

ZIP Code® Tabulation Areas (ZCTAs™) For Census 2000 Dress Rehearsal Data Products

Introduction

The ZCTA technical documentation consists of two parts. The first part introduces the concept of a ZCTA and describes its technical characteristics. The second part outlines the automated ZCTA delineation methodology used for the Census 2000 Dress Rehearsal products.

Part 1: Technical Description

Introduction and Definition

ZCTAs are approximate area representations of United States Postal Service (USPS) ZIP Code® service areas. The Bureau of the Census has created ZCTAs to meet requests by data users for statistical data by ZIP Code area. In actuality, ZIP Codes are attributes of addresses. The term ZCTA was originated to differentiate between true USPS ZIP Codes and these approximations.

The Bureau of the Census is using a new, automated methodology to delineate ZIP Code areas for the Census 2000 Dress Rehearsal products. This methodology builds these areas using ZIP Codes associated with addresses collected during census operations. ZCTAs represent the majority ZIP Code for addresses within a tabulation block. Where no ZIP Code data are available for a block, the assignment process will spatially extend ZCTA coverage from an adjacent area to cover the block.

Although the delineation methodology is new, the requirements for ZIP Code areas remain largely unchanged from the 1990 Census. For the Census 2000 Dress Rehearsal products, ZCTAs will have the following basic characteristics:

- 1) Every tabulation block has a single ZCTA code;
- 2) ZCTAs cover every tabulation block;
- 3) ZCTAs may consist of discontinuous areas;
- 4) A ZCTA code represents either a five- or three-digit USPS ZIP Code; and
- 5) Water features may have a special three-digit ZIP Code for water features or the five- or three-digit ZIP Code of the surrounding land area.

The Bureau of the Census designed the ZCTA as a single layer, statistical entity to support simple tabulations of statistical data. Unlike census tracts and other program defined entities, ZCTAs are based on address data and they are not necessarily stable over time.

A ZCTA may not exist for every valid USPS ZIP Code. In most cases, the excluded ZIP Code identifies a single delivery point location for which we have insufficient data to establish a distinct ZIP Code area.

The Bureau of the Census has tried to create full ZCTA coverage. In doing so, the delineation process extends ZCTA coverage to areas with no ZIP Code or address data. The delineation process attempts to assign a full five-digit ZCTA code to the unassigned area. Where reliable data are unavailable for extensive areas, the ZCTA code may represent the more general three-digit ZIP Code. These extensions eliminate some of the spatial fragmentation that results from creating ZCTAs using the available address data alone. Even with extension, ZCTAs may consist of discontinuous parts.

The Bureau of the Census has prepared one-time-only, supplementary ZCTA data files to help the data users evaluate and use the Census 2000 Dress Rehearsal ZCTAs. These files show where the delineation is based on actual ZIP Code data and where blocks contain data for several ZIP Codes. These files are available upon request.

The ZCTAs for the Census 2000 Dress Rehearsal data products represent a prototype release, and may not reflect the final specifications for the Census 2000 ZCTA products currently under development. Data users should not assume that the Bureau of the Census will provide supplementary ZCTA data files for the Census 2000 ZCTAs.

Distinction Between ZCTAs and ZIP Codes

There are some significant differences between actual USPS ZIP Codes and ZCTAs. The USPS assigns ZIP Codes to delivery points to facilitate mail delivery. ZIP Codes belonging to city-style (house number/street name) and rural delivery addresses generally extend out along delivery routes that follow the street network. Addresses on opposite sides of a street almost always have the same ZIP Code. ZIP Codes may split census blocks. For instance, city-style addresses located on cross streets may have different ZIP Codes. ZIP Code boundaries often follow imaginary lines that are offset from the block face. Rural addresses may be dispersed a considerable distance away from the mail delivery point located along the street making it harder to define the offset boundary.

Addresses with dedicated post office box ZIP Codes may appear dispersed over a wide area that is coextensive with other delivery type ZIP Codes. These ZIP Codes have inexact boundaries and are usually considered delivery point locations rather than areas. However, some rural communities may only have dedicated post office box ZIP Codes. In these cases, nearly all of the addresses in a post office's service area will have a post office box ZIP Code. Because other ZIP Codes do not serve the same area, we can define a ZCTA for the post office box ZIP Code.

