

---

## Appendix A. Census of Agriculture Methodology

---

## Apéndice A. Metodología del Censo Agrícola

---

### INTRODUCTION

The 2012 Puerto Rico Census of Agriculture was primarily conducted by mail, with report forms sent to farm operators on the census mail list (CML). This mail list was supplemented by an area sample which accounted for farms not included in the mail list (NML). A single version of the report form, in Spanish, was used for the CML and NML contingents of the Census. Combining data from the CML and the area sample NML should provide complete coverage for the Puerto Rico Census of Agriculture.

### CENSUS MAIL LIST COVERAGE AND WHOLE FARM NONRESPONSE

The initial CML consisted of 16,971 names and addresses stratified by size and type of farm. Nonresponse weighting was applied to account for farm operators who did not respond. For each municipio containing agricultural operations, nonresponse weights were computed by response homogeneity group (RHG); the RHGs were six size categories for historical total values of production reported for the 2012 farm records in that municipio. Each nonresponse weight was calculated as the total number of all agricultural operations listed on the CML for that RHG and municipio, divided by the total number of responding farms within that RHG and municipio. For each RHG found in a municipio, this weight, integerized, was applied to the data items of each responding operation of the RHG. This estimation procedure assumes that the distribution of farms in an RHG is the same for census nonrespondents as for census respondents. Records that were found to be undeliverable as addressed (“UAA”) by the U.S. Postal Service and which failed to respond to any follow-up attempts, were excluded from the nonresponse calculations. Note that NML

### INTRODUCCIÓN

El Censo de Agricultura de 2012 en Puerto Rico se llevó a cabo en su mayoría a través del correo. Los formularios fueron enviados a los agricultores, usando el listado de correo del censo (CML, por sus siglas en inglés). Este listado fue complementado con una muestra de área que identificó a los agricultores que no se encontraban en la lista (NML, por sus siglas en inglés). Se utilizó un solo formulario, en español para ambos contingentes del censo (CML y NML). Combinando la información de la lista (CML) y de la muestra (NML) debemos obtener una cobertura completa para el Censo de Agricultura de Puerto Rico.

### COBERTURA DE LA LISTA CENSAL Y LA NO RESPUESTA

La lista inicial (CML) consistió de 16,971 nombres y direcciones estratificados por tamaño y tipo de finca. Se aplicó la ponderación de no respuesta para incluir aquellos agricultores que no respondieron al censo. Para cada municipio donde hubo fincas, la ponderación fue calculada para respuestas agrupadas de forma homogénea (response homogeneity groups, o RHG). Los RHGs contenían seis categorías de valores de producción según informados en el 2012 en los récords de las fincas del municipio. Cada ponderación fue calculada como el total de todas las fincas listadas en CML, para todo RHG y su municipio, dividido por el número total de fincas que respondieron, dentro del RHG y el municipio. Por cada RHG encontrado en un municipio, esta ponderación, totalizada, fue aplicada a las partidas de cada finca dentro del RHG. Este procedimiento de estimación asume que la distribución de fincas en un RHG es la misma para la respuesta y la no respuesta. Récords que fueron devueltos por mala dirección postal (“UAA”) por el Servicio Postal de los EE. UU. y que no respondieron a ningún procedimiento de seguimiento, fueron excluidos de las calculaciones de la no respuesta. Note que los récords de NML no fueron sometidos a la ponderación debido a que estos fueron sometidos al proceso de seguimiento. Si no pudieron

records were not subject to nonresponse weighting because they were subjected to complete nonresponse follow-up. If the data were not captured from the nonrespondents, the records were then manually estimated using data from respondents.

Nonresponse weighting was monitored in an effort to keep nonresponse weights from exceeding 2.0; if an RHG's nonresponse weight exceeded this upper limit, a collapsing algorithm was applied to combine neighboring RHGs within a municipio, and recomputed the weights for the combined groups. For published tabulations of complete count items, the noninteger nonresponse weight was rounded to an integer weight of 1, 2, or 3. A random rounding algorithm based on cumulated fractional parts of the noninteger nonresponse weight values was used to produce an unbiased result at island level.

## **COVERAGE ADJUSTMENT**

Due to the dynamic nature of mail lists, some farm operators may not be included in the CML. To account for this undercoverage, an area frame consisting of the entire island of Puerto Rico was sampled. For sampling, NASS stratified the Puerto Rico area frame on the basis of agricultural intensity, with strata consisting of: 1) land area with dense agriculture; 2) land area with sparse agriculture and few houses; 3) land area with sparse agriculture and many houses; 4) cities with no apparent agricultural activity; and 5) ostensibly nonagricultural land such as parks and military reservations. Primary Sampling Units (PSUs) were created based on specific size requirements and permanent boundaries. An additional sampling enhancement involved the grouping of municipios with similar agriculture into nine clusters. Within each stratum and cluster, a random sample of PSUs was selected and then further subdivided into target sampling units called segments. Of approximately 7,500 segments available for sampling, 300 segments were selected. Aerial photography and maps for the 300 segments were provided to support field data collection. All NML farms discovered within the 300 sampled segments were included

obtener los datos del agricultor, la información para la finca fue estimada por el enumerador basada en observación y datos administrativos.

La ponderación aplicada fue monitoreada para el fin de no excediera un valor mayor de 2.0; si una ponderación de RHG excedía el límite superior, se le aplicó un algoritmo para combinar RHGs en la vecindad dentro del municipio, y se computó la ponderación para el grupo combinado. Para propósitos de la tabulación, en cada record, la ponderación de no respuesta para fracciones de finca se redondeó a una ponderación de 1, 2 ó 3. Para garantizar un resultado sin sesgo, a nivel de Isla, utilizamos un algoritmo de redondeo aleatorio para calcular los remanentes acumulativos.

