaron indispensables para garantizar el proceso de validación. En seguida se enuncia brevemente cada una de ellas.

#### Clasificación de preguntas por tipo

Se realizó una clasificación de las preguntas del cuestionario de acuerdo con el tipo de respuesta que se solicita y se identificaron las relaciones que guardan con otras preguntas. Se identificaron 22 tipos.

Esto permitió contar con elementos precisos para definir los criterios de validación tomando en consideración las necesidades derivadas de cada tipo de pregunta y asegurar que todas ellas también contaran con criterios definidos de manera particular. A su vez, la agrupación de las preguntas permitió aplicar los mismos criterios a todas las incluidas en la clasificación.

En seguida se describió brevemente cada tipo de pregunta y se establecieron los criterios mínimos de validación por cada uno de ellos. La siguiente tabla muestra los resultados.

#### Tipología de preguntas

Cuadro 8

Núm.	Tipo	Descripción	Validación mínima
1	Selección única	Hace referencia a las preguntas que involucran la selección de un solo código.	1) La variable que guarda el código de selección, no debe ser nulo.
2	Selección múltiple	Hace referencia a las preguntas que involucran la selección de uno o más códigos.	1) Verificar que la variable que guarda los códigos de selección, al menos tenga un valor. 2) Si la variable contiene el código referente a "No se sabe", no debe existir otro código para la entidad de referencia. 3) Si la variable contiene el código referente a "Ninguno" o "No tiene", no debe existir otro código para la entidad de referencia.

(Continúa)

**Tipología de preguntas**

Núm.	Tipo	Descripción	Validación mínima
3	Selección múltiple condicionada	Hace referencia a las preguntas que involucran la selección de uno o más códigos, pero existen conjuntos de valores que no pueden seleccionarse a la vez.	1) Verificar que la variable que guarda los códigos de selección, al menos tenga un valor. 2) Verificar que solo exista un código de selección, de los valores que integran el conjunto que son excluyentes. 3) Si la variable contiene el código referente a "No se sabe", no debe existir otro código para la entidad de referencia. 4) Si la variable contiene el código referente a "Ninguno" o "No tiene", no debe existir otro código para la entidad de referencia.
4	Selección con salto	Hace referencia a las preguntas que involucran la selección de un solo código y de acuerdo a éste, se continúa respondiendo o se pasa a la pregunta indicada en la que se debe continuar.	1) La variable que guarda el código de selección, no debe ser nulo. 2) Si la variable tiene un código que implica salto de pregunta, las variables afectadas por el salto de pregunta deben tener el valor que corresponda según sea el caso: -Si la respuesta es No el valor debe ser "NA" (-2). -Si la respuesta es En proceso... el valor debe ser "NA" (-2). - Si la respuesta es No se sabe el valor debe ser "NS" (-1).
5	Selección/abierta	Hace referencia a las preguntas que involucran la selección de uno o varios códigos, combinada con recuadros para la introducción de cantidades.	1) Debe haber una correspondencia entre la opción seleccionada y la variable que guarda una cantidad.
6	Tabla simple	Hace referencia a las preguntas conformadas por una tabla donde se tiene datos registrados de manera independiente, sin total y sin no aplica, y con un catálogo asociado y ninguna relación con otra pregunta.	1) Solo se corrobora que la tabla no tenga algún registro nulo.
7	Tabla suma horizontal	Hace referencia a las preguntas conformadas por una tabla donde una columna es el total.	1) Para cada registro, el valor de la variable que guarda el total debe ser igual a la suma de sus desgloses correspondientes (sumas horizontales). 2) Generalmente existe una pregunta que capta el valor total que se desglosa en una tabla, por lo tanto, la suma vertical del total de la tabla debe ser igual al de la pregunta relacionada. 3) Verificar que el valor de la variable en la categoría de total, sea igual a la suma de los valores correspondientes al resto de las categorías (suma vertical).

(Continúa)