Individual firms or organizations may have their own dedicated ZIP Code depending on the volume of mail they receive. We consider these ZIP Codes as point ZIP Codes (representing a single delivery point) rather than areas

even though the land area covered by these organizations can be extensive. As a result, these ZIP Codes will not become ZCTAs.

ZCTAs represent generalized ZIP Code service areas rather than the ZIP Codes assigned to individual delivery points. ZCTAs avoid some of the complexity of USPS ZIP Code boundaries by following census tabulation block boundaries. As a consequence, a census block that contains addresses with many ZIP Codes will only show one ZCTA code; ZCTAs will not represent the ZIP Code for some addresses. Statistics based on tabulation by mailing ZIP Code will not be the same as those calculated based on ZCTA. Data users should never substitute a mailing ZIP Code with a ZCTA in address range or address list data.

ZCTA Codes

The Bureau of the Census identifies a ZCTA with a five-character alphanumeric code containing numeric values and trailing blank spaces. It is important to use the full five-characters of the code. The first three characters represents the three-digit ZIP Code. This code may contain leading zeros. For ZCTAs defined only by a three-digit ZIP Code the last two characters of the code are blank spaces. For example, code "223 " represents the generic three-digit ZIP 223 where no five-digit ZIP Code is available.

For ZCTAs that reflect a full five-digit ZIP Code, the last two characters of the code are numeric. For example, the code "00601" represents five-digit ZIP Code 00601. The ZCTA delineation process does not recognize codes ending in "00", such as "29000", as valid five-digit ZCTA codes.

Some water features will have a three-digit ZCTA code followed by the letters "HH", for example "290HH". These codes apply only to water features and usually belong to water features located along the edges of five-digit ZCTAs. The codes indicate that the water feature does not clearly fall within one five-digit ZCTA and is distinct from the three-digit ZCTA code assigned to land areas. In effect, these codes identify otherwise unassigned water areas.

Application of ZCTA Codes

The three-digit ZCTA code may apply to both land and water areas and always indicates the extension of ZCTA codes to large unassigned areas. The USPS may not provide five-digit ZIP Code delivery service in this area. The use of the three-digit ZCTA code indicates that we have insufficient information to determine the five-digit codes and avoids overextending the surrounding five-digit ZCTAs.

The size of the unassigned area determines if the extension process will assign five- or three-digit ZCTA codes. For contiguous unassigned land areas covering more than 25 square miles, the process assigns a three-digit

ZCTA code. For smaller areas the extension process assigns a five-digit code.

Water features may show a variety of different codes. Inland water features that share the same tabulation block number with the surrounding land area have the same ZCTA code as the land area within the block. For coastal Great Lakes and some larger inland water features, the ZCTA code will be the nearest adjacent three-digit ZCTA followed by the letters "HH".

Water features located along the borders of a ZCTA may also have the three-digit+HH ZCTA code. Some very small lakes or ponds along the edge of two or more ZCTAs may have a three- or five-digit ZCTA codes. The intent is to prevent large lakes or long double line drainage features from distorting the shape and extent of three- and five-digit ZCTAs.

The following table summarizes the criteria for assigning ZCTA codes to water features.

Table 1: ZCTA Codes on Water Features

Type of water feature	ZCTA Code
territorial	3-digit+HH ZCTA code
coastal	3-digit+HH ZCTA code
Great Lakes	3-digit+HH ZCTA code
inland features ≥ 10 square miles*	3-digit+HH ZCTA code
inland features bordering 2 or more ZCTAs > 0.25 square miles*	3-digit+HH ZCTA code
Remaining features	variable**

* based on contiguous water area for the feature within a county

** All remaining water features either receive the same ZCTA code as the surrounding or adjacent land sharing the same tabulation block as the water area, or a ZCTA code extended from the surrounding area.

ZIP Code Content

As stated earlier, not all ZIP Codes may appear as ZCTAs. The delineation process includes a filter to eliminate the following codes:

- 1) Invalid codes;
- 2) ZIP Codes unique to firms or organizations;
- 3) Discontinued ZIP Codes;
- 4) ZIP Codes not valid for the state and county; and
- 5) Dedicated post office box ZIP Codes associated with post offices with either city-style, highway contract, or rural route delivery.