## **AJUSTE DE COBERTURA**

Se presume que una lista de nombres y direcciones nunca está completa. Por lo tanto; y para ampliar la cobertura, se utilizó un marco muestral de área representativo de toda la Isla de Puerto Rico. El marco muestral se estratificó por el NASS de acuerdo con la actividad agrícola. Los estratos se clasificaron como: 1) áreas con mucha agricultura, 2) agricultura dispersa y pocas viviendas, 3) agricultura dispersa y muchas viviendas, 4) ciudades con cero agricultura, y 5) áreas con cero agricultura tales como parques e instalaciones militares. Las unidades muestrales primarias, se crearon utilizando criterios de tamaño y límites permanentes. Además, se crearon nueve grupos de municipios con características agrícolas similares. Dentro de cada estrato y cada grupo, se seleccionó una muestra aleatoria de unidades primarias que luego se subdividió en unidades de muestreo de enfoque llamadas segmentos. Se seleccionaron 300 segmentos de alrededor de 7,500 segmentos disponibles. A los enumeradores se les proveyó una serie de fotografías aéreas y mapas de los 300 segmentos para el trabajo de campo. Toda operación NML que se encontró dentro de los 300 segmentos y que no estaba en la lista fue designada como operación de la muestra.

in the area sample.

## **PRE-CENSUS AREA SCREENING**

Prior to the actual census data collection, enumerators using aerial photos and municipio maps identified all farm operators within each assigned area segment. Enumerators recorded the farm operator's name, address, and cuerdas operated within the segment. Farm operators from the selected segments were then checked against the Census Mail List. If no match was found, a census report form was sent to the NML operator, with enumerators making follow-up visits to collect the data. For those farm operators discovered in an area segment but included in the Census Mail List, only the CML report was used, to avoid duplication. During the pre-screening process, 589 NML farm operators were found in the 300 sampled area segments, but only 295 of these original NML records were determined to be actual farm operators.

## **NML DISTRIBUTION AT THE MUNICIPIO LEVEL**

Although the area sample size that determined the NML component ensured acceptable precision at the island level, the sample was not designed for reliable estimation of municipio-level data. To redistribute the island-level NML component in a more reasonable manner, a statistical calibration model was developed using information from similar CML and NML records. The 2012 model, based on a prototype used for the 1997 Puerto Rico Census, but extended to include small, marginal farms which may be omitted from the Census list sources, applies the area sample design "cluster" definition cited above, as well as the census categorical variable, farm type. For each municipio and farm type, the expanded estimate of total farms of that farm type was calculated for CML farms and combined with the count of unweighted NML farms of that farm type within the municipio. An analogous total, combining the CML estimate of total farms and the unweighted NML count for that farm type, was also computed at cluster level for the municipio's cluster: the municipio's share of each NML farm found in its cluster was calculated as the ratio of

## **SELECCIÓN DE LA MUESTRA PRE-CENSO**

Antes de llevar a cabo el censo, los enumeradores, usando fotografías aéreas y mapas de los municipios, identificaron todas aquellas fincas que estaban dentro de cada segmento. Anotaron el nombre del agricultor, la dirección postal, y el total de cuerdas dentro del segmento. Los nombres de los agricultores dentro del segmento especificado fueron cotejados contra la lista postal del censo. Si el nombre no se encontraba, este operador NML se designó como para recibir un formulario y un enumerador le visitó para obtener la información. Para evitar duplicación, aquellos agricultores encontrados en el segmento, pero que estaban en la lista, solo el reporte CML fue utilizado. Durante este proceso, se encontraron un total de 589 fincas potenciales, pero solo 295 resultaron ser operaciones agrícolas.

## **DISTRIBUCIÓN DE LA MUESTRA (NML) AL NIVEL DE MUNICIPIO**

Aunque el tamaño de la muestra que determinó el componente de NML asegura una precisión aceptable a nivel de Isla, la muestra no está designada para hacer estimados adecuados a nivel de municipio. Para redistribuir el componente de NML a nivel de Isla de manera más razonable, se desarrollo un modelo de calibración usando información de récords similares en la lista y en la muestra récords CML y NML. El modelo de 2012, está basado en el modelo usado en el censo de Puerto Rico de 1997, pero se expandió para incluir fincas pequeñas que fueran excluidas de las fuentes del listado; aplica la definición del diseño de grupos mencionados anteriormente, así como la variable censal categórica, tipo de finca. Para cada municipio y cada tipo de finca, el estimado total de fincas expandidas por tipo de finca, se calculó para fincas CML en la lista y se combinó con el total no ponderado de fincas NML en la muestra para ese tipo de finca dentro del municipio. Un total análogo, combinando el estimado CML del total de fincas y el conteo no ponderado NML para ese tipo de finca, fue también calculado a nivel de grupo para los municipios agrupados-la parte del municipio en cada finca de muestra NML, encontrado dentro de cada grupo, fue calculada como la razón del total de fincas a nivel de municipio de ese tipo de finca, al total análogo

the municipio-level total farms of that type, described above, to the analogous total farms of that type at cluster level. The municipio-level expansion weight for an NML farm in a cluster was developed as the product of its conventional area frame weight and the municipio's share of the farms in that farm type. Within an original NML record's cluster, a replicate record was created to carry the fractionalized weight and the data allocated to any other municipio in the cluster with a share in NML farms of that farm type. The integerization algorithm applied to the municipio-level NML weight was designed to prevent the sum of the redistributed municipio level farm count to deviate by more than 2 from the island level NML farm count.