Tipología de preguntas

Cuadro 8

Núm.	Tipo	Descripción	Validación mínima
8	Tabla subtotal	Hace referencia a las preguntas conformadas por una tabla con la columna de total y sus respectivas columnas de desagregación (sin la columna NA y sin saltos).	1) Para cada registro, el valor de la variable que guarda el total debe ser igual a la suma de sus desgloses correspondientes (sumas horizontales). 2) Si se presenta el caso, el valor de la variable que guarda un subtotal debe ser igual a la suma de sus desgloses correspondientes (sumas horizontales). 3) Generalmente existe una pregunta que capta el valor total que se desglosa en una tabla, por lo tanto, la suma vertical del total de la tabla debe ser igual al de la pregunta relacionada. 4) Verificar que el valor de la variable en la categoría de total, sea igual a la suma de los valores correspondientes al resto de las categorías (suma vertical).
9	Tabla Delitos FC	Hace referencia a las preguntas conformadas por una tabla con la columna de total y sus respectivas columnas de desagregación (sin la columna NA y sin saltos).	1) Para cada delito, el valor de la variable que guarda el total debe ser igual a la suma de sus desgloses correspondientes (sumas horizontales). 2) Si se presenta el caso, el valor de la variable que guarda un subtotal debe ser igual a la suma de sus desgloses correspondientes (sumas horizontales). 3) Generalmente existe una pregunta que capta el valor total que se desglosa en una tabla, por lo tanto, la suma vertical del total de la tabla debe ser igual al de la pregunta relacionada.  *Nota: considerar que esta tabla permite -2 (no aplica) en sus desagregados.
10	Tabla con total por bloque respectivo	Hace referencia a las preguntas conformadas por una tabla con la columna de total y sus respectivas columnas de desagregación (sin la columna NA y sin saltos).	1) El total de delitos reportados en el Bien Jurídico Afectado debe ser igual a la suma de los delitos que correspondan a ese bien jurídico (suma vertical). 2) Para cada delito, el valor de la variable que guarda el total debe ser igual a la suma de sus desgloses correspondientes (sumas horizontales). 3) Si se presenta el caso, el valor de la variable que guarda un subtotal debe ser igual a la suma de sus desgloses correspondientes (sumas horizontales). 4) Generalmente existe una pregunta que capta el valor total que se desglosa en una tabla, por lo tanto, la suma vertical del total de la tabla debe ser igual al de la pregunta relacionada.
11	Tablas especiales (Delitos FC y FF, Infracciones y Siniestros)	Hace referencia a las tablas especiales que se manejan en los censos de Seguridad y Procuración de Justicia, los cuales manejan diferentes códigos, para diseño y validación.	Las validaciones usuales para estas tablas son la horizontal, la vertical y la de dato a dato, según la necesidad y el diseño de la tabla.

(Continúa)

Tipología de preguntas

Núm.	Tipo	Descripción	Validación mínima
12	Tabla NA	Hace referencia a las preguntas que involucran una tabla cuya última columna sea No Aplica.	<p>1) Si para algún registro, el valor de la variable de no aplica es "1", el valor del resto de las variables de esa pregunta, para ese registro en particular, debe ser nulo, por lo que nosotros debemos llenar ese registro nulo con "NA(-2)"</p> <p>2) Generalmente estas preguntas están relacionadas con otras de similar estructura, por lo tanto, en todas ellas debe ser consistente el valor de la variable de no aplica igual a "1", para el registro correspondiente y el valor del resto de las variables igual a "NA (-2)".</p> <p>3) Si es el caso, el valor de la variable que guarda el total debe ser igual a la suma de sus desgloses correspondientes (sumas horizontales).</p> <p>4) Si se presenta el caso, el valor de la variable que guarda un subtotal debe ser igual a la suma de sus desgloses correspondientes (sumas horizontales).</p>
13	Tabla Municipios	Hace referencia a las preguntas que involucran una tabla cuyas filas representan los delitos y las columnas los municipios de la entidad correspondiente.	<p>1) En la base de datos solo se traerá el desagregado de delitos por municipio, no habrá algún campo que guarde el valor total, por lo que solo se tendrá que validar que la suma de delitos por municipio sea igual al total de la suma de delitos de la pregunta a la que esté relacionada.</p> <p>2) El valor que se encuentre en la categoría "Resto de los municipios" debe ser igual a la suma de los valores registrados en la tabla complementaria (complemento).</p>
14	Tabla suma vertical	Hace referencia a las preguntas que involucran una tabla cuyas columnas no son un total, sino que sus filas son el desagregado de otra pregunta y/o tiene un total en su desagregado.	<p>1) Generalmente existe una pregunta que capta el valor total que se desglosa en una tabla, por lo tanto, la suma vertical del total de la tabla debe ser igual al de la pregunta relacionada. En otros casos la suma de cada columna debe ser mayor o menor al valor total al que hace referencia en otra pregunta. Para los casos donde existe un total dentro de la misma tabla, en el desagregado, se tendrá que hacer una validación del total con el desagregado.</p>
15	Tabla resumen	Hace referencia a las preguntas que involucran una tabla cuyas categorías son un agregado o resumen de otra pregunta.	<p>1) Los valores de la tabla resumen deben coincidir con la frecuencia de los datos registrados en la tabla desagregada.</p>
16	Tabla Sí/NO	Hace referencia a las preguntas que involucran una tabla donde en alguna(s) de las columnas aparece Sí/NO y cuya respuesta afecta el valor de otra(s) columna(s).	<p>1) Los valores de la columna (sí/no) solo pueden ser 1,2 o 9. No deben existir nulos.</p> <p>2) Si el valor de la columna (sí/no) es 2 el resto de las columnas para ese registro deben ser nulos, por lo que nosotros debemos llenar ese registro nulo con el valor "NA -2".</p>

(Continúa)

