

# Flowcharts/tables describing the Statistical Administrative Records System—1999 and the Administrative Records Experiment 2000

(To accompany AREX 2000 presentation  
materials)

```

graph TD
    TY98_IRS_1040[TY98 IRS 1040  
119,946,193] --> Edited_IRS_1040[Edited IRS 1040  
243,260,776]
    TY98_IRS_1099[TY98 IRS 1099  
598,075,971] --> Edited_IRS_1099[Edited IRS 1099]
    Medicare[Medicare  
56,837,022] --> Edited_Medicare[Edited Medicare]
    Selective_Service[Selective Service  
13,176,234] --> Edited_Selective_Service[Edited Selective Service]
    HUD_TRACS[HUD TRACS  
3,342,234] --> Edited_HUD_TRACS[Edited HUD TRACS]
    Indian_Health_Service[Indian Health Service  
3,106,821] --> Edited_Indian_Health_Service[Edited Indian Health Service]
    NUMIDENT[NUMIDENT  
676,589,439] --> Census_NUMIDENT[Census NUMIDENT  
396,185,872]

    Edited_IRS_1040 --> Address_Processing[Address Processing  
795,742,702]
    Edited_IRS_1099 --> Address_Processing
    Edited_Medicare --> Address_Processing
    Edited_Selective_Service --> Address_Processing
    Edited_HUD_TRACS --> Person_Processing[Person Processing  
875,750,973]
    Edited_Indian_Health_Service --> Person_Processing

    Address_Processing --> Hygiene_Unduplication[Hygiene & Unduplication  
136,154,293]
    Person_Processing --> SSN_Validation_PVS[SSN Validation (PVS)  
844,945,296 Valid (96.5%)  
30,805,677 Invalid (3.5%)]
    Census_NUMIDENT --> SSN_Validation_PVS
    Census_NUMIDENT --> Person_Characteristics_File_PCF[Person Characteristics File (PCF)  
396,185,872]

    ABI[ABI] --> Hygiene_Unduplication
    Code_1[Code 1] --> Hygiene_Unduplication
    TIGER[TIGER] --> Geocoding[Geocoding  
102,965,122 (75.6% Coded)  
33,189,171 (24.4% Uncoded)]
    Hygiene_Unduplication --> Geocoding

    SSN_Validation_PVS --> Unduplication[Unduplication  
279,601,038]
    Unduplication --> Remove_Deceased_Create_Composite_Record[Remove Deceased/Create Composite Record  
257,744,909]
    Geocoding --> Remove_Deceased_Create_Composite_Record
    Person_Characteristics_File_PCF --> Remove_Deceased_Create_Composite_Record

    Remove_Deceased_Create_Composite_Record --> Extraction_OF_AREX[Test Site Records  
1,459,760 in Baltimore Site  
1,229,234 in Colorado Site]
  
```

Page 1

**Table 1. Source File Characteristics**

<b>File</b>	<b>Targeted Population Segment</b>	<b>Address Records</b>	<b>Person Records</b>
IRS 1040	Taxpayer and other members of the tax reporting unit with current address	120 million	243 million
IRS W2/1099	Persons with taxable income who may not have filed tax returns	598 million	556 million
HUD TRACS	Low income housing population (possible non-taxpayers)	3 million	3 million
Medicare File	Elderly population (possible non-taxpayers)	57 million	57 million
IHS File	Native American population (possible non-taxpayers)	3 million	3 million
SSS File	Young male population (possible non-taxpayers)	14 million	13 million
<b>Total</b>		<b>795 million</b>	<b>875 million</b>
<b>Numident</b>	<b>SSN Lookup</b>	<b>N/A</b>	<b>677/396 million</b>