The delineation process validates each address ZIP Code using ZIP Code-County relationships derived from USPS sources. These relationships come from an unduplication of the USPS ZIP+4 file. Additional ZIP Codes and relationships from the USPS City-State file supplemented these data. The information on delivery type and discontinued status comes from the USPS Delivery Type file. During the delineation process, we added some additional ZIP Code-County relationships based on address data collected during the Census 2000 Dress Rehearsal.

The validation process uses USPS files released for April 1998. These files are consistent with the address data, which were collected or verified during the Census 2000 Dress Rehearsal in mid-April 1998.

The Census 2000 Dress Rehearsal ZCTAs include some dedicated post office box ZIP Codes. In these cases, the main post office only provides post office box service. A ZCTA based on these ZIP Codes represents the majority of the addresses. These ZCTAs usually have smaller coverage areas.

In those situations where a post office has both dedicated post office box ZIP Codes and ZIP Codes with other delivery types, the post office box ZIP Codes compete with the other ZIP Codes. The ZCTA delineation process places a higher priority on non-post office box delivery and will not create ZCTAs using these post office box ZIP Codes.

The final ZCTAs for the Census 2000 Dress Rehearsal may include interactive corrections. In a small number of instances, the Bureau of the Census has changed some ZCTA assignments to eliminate widely separated ZCTA fragments. The changes typically involve small polygons (under 0.4 square mile area) that have only one or two addresses with ZIP Codes that are inconsistent with the surrounding area.

Part 2: Delineation Methodology

Overview

The purpose of the ZCTA delineation process is to assign a ZCTA code to each tabulation block. The process requires separate methodologies to:

- 1) Determine the ZCTA Codes for tabulation blocks with addresses; and
- 2) Extend ZCTA coverage to tabulation blocks with no addresses or ZIP Codes.

Where addresses are available, the delineation process determines the majority ZIP Code for a tabulation block using the address records in the Census 2000 Master Address File (MAF) geocoded to that block. The process includes ZIP Code filters to eliminate problematic codes.

The extension of the ZCTA coverage to areas with no address or ZIP Code data involves several procedures that fill specific types of unassigned areas. These areas include:

- 1) Unassigned areas within a ZCTA;
- 2) Coverage gaps between ZCTAs;
- 3) Outlying areas with no or scattered pockets of addresses; and
- 4) Water features.

Moving down the list, the assignment of codes becomes increasingly uncertain. In extreme rural areas with no available addresses, the process cannot reasonably determine a five-digit ZCTA code and, therefore, assigns a three-digit code. Water features require special consideration. The codes for inland water features depend on the codes from adjacent land areas, and are processed near the end of the delineation process.

Methods of ZCTA Extension

The delineation process assigns codes to unassigned areas by using two approaches:

- 1) Polygon edits which assign ZCTA codes to whole groups of polygons based on the topological relationship between these aggregate polygons and the surrounding area; and
- 2) Polygon extension which expands the existing coverage outward one polygon at a time using the ZCTA codes from adjacent polygons that were previously assigned.

Both approaches depend on the adjacency relationship between basic polygons (i.e., the smallest possible polygons created by the feature network technically referred to as GT-polygons) that share a common boundary segment. In the first approach, the polygon edits create ZCTA polygons by aggregating adjacent basic polygons with the same ZCTA codes. The process groups unassigned basic polygons into separate unassigned land and water ZCTA polygons. Then it attempts to fill these polygons by relating them to the neighboring ZCTA polygons. For instance, unassigned areas totally enclosed within another ZCTA can be absorbed into that ZCTA.

In the second approach, the polygon extension process fills unassigned areas by extending the ZCTA out to adjacent basic polygons that have no ZCTA codes. The process extends ZCTA coverage in a series of update cycles. Each cycle depends on the assignments from a previous cycle and adds another layer of basic polygons to the ZCTAs. Where different ZCTAs compete to fill an unassigned polygon, the process resolves the tie situation by selecting the ZCTA with the greatest contact along the edge of the polygon.

Polygon extension respects water features as natural barriers and defers the assignment of codes to these features until all possible land area is

covered. Water features receive three- or five-digit ZCTA codes during the process to enable the extension of ZCTAs to small islands. The delineation process resets the codes on some water features later.