## **ESTIMATION**

After weighting adjustment of the CML farm records for nonresponse, and of the NML farm records for sample expansion and municipio allocation, the CML and NML components were combined to provide a single estimate. Since the CML and NML contingents are mutually exclusive, our combined estimate should reflect complete and unduplicated coverage, provided that there is no significant nonresponse bias, and other nonsampling error is not operative.

## **EDITING DATA**

Captured data were processed through a format program. The program verified that record identifiers were valid and checked the basic integrity of the data fields. Rejected records were referred to analysts for correction. Accepted records were sent to a computer batch edit process. Each execution of the computer batch edit flowed as the data were received from the National Processing Center (NPC).

All 2012 census records were passed through a complex computer edit. The edit determined whether a reporting operation met the minimum criteria to be counted as a qualifying farm (in-scope). Operations failing to meet the minimum criteria (out of-scope) were referred to analysts for verification. The edit examined each in-scope record for reasonableness and completeness and

de ese tipo de finca a nivel de grupo. La ponderación a nivel de municipio, para una finca de la muestra dentro de un grupo, se crea como el producto de toda ponderación convencional y la participación del municipio en ese tipo de finca. Dentro de los records originales NML en los grupos de la muestra, se replicó un record para que llevara la ponderación fragmentada y la información recreada a cualquier otro municipio dentro del grupo con participación en ese tipo de finca en la muestra. Este algoritmo de números enteros aplicado a la muestra a nivel de municipio, se diseñó para prevenir que la suma de las fincas redistribuidas a nivel de municipio rebasara el nivel de 2 previamente establecido para la muestra en el conteo de fincas NML a nivel de isla.

## **ESTIMACIÓN**

Después del ajuste ponderado de los records de finca CML, por no respuesta, y de los records de finca NML, por expansión de la muestra y distribución de municipio, ambos componentes (CML y NML) fueron combinados para proveer un solo estimado. Como ambos componentes CML y NML son mutuamente exclusivos, nuestro estimado total debe reflejar una cobertura completa e irreproducible, siempre y cuando no exista sesgo de no respuesta y los errores de muestreo y otros sesgos sean mínimos.

## **REVISANDO LOS DATOS**

Los datos recolectados se procesaron a través de un programa de formato. El programa verificó que los identificadores del cuestionario fueran válidos y también comprobó la integridad básica de los campos de datos. Los cuestionarios rechazados fueron remitidos a los analistas para su corrección. Los cuestionarios aceptados fueron agrupados y enviados a un proceso de revisión informática. Cada ejecución de esta revisión informática se realizó según se fueron recibiendo los datos del Centro Nacional de Procesamiento (NPC por sus siglas en inglés).

Todos los cuestionarios del censo del 2012 se pasaron a través de un programa de revisión complejo. Este programa determinó si el cuestionario cumplía con los criterios mínimos establecidos para ser clasificado como una finca/graja para el censo. Los cuestionarios que no cumplían con los criterios mínimos fueron remitidos a

determined whether to accept the recorded value for each data item or take corrective action. Actions included removing erroneously reported values, replacing an unreasonable value with one consistent with other reported data, or providing a value for an overlooked item. To the extent possible, the edit determined a replacement value. Strategies for determining replacement values are discussed in the next section.

The edit systematically checked reported data section-by-section with the overall objective of achieving an internally consistent and complete report. NASS subject-matter experts defined the criteria for acceptable data. Problems that could not be resolved within the edit were referred to an analyst for intervention. Analysts used additional information sources, examined the scanned image, and determined an appropriate action. Puerto Rico analysts used an interactive version of the edit program to submit corrected data and immediately re-edit the record to ensure a satisfactory solution.

## Imputing for Missing Data

Missing data occurred whenever a respondent failed to report in a cell that should have a positive value or when the edit determined a value was not reasonable and should be changed. The edit performed a sequence of steps that determined the best value to impute for the missing item. If an item could not be calculated directly from other data reported on the current form, the edit checked for previously reported data. Operator characteristics, such as race and gender, were brought forward from the previous census if the operator had not changed in five years. When these deterministic sources failed to produce a solution, the edit invoked an automated imputation system which searched for a reporting farm of similar type, size, and location to provide a value for the missing data item. If the imputation algorithm failed to provide a solution, the record was referred to an analyst for resolution.

The guiding principal for imputation was to find

los analistas para verificación. El programa examinó cada cuestionario en el ámbito de lo razonable y lo completo, y determinó si aceptaba el valor registrado para cada partida o tomaba medidas correctivas. Las acciones incluyeron la eliminación de valores reportados erróneamente, la sustitución de un valor no razonable por uno consistente con los otros datos registrados, o el proveer valores adecuados para partidas dejadas en blanco. Según fuese necesario, el programa determinó valores sustitutos. Las estrategias para determinar los valores sustitutos se discuten en la siguiente sección.

El programa verificó todos los datos reportados en el cuestionario de una manera sistemática, sección por sección, con el objetivo general de lograr un informe de datos que fueran internamente consistentes y completos. Los analistas de NASS definieron los criterios para datos aceptables. Los problemas que el programa no pudo resolver fueron remitidos a un analista para su intervención. Los analistas utilizaron fuentes adicionales de información, examinaron la imagen escaneada, y determinaron la acción apropiada. Los analistas en Puerto Rico utilizaron una versión interactiva del programa que les permitía hacer correcciones a los datos reportados e inmediatamente verificar que los cambios sometidos ofrecían una solución satisfactoria.