**Tipología de preguntas**

Cuadro 8

Núm.	Tipo	Descripción	Validación mínima
16	Tabla SÍ/NO	Hace referencia a las preguntas que involucran una tabla donde en alguna(s) de las columnas aparece SÍ/NO y cuya respuesta afecta el valor de otra(s) columna(s).	3) Si el valor de la columna (sí/no) es 9 el resto de las columnas para ese registro deben ser nulos, por lo que nosotros debemos llenar ese registro nulo con el valor "NS -1". 4) Si el valor de la columna (sí/no) es 1 el resto de las columnas para ese registro deben tener valores (de acuerdo con los rangos validos en cada caso).
17	Tabla de registros	Hace referencia a las preguntas que involucran una tabla, pero que tienen la particularidad de no contar con un catálogo definido para la variable que se encuentra en columnas.	1) Verificar que el llenado sea consecutivo y que no existan duplicidades. 2) Si se presenta el caso en que esta tabla esté relacionada con otras verificar que se mantenga la correspondencia por registro.
18	Campo independiente (número o texto)	Hace referencia a las preguntas con recuadros para la introducción de cantidades o texto, pero que no constituyen una tabla.	1) El valor de la variable no debe ser nulo.
19	Suma horizontal simple	Hace referencia a las preguntas con la clasificación de abiertas, donde un recuadro representa el total y los restantes su desglose.	1) El valor de la variable que guarda el total debe ser igual a la suma de las variables que corresponden al desglose. 2) Si se presenta el caso, el valor de la variable que guarda un subtotal debe ser igual a la suma de sus desgloses correspondientes. 3) No debe haber variables con valor nulo, a excepción de las que guardan los valores no especificados; para llenar los nulos que pudieran encontrarse, se deben relacionar el desglose y el total para deducir si se les asigna valor "0" o "NS(-1)" al valor nulo.
20	Suma vertical simple	Hace referencia a las preguntas que fueron estructuradas en BD como una variable que almacena en total y su respectivo desglose.	1) Verificar que el valor de la variable en la categoría de total, sea igual a la suma de los valores correspondientes al resto de las categorías (suma vertical). 2) No debe haber valor nulo en la variable, a excepción de la categoría que guarda el no especificado.
21	Selección de catálogo	Hace referencia a tablas en las cuales en sus columnas se tiene que seleccionar el valor de un catálogo.	1) Verificar que en las columnas que vengan relacionadas a un catálogo, se seleccione alguna opción, de lo contrario poner el valor estándar al No se sabe.
22	Tabla fecha	Hace referencia a tablas en las cuales una columna involucra una fecha y la otra columna trata la existencia o no de leyes en el tema.	1) Las columnas son excluyentes, es decir, solo una puede tener dato dentro de sus rangos válidos.

A continuación se identificaron aquellas preguntas que correspondían a cada uno de los 22 tipos definidos en

el cuadro anterior, para los cuatro módulos del cuestionario. En el siguiente cuadro se muestran los resultados.

**Tipo de preguntas por módulo**

Cuadro 9

Núm.	Tipo	Módulo 1	Módulo 2	Módulo 3
1.	Selección Única	----	----	----
2.	Selección Múltiple	20.1, 22	2, 4, 7	2
3.	Selección Múltiple condicionada	16.1, 19, 19.2	----	----
4.	Selección con salto	16, 17, 18	----	----
5.	Selección/abierta	----	----	----
6.	Tabla simple	----	3.1	----
7.	Tabla suma horizontal	4, 5, 6, 7, 8, 9, 12, 13, 14, 29.2	----	----
8.	Tabla subtotal	----	----	----
9.	Tabla Delitos FC	----	----	----
10.	Tabla con total por bloque respectivo	----	----	----
11.	Tablas especiales (Delitos FC y FF, Infracciones y Siniestros)	----	----	----
12.	Tabla NA	15	----	1, 3, 5
13.	Tabla Municipios		----	----
14.	Tabla suma vertical	24, 25, 26, 27, 29.1, 31	1	----
15.	Tabla resumen	----	----	----
16.	Tabla SÍ/NO	16.2, 17.1, 18.1, 20, 28, 30, 32	6, 8, 8.1	----
17.	Tabla de registros	----	----	4
18.	Campo independiente (número o texto)	10, 11, 19.1, 21, 23	3, 5	----
19.	Suma horizontal simple	----	----	----
20.	Suma vertical simple	33	----	----
21.	Selección de catálogo	1, 2	----	----
22.	Tabla fecha	----	----	----

## Modelado de la base de datos

Para establecer y describir la forma en que se organizaron y ordenaron los datos para su almacenamiento de forma que la información se encuentre disponible sin redundancia y de manera entendible, se creó un modelo entidad-relación el cual define de manera precisa las relaciones entre las entidades u objetos para los que se almacenan datos y sus atributos, es decir, las características de la entidad cuyos valores se encuentran en los campos de las tablas de la base de datos.

## Migración de la base de datos de captura

Con la finalidad de empatar la base de datos de captura con el modelo de la definitiva, se efectuó un proceso de mapeo entre la base de datos preliminar (base de datos de captura) y el modelo definitivo de base de datos, para encontrar la relación entre los datos que requiere cada campo de éste último y el campo en que se encuentra en la base de datos preliminar.

Asimismo, se creó un esquema de consultas (vistas) en SQL para acceder a los datos de la base preliminar y ordenarlos en tablas y campos siguiendo la estructura de la base de datos definitiva.