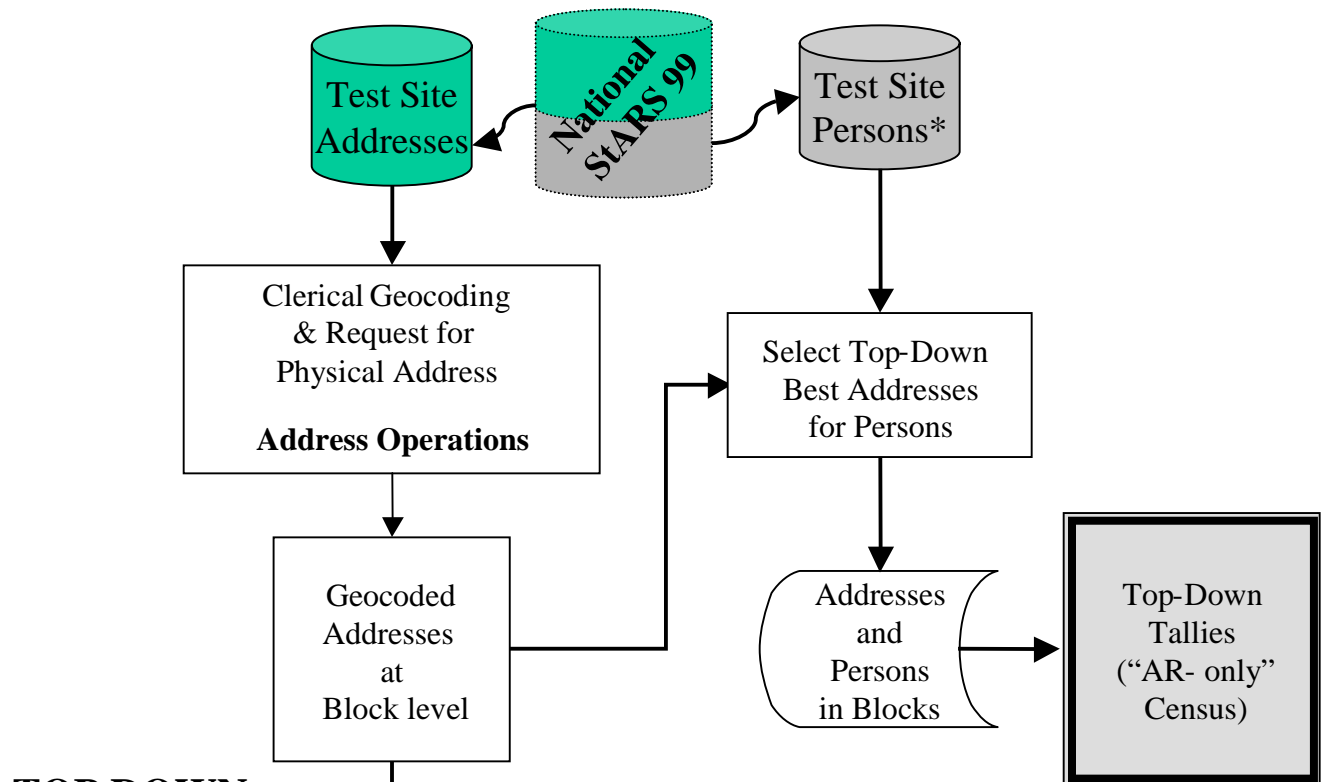
Notes: Variance between the number of address records and person records within input source files is a result of the following file anomalies:

1. The number of address records column is generally synonymous with the total record count on the input file.
2. Each IRS 1040 input record may reflect up to six persons (primary filer, secondary, and four dependents).
3. Each SSS input record may reflect two addresses - defined as current and/or permanent address.
4. The IRS W-2/1099 file undergoes a preliminary unduplication and clean-up prior to the initial file edit process. Prior to person processing, records are written “out of scope” if the SSN field is blank, the name standardizer returns a “bad name”, or the edited or input name field is blank. “Bad names” include institutional or firm names not recognized by the standardizer.