## La Imputación de Datos

La imputación de datos ocurrió cuando el respondedor no proveyó información para una partida que requería un valor positivo o cuando el programa determinó que el valor reportado no era razonable y se tenía que cambiar. El programa realizó una secuencia de pasos que determinó el mejor valor para sustituir por el valor que faltaba. Si el valor de una partida no puede ser calculado directamente utilizando los otros datos presentes en el cuestionario, el programa utilizó los datos reportados en cuestionarios anteriores. Las características demográficas del operador, tales como raza y género, fueron reproducidas utilizando los datos del censo anterior si el operador no había cambiado en cinco años. Cuando estas fuentes deterministas no podían llegar a una solución, el programa invocaba un sistema de imputación automática que buscaba una finca de características similares en cuanto al tipo, el tamaño y la ubicación para proporcionar un valor para esa partida en específico. Si el algoritmo de imputación no proveyó una solución, el cuestionario se envió a un analista para su resolución.

a close match to the farm with the missing item. The census imputation algorithm relied on a pre-established donor pool. The donor pool included a collection of completed reports that had successfully navigated the edit. The pool was further divided into groups of similar type and size, referred to as profiles. When the edit determined the need to impute an item, it went to the appropriate profile and searched for the best fit. Best fit was determined by calculating "distance" between the incomplete report and each candidate donor using a set of match variables. Match variables were specific to each section of the report form and included the latitude and longitude of the principal county of operation. The distance was the sum of the squared differences between the reported values of the match variables. The donor with the smallest distance was considered the "nearest neighbor" and became the source for the imputation action. The value returned may have been a direct copy of the donor's value. In many cases, a relationship between two related variables on the donor record was applied to a reported value on the incomplete record. Using crop production as an example, the donor's production was divided by its harvested cuerdas (yield) and multiplied by the recipient's harvested cuerdas to obtain imputed production.

The imputation process was imbedded in the edit. When the edit determined an item required imputation, the edit program launched the algorithm, waited for a value to be returned, validated that the returned value was satisfactory, and resumed editing. Since imputation was conducted independently for each occurrence, reports requiring multiple imputations drew from multiple donors.

Initial donor pools were established before the first batch edits were run. These donor pools were "seeded" with 2007 census data that were "mapped" to look like 2012 data and passed through the 2012 edit to ensure they were consistent using the 2012 data relationships. As 2012 data were successfully processed, new records systematically replaced the older records in the donor pool. The older records disappeared entirely from the donor pool after the first few

El principio fundamental para la imputación de valores se basa en encontrar una finca con características semejantes para una partida en específico. El algoritmo de imputación del censo dependió de un grupo de donantes preestablecido. El grupo de donantes se creó utilizando una colección de informes completos que habían navegado con éxito el programa de revisión. El grupo de donantes se dividió en grupos con características similares tales como tipo de finca, y tamaño. Estos grupos de denominaron como perfiles. Cuando el programa determinó la necesidad de imputar una partida, el programa identificó el perfil adecuado y buscó el valor de mejor ajuste. El valor de mejor ajuste se determinó mediante el cálculo de la "distancia" entre el cuestionario incompleto y cada donante disponible utilizando una serie de variables con valores semejantes. Estas variables eran específicas para cada sección del cuestionario e incluían la latitud y la longitud del municipio principal donde se encontraba localizada la finca. La distancia se calculó sumando las diferencias al cuadrado entre los valores reportados por las variables de valores semejantes. El donante con la distancia más pequeña era considerado el " vecino más cercano " y se convirtió en el valor utilizado para la imputación. El valor utilizado para la imputación puede haber sido una copia directa del valor del donante En otros casos, una proporción de valores entre variables del donante fue aplicada a un valor reportado. Utilizando como ejemplo la producción de cultivos, el rendimiento del donante se multiplicó por el número de cuerdas cosechadas del receptor para obtener la producción imputada.

El proceso de imputación se ejecutó como parte del programa de revisión. Cuando el programa de revisión determinó necesaria la imputación de una partida, inició automáticamente el algoritmo de imputación, esperó por el cálculo de un valor sustituto, verificó que el valor sustituto fuese satisfactorio, y reanudó nuevamente la revisión de datos. Dado que la imputación se realizó independientemente para cada partida, los cuestionarios que requirieron múltiples imputaciones potencialmente utilizaron múltiples donantes.

Se establecieron conjuntos de donantes iniciales antes de que se ejecutaran las primeras revisiones en grupos de cuestionarios, utilizando los datos del Censo Agropecuario del 2007. Estos datos fueron revisados para que fuesen similares a los datos del Censos Agropecuario del 2012. Según se fueron procesando los

batch edits.

The donor pool for each State was refreshed weekly during the first couple of months of editing. As the flow of new data slowed, the donor pools were refreshed biweekly.

In some cases, nearest-neighbor imputation was not possible. The requirement of a positive imputed value could have ruled out all available donors, resulting in an imputation failure. An imputation failure could have occurred if there were no donors in the same profile as the report being edited. Records with imputation failures were either held until more records were available or referred to an analyst.

## **CENSUS SURVEY ERROR**

Because the NML component of the census data is an estimate, and surveys in general are prone to human error, the possibility of both sampling and nonsampling errors is always present. Sampling errors occur when only a portion of a population is selected. For samples with known probabilities of selection, the precision for a sample can be determined and confidence intervals calculated. In addition, the precision for list nonresponse can be determined and confidence intervals calculated as well. Nonsampling errors are associated with mistakes in reporting or keying the data as well as imputing for nonresponse.

## **CENSUS SAMPLING ERROR**

The 2012 Puerto Rico Census of Agriculture, like the three Censuses preceding it, used an area frame to estimate the NML component from a sample: thus there is sampling error associated with this estimator. The 300 segments sampled represented only a small proportion of the approximately 7,500 segments potentially

datos del Censos Agropecuario del 2012, los nuevos records válidos fueron reemplazando sistemáticamente el conjunto de donantes originales, hasta que el grupo original fue reemplazado por completo.