Enseguida se llevó a cabo un proceso mediante el cual se insertaron los datos de la base preliminar, a través de las consultas o vistas, a la base de desarrollo. Para ello se construyó una serie de paquetes que generaron la estructura de las tablas de la base de datos de acuerdo con los tipos de dato permitidos. De manera paralela se creó una tabla especial para almacenar casos en que los datos no se encontraban en los rangos establecidos. Esto permitió detectar errores en los datos de la base de captura y corregirlos. En este proceso se realizó la conversión algunos datos alfanuméricos de la base de datos preliminar como NS y los NA a datos numéricos para insertarlos a la estructura de la base de datos definitiva, convirtiendo los NS a -1 y los NA a -2.

## Definición de funciones

Se definieron 28 funciones para contar con un lenguaje común formado por un conjunto de símbolos y variables para expresar las operaciones y/o procesos que debe llevar a cabo la computadora al aplicar cada uno de los vectores de validación.

En el siguiente cuadro se muestran estas funciones, su descripción, su clasificación por tipo, de acuerdo con el uso que se le dio durante el proceso, así como la cantidad mínima de argumentos que debió considerar y su valor de retorno.

## Clasificación de funciones

Cuadro 10

Núm.	Función	Descripción	Tipo	Uso
1.	DIFERENTEDE	Regresa TRUE si el primer valor es distinto al segundo valor definido.	-Condición -Vector -Auxiliar	<b>DIFERENTEDE</b> (variable_valor, variable_valor) <b>DIFERENTEDE</b> (valor, valor) <b>DIFERENTEDE</b> (valor, valor)
2.	EXISTE	Verifica si existe el valor en la columna dada como parámetro. Regresa TRUE o FALSE.	-Vector -Auxiliar	<b>EXISTE</b> (variable, valor_a_buscar)
3.	EXISTEO	Verifica si existe alguno de los valores en la columna dada como parámetro. Regresa TRUE o FALSE.	-Vector -Auxiliar	<b>EXISTEO</b> (variable, lista_de_valores)
4.	EXISTEY	Verifica si existen todos los valores en la columna dada como parámetro. Regresa TRUE o FALSE.	-Vector -Auxiliar	<b>EXISTEY</b> (variable, lista_de_valores)

(Continúa)

**Clasificación de funciones**

Cuadro 10

Núm.	Función	Descripción	Tipo	Uso
5.	IGUAL	Regresa TRUE si el valor de una variable es igual a un valor definido o FALSE si no lo es.	-Condición de entrada -Vector -Auxiliar	<b>IGUAL</b> (variable_valor, variable_valor) <b>IGUAL</b> (valor, valor) <b>IGUAL</b> (valor, valor)
6.	MAYOR	Regresa TRUE si el valor de una variable es mayor a un valor definido o FALSE si no lo es.	-Condición de entrada -Vector -Auxiliar	<b>MAYOR</b> (variable_valor, variable_valor ) <b>MAYOR</b> (valor, valor) <b>MAYOR</b> (valor, valor)
7.	MAYORIGUAL	Regresa TRUE si el primer valor es mayor o igual al segundo valor definido.	-Condición de entrada -Vector -Auxiliar	<b>MAYORIGUAL</b> (variable_valor, variable_valor) <b>MAYORIGUAL</b> (valor, valor) <b>MAYORIGUAL</b> (valor, valor)
8.	MENOR	Regresa TRUE si el primer valor es menor al segundo.	-Condición de entrada -Vector -Auxiliar	<b>MENOR</b> (variable_valor, variable_valor) <b>MENOR</b> (valor, valor) <b>MENOR</b> (valor, valor)
9.	MENORIGUAL	Regresa TRUE si el primer valor es menor o igual al segundo valor definido.	-Condición de entrada -Vector -Auxiliar	<b>MENORIGUAL</b> (variable_valor, variable_valor) <b>MENORIGUAL</b> (varlor, valor) <b>MENORIGUAL</b> (varlor, valor)
10.	BUSCARH	Busca horizontalmente el valor indicado en el rango de variable(s) dentro del argumento, indicando los índices de las categorías en las que se desea realizar la búsqueda. Regresa una cadena que representa las columnas donde se encontró el valor buscado y null o vacío cuando no encuentra nada.	-Auxiliar	<b>BUSCARH</b> (valor_a_buscar, var_cat1, indices_rest1, var_cat2, indices_rest2, lista_de_variables)
11.	BUSCARV	Busca verticalmente el valor indicado en la variable dentro del rango de valores dado por la variable categórica, y regresa una cadena que representa la variable categórica con los índices donde se encontró el valor buscado y null o vacío cuando no encuentra nada.	-Auxiliar	<b>BUSCARV</b> (valor_a_buscar, var_cat1, indices_rest1, var_cat2, indices_rest2, variable)
12.	CONTARREG	Cuenta el número de registros existentes en una tabla.	-Auxiliar	<b>CONTARREG</b> (variable)

(Continúa)