Polygon extension is a simple method that can expand ZCTAs to cover large areas with relatively little processing. In one test area, we used Thiessen polygons to allocate the unassigned areas to ZCTAs and found the results comparable to polygon extension. However, the results of polygon extension are affected by differences in polygon size and become increasingly uncertain with each update cycle. To limit the extension of ZCTAs, the process blocks large unassigned areas from the extension process. These areas receive three-digit ZCTA codes as noted earlier.

The ZCTA delineation process tailors the delineation approach to the type of unassigned area. The sequence of operations is critical. After the initial assignments from the MAF, the process builds the initial ZCTA polygons and attempts to fill coverage holes. It then fills in the spaces between ZCTAs with five- and then three-digit ZCTAs. The process resets the codes on water areas and edits the results to ensure that ZCTAs do not split tabulation blocks. Attachment A outlines the sequence of processing steps.

Initial ZCTA Assignments

The majority ZIP Code becomes the ZCTA for tabulation blocks with address data. In rural areas typically with no city-style addresses, the data reflect the residential addresses collected and geocoded from field operations. In city-style delivery areas, both commercial and residential addresses contribute equally to ZCTA delineation. Commercial addresses supply ZIP code data for business districts where residential address may be sparse or missing. Note that the process suppresses unique commercial ZIP Codes dedicated to specific firms or organizations. The process suppresses other ZIP Codes that are neither valid nor current according to USPS sources.

In some tabulation blocks, the addresses may have either no ZIP Code or an unacceptable ZIP Code for ZCTA delineation. The delineation process considers these as unassigned areas and assigns them to a ZCTA using one of the extension methods outlined above. Tabulation blocks with no majority ZIP Code (e.g., “tie” situations) require resolution later during polygon extension when results from adjacent polygons can break the tie.

The process distributes the ZCTA codes from the tabulation block to each of the basic polygons within the block. This transfer facilitates polygon building and automated ZCTA extension which are polygon-based processes.

Polygon Building and Coverage Holes within ZCTAs

With the address-based ZCTA assignments in place, the delineation process begins to fill in the unassigned areas starting with coverage holes

within ZCTAs. The process builds ZCTA polygons and runs edits to absorb smaller unassigned areas into larger assigned areas. The edit performs the following:

- 1) Assigns codes to unassigned areas that are totally contained within a single ZCTA to that ZCTA;
- 2) Assigns codes to small unassigned areas (less than 1 square mile) that border the edge of a county and a single, surrounding ZCTA;
- 3) Absorbs small (0.25 square mile) isolated ZCTA fragments that consist of a single basic polygon that is totally contained within another ZCTA; and
- 4) Identifies large unassigned areas greater than 25 square miles land area to block automated extension of five-digit ZCTAs into these areas; these areas are candidates to receive a three-digit ZCTA code later.

The first edit fills holes within a single ZCTA. In the second edit, the process fills small gaps that are surrounded by a single ZCTA, but also touch the county or equivalent entity's boundary. Because ZCTA delineation is a county-based process, ZCTAs will not cross county boundaries unless actual address data indicate a county crossover.

The third edit removes any small ZCTA fragments totally enclosed in another ZCTA. Evidence suggests these anomalies are created by blocks with small numbers of addresses that are associated with incorrect or nonrepresentative ZIP Codes for an area.

The process also flags large contiguous land areas to prevent polygon extension as discussed earlier. The 25 square mile threshold is an arbitrary value based on limited observations. Water areas remain as separate polygons that are excluded from the calculations. This provision allows ZCTAs to expand outward to natural barriers such as rivers or lakes, but not beyond. It also prevents drainage features from connecting scattered unassigned land areas and forming large irregular areas that might be inappropriately excluded from five-digit ZCTA extension.

Rural Fragments

Large unassigned areas may contain small isolated ZCTA fragments that represent one or more five-digit ZCTAs. The fragments may reflect small communities or isolated homesteads. Rather than surround these fragments with a three-digit ZCTA, the process attempts to consolidate the fragments by coding the surrounding area to the appropriate five-digit ZCTA.

To define the surrounding area, the process constructs a "window" around the fragments using a Triangular Irregular Network (TIN). The process creates separate windows for each five-digit ZCTA and requires a minimum of three basic polygon fragments for each. All basic polygons that fall within the boundaries of the TIN become part of the window. To complete the window, the process constructs a one-basic polygon buffer

around the polygons that fall within the TIN. It is possible that the window may include all or most of the original unassigned area.