Los valores del grupo de donantes se actualizaron aproximadamente cada dos semanas durante los primeros meses de revisión. A medida que el flujo de nuevos datos fue disminuyendo, los valores del grupo de donantes se actualizaron con menor frecuencia.

En los casos relativamente raros en que el programa de revisión fue incapaz de proporcionar un donante, el proceso de selección de donantes emitió un mensaje de fallo apropiado para la revisión. Los fracasos de imputación se produjeron por diferentes razones. El requisito de un valor imputado positivo podría haber descartado todos los donantes disponibles, lo que resulta en un fracaso imputación. Un fracaso de imputación también podría haber ocurrido si no había donantes en el mismo perfil del cuestionario que se estaba siendo revisado. Los cuestionarios con fracasos de imputación se revisaron cuando hubo más cuestionarios disponibles o se refirieron a un analista para su resolución.

## **ERROR DE LA ENCUESTA**

Dado que el componente NML de los datos de censo es un estimado, y las encuestas están usualmente expuestas a errores humanos, la posibilidad de errores de muestreo siempre existe. Los errores de muestreo ocurren cuando solo se selecciona una porción de la población total. En muestras con una probabilidad conocida de selección, la precisión de la muestra puede determinarse y los intervalos de confiabilidad pueden calcularse. Adicionalmente, la precisión a la no respuesta puede determinarse y los intervalos de confiabilidad pueden calcularse también. Los errores ajenos al muestreo se asocian con errores cometidos al anotar o entrar los datos y con imputaciones a la no respuesta.

## **ERROR DE MUESTREO**

El Censo Agrícola de Puerto Rico del 2007, al igual que los dos censos anteriores, usó un marco muestral de área para estimar el componente de NML, por lo tanto existe error de muestreo asociado con este estimador. En Puerto Rico, los 300 segmentos seleccionados representan solamente una pequeña porción de los casi

available. If a different 300 had been selected, the results would have been different. The error resulting from the difference between the sample actually selected, and the universe it was designed to represent, is called sampling error. If the sample of 300 is selected with known probabilities, then the sampling errors can be estimated from just one sample. The sampling error of an estimate is generally expressed in terms of the variance of the estimate, or its square root, the standard error, a measure of precision allowing the user to gauge the accuracy of the estimate. These measures are based on the average sum of the squared differences between each value of the reported data item and the mean value for that item.

Assuming that for each municipio, nonresponding CML farms in a given CML nonresponse stratum resemble the responding farms in that stratum, we can also apply the results of sampling theory to quantify the variability generated by nonresponse in the portion of any estimate coming from the CML contingent. For that estimate, we compute CML-based nonresponse variance for a stratified sample, assuming that when there is nonresponse within a nonresponse stratum, the respondents represent a random sample of all eligible farms, responding or nonresponding, within that stratum, and that, within municipio and stratum, response is independent of the response occurring within any other municipio and stratum combination. In our Census, nonresponse can occur only for the non-selfrepresenting CML cases. Under the assumptions of random and independent nonresponse, we sum the individual CML variances for an estimate, from each combination of municipio and stratum, to obtain the total nonresponse variance of that estimate at Island level. For each estimate, the CML-based nonresponse variance is added to the sampling variance for that estimate from the NML contingent, to obtain the total variance of the estimate for the 2012 Puerto Rico Census of Agriculture.

The figures presented in Table A-1 are estimates of the percent relative standard error of selected quantities for Puerto Rico. The percent relative

7,500 segmentos disponibles. Si se hubiesen seleccionado otros 300 segmentos diferentes, el resultado hubiese sido diferente al obtenido. A la magnitud de estas diferencias se le llama error de muestreo. Si la muestra de 300 segmentos es seleccionada con una probabilidad conocida, entonces el error de muestreo podría estimarse con una sola muestra. El error de muestreo en un estimador es generalmente expresado en términos de la varianza del estimado o su raíz cuadrada, el error estándar, que es una medida que provee al usuario una mejor comprensión de la medida de la precisión del estimador. Estas medidas se basan en la suma promedio de los cuadrados de las diferencias entre cada valor de las partidas reportadas y el valor promedio para esa partida.

Asumiendo que para cada municipio, las fincas CML en la lista que no respondieron y que estaban dentro de un mismo estrato CML, tenían características similares a las fincas que completaron el formulario, también podemos aplicarles los resultados de teoría de muestreo para demostrar la varianza generada por la no respuesta en la parte del estimado proveniente del contingente de CML. Para este estimado, calculamos la varianza para una muestra estratificada basándonos en el CML, asumiendo que cuando existe la no respuesta dentro de un estrato de no respuesta, el entrelistado representa una muestra aleatoria de todas las fincas elegibles, respuesta o no respuesta, dentro del estrato, y esta respuesta, dentro del municipio y el estrato, es independiente de la respuesta que ocurre dentro de la combinación de cualquier otro municipio y el estrato. En este Censo, la no respuesta solo puede darse para los casos de CML que no son representativos de por sí. Bajo esta premisa, de no respuesta independiente y aleatoria, sumamos la varianza individual de CML para un estimado, de cada combinación de municipio y estrato, para obtener el total de varianza de no respuesta del estimado a nivel de Isla. Para cada estimado, se sumo la varianza de no respuesta basada en la lista, a la varianza de muestra para este estimado en el contingente de la muestra NML, para obtener la varianza total del estimado para el Censo de Agricultura de 2007 en Puerto Rico.

Las cifras presentadas en la Tabla A-1 son estimados del por ciento relativo de error estándar para ciertas cantidades en Puerto Rico. El por ciento relativo de error estándar es igual al error estándar del estimador, dividido por el estimado y multiplicado por 100. Los estimados del error estándar relativo presentados en la



standard error is equal to the standard error of the estimate divided by the estimate itself, multiplied by 100. The estimates of relative standard error displayed in Table A-1 were derived to reflect the total sampling error, which comes from Area Sampling, and also, under our assumptions, from the nonresponse process operating on CML records.