**Clasificación de funciones**

Cuadro 10

Núm.	Función	Descripción	Tipo	Uso
13.	CONTARH	Para cada registro, cuenta el número de Vectores que encontró el valor del parámetro, en las variables indicadas.	-Auxiliar	<b>CONTARH</b> (valor_a_buscar, var_cat1, indice_res1, var_cat2, indice_res2, lista_de_variables)
14.	CONTARV	Cuenta el número de registros que contienen el valor del parámetro en la variable indicada, estableciendo una posición.	-Auxiliar	<b>CONTARV</b> (valor_a_buscar, var_cat1, indice_rest1, var_cat2, indice_rest2, variable)
15.	BUSCAREP	Función que regresa un valor lógico si hay valores repetidos en las columnas indicadas. Cuando no se encuentre valor repetido regresa.  FALSE. TRUE en otro caso.  Nota: la lista de variables donde se va a buscar el valor debe ser en el orden definido en la tabla.  Observación: no se consideran los valores <i>Null</i> .	-Auxiliar	<b>REPETIR</b> (var_cat1, indice_rest1, var_cat2, indice_rest2, lista_de_variables)
16.	RESTA	Regresa la diferencia de los valores de las variables indicadas.	-Auxiliar	<b>RESTA</b> (valor, valor)
17.	REVSALTO	Función que nos permitirá identificar la consistencia del salto de pregunta.  Si el primer parámetro es TRUE, regresa TRUE si sólo hay <i>Null</i> y valor válido; y FALSE si encuentra un dato no válido dentro de las variables que se definen en el argumento.  Si el primer parámetro es FALSE, regresa TRUE si sólo hay valor válido; y FALSE si encuentra un dato no válido dentro de las variables que se definen en el argumento, incluso <i>Null</i> .  Si el primer argumento lleva un asterisco, indicaría al criterio que debe usar la tabla principal para obtener las claves de las entidades.	-Auxiliar	<b>REVSALTO</b> (valor_logico, valor, lista_de_variables)

(Continúa)

**Clasificación de funciones**

Cuadro 10

Núm.	Función	Descripción	Tipo	Uso
18.	REVSALTOMOD	<p>Función que nos permitirá identificar la consistencia del salto de pregunta, en el caso donde se salte todo el módulo, excepto en la primera pregunta.</p> <p>Si el primer parámetro es TRUE, regresa TRUE si sólo hay <i>Null</i> y valor válido; y FALSE si encuentra un dato no válido dentro de las variables del módulo.</p> <p>Si el primer parámetro es FALSE, regresa TRUE si sólo hay valor válido; y FALSE si encuentra un dato no válido dentro de las variables del módulo, incluso <i>Null</i>.</p> <p>Nota: esta función no está soportada tal cual, se implementó usando REVSALTO definiendo de manera automatizada todas las variables involucradas en el módulo del censo, excepto la indicada como parámetro.</p>	-Auxiliar	<b>REVSALTOMOD</b> (valor_logico, valor, variable)
19.	SUMA	Regresa la adición de los valores dados como parámetros.	-Auxiliar	<b>SUMA</b> (valor, valor)
20.	SUMAH	<p>Suma todos los valores en el rango de variable(s) dentro del argumento, indicándole los índices de las categorías que debe considerar.</p> <p>Observación: solo se consideran los valores mayores o iguales a CERO, excluye números negativos y <i>Null</i>.</p>	-Auxiliar	<b>SUMAH</b> (var_cat1, indice_rest1, var_cat2, indice_rest2, lista_variables)
21.	SUMAV	<p>Suma todos los valores de la variable dentro del argumento, indicándole el índice en el que inicia y termina.</p> <p>Observación: solo se consideran los valores mayores o iguales a CERO, excluye números negativos y <i>Null</i>.</p>	-Auxiliar	<b>SUMAV</b> (var_cat 1, indice_rest1, var_cat2, indice_rest2, lista_de_variables)
22.	VALOR	Regresa el valor numérico de la variable indicada, dándole la posición del registro a considerar.	-Auxiliar	<b>VALOR</b> (variable, var_cat1, indice_rest1, var_cat2, indice_rest2)

(Continúa)

**Clasificación de funciones**

Cuadro 10

Núm.	Función	Descripción	Tipo	Uso
23.	VALORCAD	Regresa la cadena de la variable indicada, dándole la posición del registro a considerar.	-Auxiliar	<b>VALORCAD</b> (variable, var_cat1, indice_rest1, var_cat2, indice_rest2)
24.	ASIGNAVAL	Cambia el valor de la(s) variable(s), en una posición determinada, utilizando más de índice de posición.  Cuando algún parámetro de índice de restricción (indice_rest1, indice_rest2) lleva al final un asterisco, esta función asigna el valor a la(s) variable(s) de manera vertical.	-Tratamiento	<b>ASIGNAVAL</b> (valor, var_cat1, indice_rest1, var_cat2, indice_rest2, lista_de_variables)
25.	IMPOSIBLE	Constante para un caso imposible.	-Tratamiento	<b>IMPOSIBLE</b>
26.	REVISION_DE_CASO	Constante para revisión de caso.	-Tratamiento	<b>REVISION_DE_CASO</b>
27.	SIN_CAMBIOS	Constante para indicar no hacer nada.	-Tratamiento	<b>SIN_CAMBIOS</b>
28	CONTAR	Es usada como condición de entrada, cuenta los valores que el usuario desea buscar en un número de variables. Si al contar los valores estos equivalen a la cantidad de veces que el usuario desea encontrarlo, entonces se aplica la función de comparación establecida.  Nota: 'comparación' puede tomar los siguientes valores que equivalen a los operadores de comparación usuales (IGUAL, DIFERENTE, MAYOR, MENOR, MAYORIGUAL, MENORIGUAL).	-Condición de entrada	<b>CONTAR</b> (comparacion, valor_a_contar, valor_a_comparar, lista_variables)