The resulting windows for several ZCTAs may merge together or connect with ZCTAs on the periphery of the surrounding unassigned area. Because several ZCTAs may be in contact, the delineation process uses the polygon extension routines to fill the window(s).

ZCTA Extension

The polygon extension process described earlier extends five-digit ZCTA coverage to all eligible basic polygons. The remaining unassigned areas receive a three-digit ZCTA code. If the USPS files indicate that a county has a single three-digit ZCTA, all remaining unassigned areas automatically receive this three-digit code. Otherwise, the process assigns the ZCTA code by extending out the first three-digits of established five-digit ZCTAs. The boundaries between three-digit ZCTAs may not match USPS three-digit ZIP Code boundaries due to the ZCTA extension methodology. Note that any windows around isolated ZCTA fragments may reduce the area receiving a three-digit ZCTA to below the 25 square mile threshold.

Water Feature Edit

The ZCTA delineation process contains an edit for water features. All codes on water features come from an extension process. These features do not require ZCTA codes for tabulation purposes. However, the delineation process assigns codes to some water features to avoid splitting tabulation blocks and to improve the ZCTA coverage for GIS applications. ZCTA codes on inland water features within ZCTAs eliminate coverage holes. Codes on coastal water features and along the boundaries of ZCTAs distort the size and shape of the ZCTA. River features are particularly problematic, because the basic polygons in these features may run for miles and create highly irregular ZCTAs.

The water feature edit selectively changes the three- or five-digit ZCTA code to a water code depending on the type of area, the surrounding ZCTAs, and the feature size (see Table 1). The edit initially identifies each contiguous water area as a distinct feature. Only inland water features may retain five-digit ZCTA codes. The process retains the three- or five-digit ZCTA codes (without using HH) on features smaller than 0.25 square miles to avoid creating “micro gaps” along ZCTA boundaries.

The assignment process may divide a feature between different three-digit+HH ZCTA codes where the water feature borders several three-digit ZCTAs. The assignment of the three-digit+HH ZCTA codes depends on the ZCTA assigned to the nearest adjacent land area. In the Census 2000 Dress Rehearsal sites, the application of some three-digit+HH ZCTA codes involved manual intervention.

Tabulation Block Edits

Before generating data products or summary data, the delineation process runs an edit to ensure that ZCTAs conform to Census 2000 tabulation blocks. The delineation process uses the MAF addresses geocoded to tabulation block to make the initial ZCTA assignments. As a result, assignments based on address data do not split tabulation blocks. In areas without address data, the ZCTA extension process makes assignments by basic polygon and not by block, creating the potential for some split blocks. The edit resolves these situations using the following rules:

- 1) Select the ZCTA code assigned during the earliest round or cycle of processing; and
- 2) Select the ZCTA code with the largest area where the updates reflect the same round of processing.

The process gives preference to assignments made during the earlier rounds of the polygon growing process. As each round of processing moves farther out from the areas with addresses, the accuracy of the assignments becomes more uncertain. Where all factors are equal, the edits select the ZCTA by majority area.

Product Generation

The tabulation block edit is the final update operation before product generation. The edit produces a ZCTA-tabulation block equivalency file for data tabulations. The delineation process also includes a final ZCTA polygon build to produce the summary ZCTA data for the Census 2000 Dress Rehearsal sites. An extract process produces a ZCTA boundary file that identifies all line features by TLID where the left and right side ZCTA Codes are unequal. The final product is a ZCTA-polygon data file. This file links the ZCTA codes to the CENID/POLYID for the TIGER/Line® files, and contains some additional data about the ZCTA delineation not included in the TIGER/Line files. For the Census 2000 Dress Rehearsal, we are making these supplementary files available to the public upon request.

Technical Contacts

For product information, contact the Geography Division's Products and Services Staff on (301) 457-1128 or through the internet at tiger@census.gov.

For comments or questions on the ZCTA delineation methodology contact Andrew Flora of TIGER Operations Branch on (301) 457-1100 or through the internet at aflora@geo.census.gov.

ZCTA™ is a trademark of the U. S. Bureau of the Census.

ZIP Code® is a registered trademark of the United States Postal Service.

TIGER/Line® is a registered trademark of the U. S. Bureau of the Census.

ZCTA Delineation

Census 2000 Dress Rehearsal