The particular sample selected for this census is one of many different samples, which could have been chosen. Had many different samples been taken - that is, had sampling been performed repeatedly under the same general conditions with the same sample design - many different estimates of each of the quantities being estimated would have been obtained

The estimates of relative standard error displayed in Table A-1 are for Island-level estimated totals and farm counts and are a measure only of the sampling error of the estimates; they do not account for error in the estimates due to other sources of error, called nonsampling error. The following sections discuss different types of nonsampling errors that may be present in the estimates.

## **CENSUS NONSAMPLING ERROR**

Many other types of error are not caused by sampling and are not so easily quantified: *nonsampling error* may be introduced into the census through incorrect responses by reporters, or through mistakes made by programmers or data entry staffers, as in inappropriate adjustment of missing data. Nonsampling error due to mail list incompleteness and duplication or misclassification of records on the mail list is called coverage error, discussed in greater detail below. The purpose of the area frame is to eliminate nonsampling error associated with CML incompleteness, coverage error, but not nonresponse.

## **RESPONDENT AND ENUMERATOR ERROR**

Incorrect or incomplete responses to the census report form or to the questions posed by an

tabla B-1 fueron derivados para reflejar el error de muestreo, que resulta del muestreo por área, y también, bajo lo asumido, del proceso de no respuesta que opera en los récords CML.

La muestra de segmentos seleccionada para este Censo de Agricultura es una de varias muestras que pudieron haberse usado. Si se hubiesen usado varias muestras diferentes- esto es, si el muestreo se hubiese ejecutado repetidamente bajo las mismas condiciones con el mismo diseño de muestra - pudieron haberse obtenido muchos estimados diferentes de cada una de las cantidades estimadas.

Los estimados del error estándar relativo presentados en la Tabla A-1 son solo para totales estimados a nivel de Isla y totales de fincas y son una medida del error de muestreo en el estimado; no toman en cuenta por errores en el estimado debido a otras fuentes, llamados errores no debidos a la muestra. Las secciones siguientes discuten otros tipos de errores no de muestreo y que podrían estar presentes en el estimado.

## **ERRORES DE CENSO AJENOS AL MUESTREO**

El otro tipo de error en un censo ocurre como resultado de obtener respuestas incorrectas, de errores cometidos por las personas entrando los datos o de ajustes inapropiados cuando las partidas están en blanco. A estos se les conoce como errores ajenos al muestreo. También pueden surgir de la insuficiencia y duplicación de la lista, así como de clasificación errónea de los récords, y se les conoce como errores de cobertura, se discute con mayor detalle más adelante. Más aún, el marco muestral de área se usó para eliminar errores ajenos al muestreo asociados con la insuficiencia de la lista (CML) errores de cobertura, pero no por la no respuesta.

## **ERROR DEL ENUMERADOR O DEL ENTREVISTADO**

Respuestas incorrectas o incompletas en el formulario

numerator can introduce error into the census data. To reduce reporting error, detailed instructions for completing the report form were provided to each respondent. Questions were phrased as clearly as possible, based on previous tests of the report form. In addition, each

respondent's answers were checked for completeness and consistency by the complex edit program.

## **ITEM NONRESPONSE**

Nonresponse to particular questions on the census report form, which we would logically or statistically expect to be present, may result in a type of nonsampling error. When information reported for another farm with ostensibly similar characteristics is used to impute for item nonresponse, the final estimates may be biased, if the unobserved characteristics of the nonrespondents differ systematically from those reported by respondents at either the record level (for an individual farm operation) or collectively, if respondent records are averaged or totaled.

## **PROCESSING ERROR**

All phases of processing a census report form may introduce nonsampling error into the estimates. The processing of census report forms includes clerical screening for farm activity, computerized check-in of report forms and follow-up of nonrespondents, keying and transmittal of completed report forms, computerized editing of inconsistent and missing data, review and correction of individual records referred from the computer edit, review and correction of tabulated data, and electronic data processing. These operations undergo a number of quality control checks to ensure as accurate an application as possible, yet some errors may ultimately escape detection.

## **COVERAGE ERROR**

Farms incorrectly excluded or included in the census result in coverage errors. Possible

censal o las preguntas formuladas por el enumerador podrían introducir error en la información censal. Para reducir este error, se dieron instrucciones detalladas de como completar el formulario a cada uno de los agricultores. Las preguntas fueron fraseadas lo más claramente posible basándonos en evaluaciones anteriores del formulario. Además, cada respuesta fue

cotejada para verificar su integridad y consistencia.

## **PARTIDAS EN BLANCO**

Partidas en blanco en ciertas preguntas del cuestionario, para las cuales se esperaba una respuesta, pueden causar errores ajenos al muestreo. Cuando imputamos los datos para la partida en blanco usando información recopilada para otra finca con características similares, el estimado final puede estar sesgado. Esto resulta porque las características omitidas por los casos que no han sido reportados, pueden ser diferentes de las características reportadas por los operadores, lo mismo al nivel del record (operación de finca individual) o en el promedio o total.

## **ERROR DE PROCESAMIENTO**

Todas las fases del procesamiento para cada formulario del censo son fuentes para la introducción de errores ajenos al muestreo. El procesamiento de las formas censales incluye una revisión clerical de las actividades agrícolas, el registro computadorizado del recibo de los formularios y el seguimiento de los casos que no han respondido, la entrada y transmisión electrónica de los datos en archivos computadorizados; la depuración computadorizada de la información inconsistente u omitida, el repaso y corrección de casos con errores rechazados por el proceso de depuración de datos, un análisis y corrección de los datos tabulados; y el procesamiento electrónico de los datos. Estas operaciones conllevan un número de cotejo de control de calidad para asegurar que la aplicación sea lo más precisa posible. Sin embargo, varios de estos errores pueden escaparse sin ser detectados.