**Definición de vectores**

Se definió un conjunto de tratamientos (vectores de validación) para corregir las posibles inconsistencias en la base de datos preliminar, considerando las distintas combinaciones de valores que se pudieran encontrar dentro de una misma pregunta y entre las preguntas. Los criterios de validación son necesarios para que todos los casos se resuelvan de la misma forma y así los datos sean consistentes.

En total se crearon diez escenarios; siete para el primer módulo, dos para el segundo y uno para el tercero. Asimismo, se desarrollaron 970 vectores de validación: 889 para el primer módulo, 50 para el segundo, y 31 para el tercero.

En el cuadro siguiente se muestra la cantidad de escenarios definidos para cada módulo, su nomenclatura y el número de vectores de validación desarrollados para cada uno de ellos.

**Vectores por módulo**

Cuadro 11

Módulo	Número de escenario por módulo	Escenario	Número de criterios por escenario	Total de criterios
1	1	VT_GOF_M1_S1	16	889
	2	VT_GOF_M1_S2	213	
	3	VT_GOF_M1_S3	259	
	4	VT_GOF_M1_S4	62	
	5	VT_GOF_M1_S5	12	
	6	VT_GOF_M1_S6	312	
	7	VT_GOF_M1_S7	15	
2	1	VT_GOF_M2_S1	37	50
	2	VT_GOF_M2_S2	13	
3	1	VT_GOF_M3_S1	31	31
<b>TOTAL DE CRITERIOS:</b>				<b>970</b>

**Vectores por módulo**

Con el fin de estandarizar el proceso para definir los vectores de validación y su programación, se creó una aplicación informática denominada Entorno de Desarrollo para Criterios de Validación Exhaustiva, mediante la cual, por un lado, se realizó la definición de vectores de validación, en la que se tomaron en cuenta todas las posibles combinaciones que se pueden dar en la pregunta o preguntas que se están validando, y se definió un tratamiento para cada una de esas combinaciones.

Por otro lado, esta aplicación permitió la programación sistematizada de los vectores de validación.

A partir de las funciones definidas se crearon sus correspondientes códigos en lenguaje PL/SQL, es decir fragmentos de código predefinidos, que se generan de manera automática a partir de la función y las variables definidas en la aplicación para cada vector.

**Definición de criterios de validación**

Para poder validar la información se definieron dos tipos de criterios para aplicarlos de manera homogénea a lo largo del proceso. Por un lado, un conjunto de criterios básicos que permitieron establecer relaciones primarias entre respuestas que involucraron los caracteres alfanuméricos NS (no contó con elementos para responder) y NA (no le aplica). Esta definición quedó como sigue:

**Criterios básicos**

- 1 + -1 = -1 Sumatoria de NS da como resultado NS
- 1 + -2 = -1 Sumatoria de NS y NA da como resultado NS
- 1 + X = X Sumatoria de NS y un número positivo da como resultado el número positivo
- 0 + -1 = -1 Sumatoria de cero y NS da como resultado NS
- 2 + -2 = -2 Sumatoria de NA da como resultado NA

**Criterios específicos**

Enseguida se crearon criterios específicos para situaciones particulares que se presentaron a lo largo del proceso, mismos que se muestran a continuación.

## I. Para validación de sumas

### Total Nulo

- **Caso 1.** Total nulo y datos de sumatoria completos. Se coloca el valor de la suma en el total.

	Total	V1	V2	V3	V4	NE
Inconsistencia		20	10	10	10	
Solución	50	20	10	10	10	

- **Caso 2.** Total nulo (*Null*) y todos los datos de sumatoria nulos. Se imputan todos los nulos como -1.

	Total	V1	V2	V3	V4	NE
Inconsistencia						
Solución	-1	-1	-1	-1	-1	

- **Caso 3.** Total nulo y algunos datos de sumatoria nulos. Los nulos se vuelven ceros y siguiendo con el criterio del caso 1 el total se imputa con la sumatoria.

	Total	V1	V2	V3	V4	NE
Inconsistencia		5	5		5	
Solución	15	5	5	0	5	

- **Caso 4.** Total nulo y algunos datos de sumatoria nulos y otros -1. Se pone -1 el total y los nulos se imputan como -1.

	Total	V1	V2	V3	V4	NE
Inconsistencia			-1	-1	-1	
Solución	-1	-1	-1	-1	-1	

- **Caso 5.** Total nulo y todos los datos de la sumatoria -1. Se pone -1 el total.

	Total	V1	V2	V3	V4	NE
Inconsistencia		-1	-1	-1	-1	
Solución	-1	-1	-1	-1	-1	

### Total -1 (NS)

- **Caso 6.** Total -1 y todos los datos de la sumatoria ceros. Se imputa como cero el total.