**Table 1.6.1: AREX 2000 Source File Reference Dates**

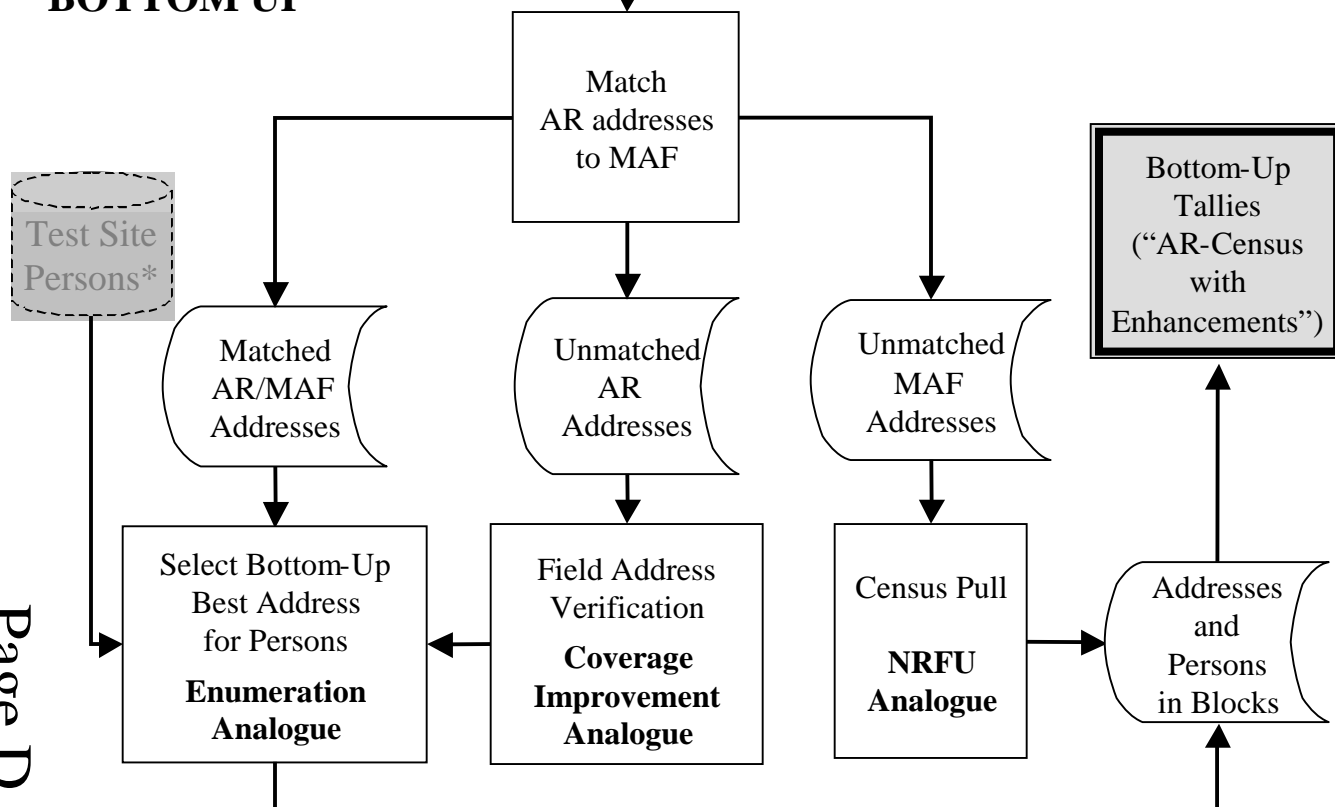
<b>Source File</b>	<b>Cut-off Date</b>	<b>Requested Cut Date</b>	<b>Universe</b>
Indian Health Svc.	04/01/99	04/01/99	All persons alive at cut-off date
Selective Service	Note 2	04/01/99	Males between the age of 18 - 25 <sup>2</sup>
HUD TRACS	04/01/99	04/01/99	All persons on file as of cut-off date
Medicare	Note 3	04/01/99	All persons alive at cut-off date
IRS 1040	12/98	09/30/99 <sup>1</sup>	Individual tax returns for tax year 1998
IRS W-2 / 1099	12/98	04/01/99	Forms W-2 and all 1099 forms tax year 1998

1. File Cut date is for posting cycle weeks 1-39 only for IRS 1040, and weeks 1-41 for IRS 1099 files. Weeks 40-52 (and 42-52 respectively) were not included in StARS '99. This file reflects the most current address on file for the taxpayer. It could be an address that has been updated since the 1998 tax return was posted.
2. Cut-off date is same as dates used to define universe: persons born after April 2, 1972 and on (or before) April 1, 1980.
3. Universe also defined as persons with a death date of 12/31/1989 or later.



**TOP DOWN**

**BOTTOM UP**



# AREX county population results

Census 2000		Top-Down		Bottom-Up	
		A-C Diff	%	A-C Diff	%
Total	2,558,212	-212,725	92	-25,763	99
Baltimore County	736,652	-40,469	95	-8,447	99
Baltimore City	625,401	-54,753	91	11,328	102
Douglas County	175,300	-27,030	85	-5,660	97
El Paso County	501,533	-44,642	91	-7,280	99
Jefferson County	519,326	-45,831	91	-15,704	97

## Household-level:

Effect on odds of “demographically matching”

- Estimated odds ratios

- Single unit: 2.6\*
- One or two persons in HH: 3.5
- No AREX imputed race: 2.1
- AREX one or more white: 2.1
- All AREX 65 and older: 1.7

- Interaction effects:

- Total effect of 65+,singleunit,nonimputed: **5.2**
- Total effect of 65+,1+white, 1-2 persons: **19.2**

Page F

\*(I.e. A single unit address is 2.6 times more likely than a multi-unit address to have exactly the same demographic distribution between the AREX and Census 2000 household, holding other effects constant; 1.0 = no positive or negative effect.)