## **ERROR DE COBERTURA**

Fincas incorrectamente excluidas o incluidas en el censo

coverage errors include the following:

Overcount due to farms duplicated or enumerated more than once.

Undercount due to farms misclassified as nonfarms.

Overcount due to nonfarms misclassified as farms.

Record mismatch between the CML and NML frames.

These coverage errors were minimized with computerized duplication identification programs and careful enumeration of all land area within a sampled area segment. In addition, field enumerators were required to follow up CML cases identified as major agricultural operations, if they did not respond by mail. If all follow-up attempts failed, either the NASS survey database, the census historic database, or other more current sources were used to impute data for the record.

## **BIAS AND PRECISION OF THE ESTIMATES**

The nonsampling errors discussed above can be sources of bias leading to underestimates or overestimates, based on the actual processed data. Since the formulas used to calculate estimates of relative standard error do not account for the nonsampling error sources, the figures published in the following tables may also be somewhat biased. It is important to note that the largest portion of the variability in the estimates probably comes from sampling error; it is the relative standard error of the estimates due to this source of error that is reported in the following table.

resultan como error de la cobertura. Los siguientes son posibles errores de cobertura:

Cifras exageradas debido a duplicidad o enumeración de fincas por más de una vez.

Cifras subestimadas debido a fincas erróneamente clasificadas como no fincas.

Cifras exageradas debido a no fincas erróneamente clasificadas como fincas.

Errores de identificación de fincas entre la CML y la NML.

Estos errores de cobertura se minimizaron a través de una revisión ejecutada por las computadoras para identificar y eliminar récords duplicados y una enumeración somera de todo terreno agrícola dentro de cada segmento en la muestra. Además, los enumeradores visitaron agricultores de CML identificados como de una gran producción agrícola y que no respondieron a través del correo. Si después del seguimiento no hubo respuesta, se utilizó o la encuesta del NASS o la fuente de datos del censo anterior o otras fuentes más recientes para llenar los datos del record.

## **SESGO Y PRECISION DE LOS ESTIMADOS**

Los errores ajenos al muestreo discutidos anteriormente, pueden ser fuentes de sesgo que podrían causar que los datos publicados contengan estimados por debajo o sobre lo normal. Como las fórmulas usadas para calcular estimados del error estándar relativo no son responsables por los errores ajenos al muestreo, las cifras publicadas podrían estar un tanto sesgadas. Es importante tomar en cuenta que la porción mayor de variabilidad en los estimados proviene de errores de muestreo. Las tablas siguientes contienen el error estándar relativo de los estimados procedente de estas fuentes.

**Table A. Estimates of Precision for Puerto Rico Totals: 2012**

[For meaning of abbreviations and symbols, see introductory text.]

Item	All Farms	Relative Standard Error of Estimate
<b>FARMS AND LAND IN FARMS</b>		
Farms .....	13,159	0.7
Land in farms .....	584,988	1.7
Average size of farm .....	44.5	3.6
Farms by size:		
Less than 10 cuerdas .....	5,129	2
..... cuerdas	21,002	2
10 to 19 cuerdas .....	2,859	2.4
..... farms	39,083	2.3
..... cuerdas	2,872	3
20 to 49 cuerdas .....	85,023	3.3
..... farms	940	5.2
..... cuerdas	63,789	5.1
50 to 99 cuerdas .....	563	4
..... farms	73,170	4.3
..... cuerdas	401	4.7
100 to 174 cuerdas .....	81,321	4.9
..... farms	395	3.3
..... cuerdas	221,601	4.2
260 cuerdas or more .....		
<b>TENURE OF PRINCIPAL OPERATOR</b>		
Full owners .....	9,362	1.1
..... farms	275,689	2.2
..... cuerdas	1,069	4.3
Part owners .....	143,371	5.9
..... farms	2,728	3.3
..... cuerdas	165,928	2.5
<b>PRINCIPAL OCCUPATION OF PRIMARY OPERATOR</b>		
Agricultural .....	5,463	1.5
Nonagricultural .....	7,696	1.5
<b>TYPE OF ORGANIZATION</b>		
Individual or family .....	11,938	0.9
..... farms	406,461	2
..... cuerdas	117	8.7
Partnership .....	18,414	11
..... farms	738	3
..... cuerdas	138,152	4.6
Corporation .....	366	3.8
..... farms	21,962	10.6
..... cuerdas		
<b>FARMS BY MARKET VALUE OF AGRICULTURAL PRODUCTS SOLD</b>		
Less than \$1,000 .....	2,973	2.7
\$1,000 to \$2,499 .....	2,015	2.6
\$2,500 to \$4,999 .....	1,986	3.3
\$5,000 to \$7,499 .....	1,209	4.5
\$7,500 to \$9,999 .....	771	4.4
\$10,000 to \$19,999 .....	1,521	3.8
\$20,000 to \$39,999 .....	968	4
\$40,000 to \$59,999 .....	394	11.7
\$60,000 or more .....	1,322	3.9
<b>MARKET VALUE OF AGRICULTURAL PRODUCTS SOLD</b>		
Total sales .....	13,159	0.8
..... farms	547,628,613	2.1
..... dollars	41,616.30	4.8
Average per farm .....		
Crops, including nursery and greenhouse crops .....	9,367	1.1
..... farms	271,320,069	3.1
..... dollars	4,478	1.1
Coffee .....	29,273,215	3.2
..... farms	42	15
..... dollars	1,905,743	2.4
Pineapples .....	3,628	1.9
..... farms	80,505,103	6.6
..... dollars	1,474	2.6
Plantains .....	12,106,293	3.5
..... farms	781	7.2
..... dollars	8,539,117	3.1
Root crops or tubers .....	1,242	5.5
..... farms	8,191,970	8.4
..... dollars	1,926	3.2
Fruits and coconuts .....	31,541,191	8.6
..... farms	1,528	5.3
..... dollars	47,134,996	10.1
Vegetables and melons, including hydroponic crops .....	657	7.7
..... farms	37,533,811	4.5
..... dollars	532	8.9
Nursery, greenhouse, floriculture, and sod (see text) .....	14,588,630	3.2
..... farms		
..... dollars		
Grasses .....		
..... farms		
..... dollars		
Livestock, poultry, and their products .....	4,119	2.2
..... farms	276,308,544	2.8
..... dollars	2,911	2.2
Cattle and calves .....	33,605,544	2.2
..... farms	706	7.2
..... dollars	40,375,759	4.7
Poultry and poultry products .....	318	3.6
..... farms	189,425,127	2.4
..... dollars	532	10.6
Milk and other dairy products from cows .....	8,019,245	5.7
..... farms		
..... dollars		
Hogs and pigs .....		
..... farms		
..... dollars		