	Total	V1	V2	V3	V4	NE
Inconsistencia	-1	0	0	0	0	
Solución	0	0	0	0	0	

**Total ceros**

- **Caso 7.** Total cero y al menos un -1 en la sumatoria. Se imputa como -1 el total.

	V1	V2	V3	V4	NE	Total
Inconsistencia	0	-1	0	0		0
Solución	0	-1	0	0		-1

- **Caso 8.** Total cero y al menos un nulo en la sumatoria. Se imputan los nulos con cero.

	Total	V1	V2	V3	V4	NE
Inconsistencia	0	0		0		
Solución	0	0	0	0	0	

**Sumatoria igual al total**

- **Caso 9.** Sumatoria igual al total y uno o más -1. Los -1 se imputan como ceros.

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	15	15	-1	-1	
Solución	30	15	15	0	0	

- **Caso 10.** Sumatoria igual al total y uno o más espacios nulos. Se imputan los nulos como ceros.

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	15	15			
Solución	30	15	15	0	0	

**Sumatoria menor al total**

- **Caso 11.1.** Sumatoria menor al total y un -1. Se imputa el valor del -1 por la diferencia.

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	10	5	-1	5	
Solución	30	10	5	10	5	

- **Caso 11.2.** Sumatoria menor al total y un nulo. Se imputa el valor del nulo por la diferencia.

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	10	5		5	
Solución	30	10	5	10	5	

• **Caso 12.** Sumatoria menor al total y más de un -1. La diferencia se va al *No Especificado* (NE) y los -1 se mantienen.

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	10	-1	-1	5	
Solución	30	10	-1	-1	5	15

• **Caso 13.1** Sumatoria menor al total y no -1 ni nulo. Si sólo es un desglose se da prioridad a la desagregación y se imputa el total.

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	5	5	5	5	
Solución	20	5	5	5	5	

• **Caso 13.2** Sumatoria menor al total y no -1 ni nulo. Si tiene relación con varias tablas se da prioridad al total que sirve de base para la comparación y se calcula el *No Especificado* (NE) donde corresponda.

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	5	5	5	5	
Solución	30	5	5	5	5	10

• **Caso 13.3** Sumatoria menor al total y no -1 ni nulo. Si el total de la tabla base es modificado por un valor mayor se borra el no especificado para que pueda ser calculado nuevamente.

Tabla base	Total	V1	V2	V3	V4	NE
Inconsistencia	30	5	5	5	5	10
Solución	40	5	5	5	5	Se elimina

Tabla desglose	Total	V1	V2	V3	V4	NE
Inconsistencia	40	15	5	15	5	

• **Caso 14.** Sumatoria menor al total y uno o más nulos. Se imputan los nulos como -1 y la diferencia entra en el *No Especificado* (NE).

Tabla base	Total	V1	V2	V3	V4	NE
Inconsistencia	30	5	5			
Solución	30	5	5	-1	-1	20

### Sumatoria mayor al total

• **Caso 15.** Sumatoria mayor al total y no hay nulos ni -1. Se imputa el total, tomando como válida la sumatoria.

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	20	10	10	0	
Solución	40	20	10	10	0	

- **Caso 16.** Sumatoria mayor al total y uno o más -1. Se imputa el total con la sumatoria y los -1 se vuelven ceros.

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	10	20	-1	10	
Solución	40	10	20	0	10	

- **Caso 17.** Sumatoria mayor al total y uno o más nulos. Los nulos se vuelven cero y siguiendo con el criterio del caso 16 el total se imputa con la sumatoria.

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	20		10	10	
Solución	40	20	0	10	10	

### Desglose ceros

- **Caso 18.** Existe un total y el desglose son ceros. Se imputan los ceros como -1 y se pone el valor total en el *No Especificado* (NE).

	Total	V1	V2	V3	V4	NE
Inconsistencia	30	0	0	0	0	
Solución	30	-1	-1	-1	-1	30

## II. Criterios para llenado de preguntas de opción múltiple

- **Caso 1.** Si la opción marcada es **No** las demás opciones deben venir como -2

ENTIDAD	1	2	3	4 NO	5 No sabe
1				1	
1	-2	-2	-2	1	-2

- **Caso 2.** Si la opción marcada es **No sabe** las demás opciones deben venir como -1

ENTIDAD	1	2	3	4 NO	5 No sabe
1					1
1	-1	-1	-1	-1	1

## III. Criterios para llenado de preguntas con tablas sí/no

- **Caso 1.** Si la opción registrada es **2. No** las demás opciones deben venir como -2

ENTIDAD	1. Cuenta con ...	2	3	4	5
1	2				
1	2	-2	-2	-2	-2

- **Caso 2.** Si la opción registrada es **9. No sabe** las demás opciones deben venir como -1

ENTIDAD	1. Cuenta con ...	2	3	4	5
1	9				
1	9	-1	-1	-1	-1

#### IV. Criterios para salto de pregunta

- **Caso 1.** Si la respuesta es No, las preguntas que se saltan deben llenarse con -2. Excepto en las tablas NA, la columna NA debe quedar nula.
- **Caso 2.** Si la respuesta es No se sabe, las preguntas que se saltan se deben llenar con (-1) para el caso en que no exista un catálogo que contenga la opción no sabe, y con el número que corresponda en los casos en que sí exista.