--continued

**Table A. Estimates of Precision for Puerto Rico Totals: 2012 (continued)**

[For meaning of abbreviations and symbols, see introductory text.]

Item	All Farms	Relative Standard Error of Estimate
<b>FARM-RELATED INCOME</b>		
Income from custom farmwork done for others ..... farms	464	5
dollars	11,288,566	6.5
Income from renting out farmland ..... farms	216	11.8
dollars	2,406,796	3.4
Income from participation in government farm programs ..... farms	2,241	2.6
dollars	35,351,806	2.7
Income from sale of farm by-products or waste materials ..... farms	43	12
dollars	611,359	0.5
Other farm-related income ..... farms	248	5.3
dollars	2,678,700	6.9
<b>SELECTED PRODUCTION EXPENSES</b>		
Livestock, poultry, and fish purchased ..... farms	2,168	2.7
dollars	24,507,205	2.8
Feed purchased for livestock and poultry ..... farms	4,418	2.1
dollars	107,502,439	3.3
Medicines and drugs purchased for livestock and poultry ..... farms	3,887	1.9
dollars	8,026,235	4.1
Veterinarian services ..... farms	1,336	2.7
dollars	2,287,289	5.9
Professional services ..... farms	2,708	2.2
dollars	5,774,613	4.8
Seeds, bulbs, plants, and trees purchased ..... farms	4,636	2.1
dollars	11,500,410	3.9
Commercial fertilizer purchased ..... farms	7,696	1.3
dollars	17,826,553	2.8
Gasoline and other fuel and oil products purchased ..... farms	10,983	1
dollars	25,220,663	2
Wages and salaries paid to employees or hired farm workers ..... farms	8,337	1.1
dollars	168,143,549	4.9
Contract labor ..... farms	657	2.9
dollars	2,852,309	3.5
Machine hire and customwork ..... farms	2,979	2.3
dollars	9,049,513	2.5
Agricultural chemicals purchased ..... farms	6,999	1.4
dollars	16,359,448	4.7
Machinery and equipment repair and maintenance ..... farms	3,034	2.8
dollars	12,683,427	2.8
Building repair and maintenance ..... farms	1,185	4.5
dollars	6,005,981	3.1
Water (see text) ..... farms	2,241	4
dollars	2,277,423	6.7
Electricity expense (see text) ..... farms	4,750	2.1
dollars	14,152,940	4.4
Interest expenses (see text) ..... farms	4,712	1.8
dollars	16,617,408	3.8
All other expenses (see text) ..... farms	8,954	1.1
dollars	45,753,671	3.7
<b>HIRED FARM WORKERS, AGREGADOS, AND SHARECROPPERS</b>		
Workers who worked 5 months or longer ..... farms	5,034	1.6
number	14,922	5.8
Workers who worked less than 5 months ..... farms	4,282	1.7
number	15,200	3.8
Agregado and sharecropper families living on place December 31 ..... farms	1,450	2.1
number	2,277	1.5
<b>INVENTORY OF LIVESTOCK, POULTRY, AND THEIR PRODUCTS</b>		
Cattle and calves ..... farms	2,522	2.2
number	130,339	1.9
Hogs and pigs ..... farms	631	9
number	48,262	3.7
Horses ..... farms	1,214	5.2
number	10,307	5.6
Sheep ..... farms	425	13.7
number	12,539	21.2
Goats ..... farms	470	8.2
number	5,655	2.5
Rabbits ..... farms	329	4.2
number	15,927	5.6
Laying hens (see text) ..... farms	343	13.4
number	304,789	10.9
Broilers and other chickens for meat production ..... farms	83	8.1
number	10,359,499	3
Fighting cocks ..... farms	949	5
number	34,612	10.6
<b>CROPS HARVESTED</b>		
Coffee grown in the shade ..... farms	1,919	1.9
cuerdas	9,337	3
cwt	32,155	5.2
Coffee grown without shade ..... farms	3,104	1.6
cuerdas	23,876	4
cwt	94,841	3.3
Pineapples ..... farms	106	7.2
cuerdas	330	5
tons	2,186	3.4
Plantains ..... farms	4,737	1.6
cuerdas	22,719	4
thousands	405,256	4.9
Bananas ..... farms	1,828	2.4
cuerdas	6,639	2.9
thousand	295,488	3.3