#### V. Comparación de totales con valor cero y/o (-1) en dos tablas

- **Caso 1.** El total de la tabla base es 0 y el total de la tabla desglose es -1. Se imputan los valores de la tabla desglose con ceros.

Caso 1 Tabla base	Total	V1	V2	V3
Datos	0	0	0	0

Caso 1 Tabla desglose	Total	V1	V2	V3
Inconsistencia	-1	-1	0	-1
Solución	0	0	0	0

- **Caso 2.** El total de la tabla base es -1 y el total de la tabla desglose es 0. Se imputan los valores de la tabla desglose con -1.

Caso 2 Tabla base	Total	V1	V2	V3
Datos	-1	0	-1	0

Caso 2 Tabla desglose	Total	V1	V2	V3
Inconsistencia	0	0	0	0
Solución	-1	-1	-1	-1

#### VI. Criterio para preguntas abiertas

- **Caso único.** Si el informante no proporcionó información en una pregunta abierta y no existe comentario para la pregunta el campo deberá llenarse con -1. De lo contrario, si existe el comentario se manda a una revisión de caso.

## 5.2 Fase de pruebas

Una vez programados los vectores se corrieron sobre la base de datos ya migrada en el esquema de desarrollo y se revisó la traza. Por medio de estas pruebas se encontraron algunos errores y se corrigieron. Los errores encontrados fueron de tres tipos:

- a) En la programación de vectores de validación (se corrigieron reprogramando las funciones del PL).
- b) En la definición de vectores de validación (se corrigieron ajustando el vector en la aplicación).
- c) Se identificaron cifras o valores que fue necesario considerar como revisiones de caso.

Estas últimas resultan de casos en que la inconsistencia de datos no puede ser corregida por los vectores aplicados y es necesaria la intervención de una persona. Aunque en algunos casos fue posible realizar una re consulta con el informante y encontrar una solución, la mayoría de ellos se resolvió revisando los datos con el equipo de campo y buscando la mejor solución con el área conceptual.

Después de realizar las correcciones se volvieron a probar los vectores sobre la base de desarrollo. Una vez que no se encontró ningún error se liberaron, considerándolos terminados y listos para ser aplicados en la base de datos productiva.

## 5.3 Fase de validación definitiva<sup>6</sup>

Al terminar de realizar las pruebas, se migraron los datos de la base preliminar a la base en el esquema de producción, con todos los ajustes realizados a partir de las revisiones de caso. A esta base de datos se le aplicaron nuevamente los vectores de validación que fueron probados previamente, para verificar la consistencia de los datos y, en su caso, se realizaron los ajustes finales. La base validada se liberó como base

definitiva el 08 de enero de 2018 para su uso en los procesos de generación de resultados.

## 5.4 Resumen de datos procesados

De acuerdo con la forma en que están diseñados los cuestionarios que corresponden al Censo Nacional de Gobierno Federal 2017, y para efectos de la integración, procesamiento y validación de la información, se genera una base de datos relacional. La base de datos generada considera un total de 38 073 datos individuales, sin considerar las relaciones entre las tablas, los catálogos, y las llaves de las tablas. Con respecto al total de datos que se contabilizan, el 85.88 por ciento corresponden al módulo 1, el 10.04 por ciento corresponden al módulo 2 y el 4.08 por ciento corresponden al módulo 3.

Durante el procesamiento de los datos que se integraron en la base de datos y que se derivaron de los cuestionarios respondidos, se incluye la discriminación de cuatro tipos de datos: Valores o datos mayores o iguales a cero y que incluyen códigos relacionados con preguntas de tipo “verdadero” y “falso”; Valores que corresponden a reactivos que, por las características de las preguntas no requieren una respuesta o no requieren el registro de datos específicos, y que se registran como “No aplica” y se codifican con “-2”; valores que corresponden a datos que no fueron proporcionados por el informante, ya fuera porque no supieron la respuesta o porque no tuvieron elementos de información en sus registros administrativos para responderla y que se registraron como “No se sabe” y se codificaron con “-1”; y por último, valores nulos que corresponden a variables que no requieren registro de información dada la construcción de las tablas y las características de las preguntas.

El siguiente cuadro muestra, por módulo, el porcentaje de datos que fueron procesados, de acuerdo con los tipos mencionados en el párrafo anterior:

Módulo	Porcentaje de datos procesados				
	Valores mayores o iguales a cero	Valores "No se sabe"	Valores "No aplica"	Valores "No disponibles"	Valores nulos
M1	54.04	7.58	3.55	0	20.7
M2	2.84	0.25	6.65	0	0.3
M3	1.21	0.86	1.46	0	0.56

<sup>6</sup>Con fecha 07 de mayo de 2018 se actualizó el presente capítulo de esta Memoria, debido a la incorporación de los apartados 5.3 Fase de validación definitiva y 5.4 Resumen de datos procesados.